

PHONOLOGY IN THE BILINGUAL AND BIDIALECTAL LEXICON

EDITED BY : Isabelle Darcy, Annie Tremblay and Miquel Simonet
PUBLISHED IN : Frontiers in Psychology





frontiers

Frontiers Copyright Statement

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-210-1

DOI 10.3389/978-2-88945-210-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

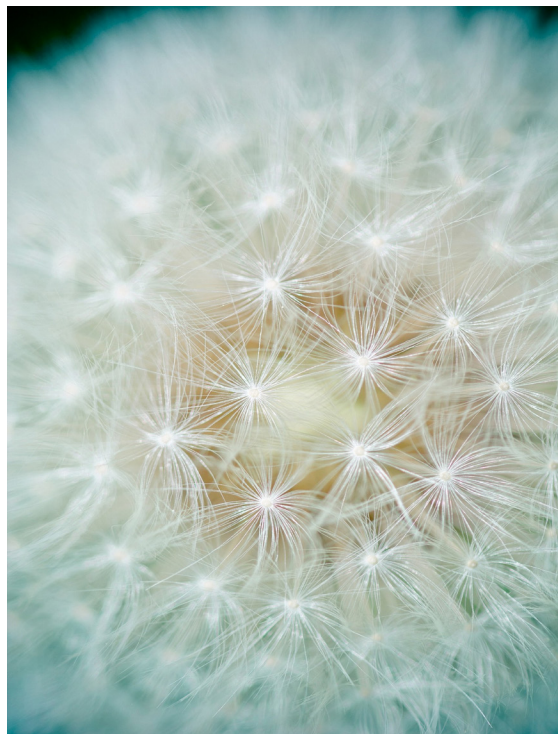
PHONOLOGY IN THE BILINGUAL AND BIDIALECTAL LEXICON

Topic Editors:

Isabelle Darcy, Indiana University, Bloomington, USA

Annie Tremblay, University of Kansas, USA

Miquel Simonet, University of Arizona, USA



Dandelion. The Ujulala. Pixabay.com

A conversation between two people can only take place if the words intended by each speaker are successfully recognized. Spoken word recognition is at the heart of language comprehension. This automatic and smooth process remains a challenge for models of spoken word recognition. Both the process of mapping the speech signal onto stored representations for words, and the format of the representation themselves are subject to debate. So far, existing research on the nature of spoken word representations has focused mainly on native speakers. The picture becomes even more complex when looking at spoken word recognition in a second language.

Given that most of the world's speakers know and use more than one language, it is crucial to reach a more precise understanding of how bilingual and multilingual individuals encode spoken words in the mental lexicon, and why spoken word recognition is more difficult in a second language than in the native language. Current models of native spoken word recognition operate under two assumptions: (i) that listeners' perception of the incoming speech signal is optimal; and (ii) that listeners' lexical representations are accurate. As a result, lexical representations are easily activated, and intended words are successfully recognized. However, these assumptions are compromised when applied to a later-learned second language. For a variety of reasons (e.g., phonetic/phonological, orthographic), second language users may not perceive the speech signal optimally, and they may still be refining the motor routines needed for articulation. Accordingly, their lexical representations may differ from those of native speakers, which may in turn inhibit their selection of the intended word forms. Second language users also have to solve a larger selection challenge—having words in more than one language to choose from. Thus, for second language users, the links between perception, lexical representations, orthography, and production are all but clear. Even for simultaneous bilinguals, important questions remain about the specificity and interdependence of their lexical representations and the factors influencing cross-language word activation. This Frontiers Research Topic seeks to further our understanding of the factors that determine how multilinguals recognize and encode spoken words in the mental lexicon, with a focus on the mapping between the input and lexical representations, and on the quality of lexical representations.

Citation: Darcy, I., Tremblay, A., Simonet, M., eds. (2017). *Phonology in the Bilingual and Bidialectal Lexicon*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-210-1

Table of Contents

06 Editorial: Phonology in the Bilingual and Bidialectal Lexicon

Isabelle Darcy, Annie Tremblay and Miquel Simonet

Interfaces and multidirectional relationships

09 Cross-Linguistic Influence in the Bilingual Mental Lexicon: Evidence of Cognate Effects in the Phonetic Production and Processing of a Vowel Contrast

Mark Amengual

26 Fuzzy Nonnative Phonolexical Representations Lead to Fuzzy Form-to-Meaning Mappings

Svetlana V. Cook, Nick B. Pandža, Alia K. Lancaster and Kira Gor

43 Phonetic Encoding of Coda Voicing Contrast under Different Focus Conditions in L1 vs. L2 English

Jiyoun Choi, Sahayng Kim and Taehong Cho

60 The Influence of the Pinyin and Zhuyin Writing Systems on the Acquisition of Mandarin Word Forms by Native English Speakers

Rachel Hayes-Harb and Hui-Wen Cheng

Lexical access

73 Phonotactic Constraints Are Activated across Languages in Bilinguals

Max R. Freeman, Henrike K. Blumenfeld and Viorica Marian

85 Cognate Costs in Bilingual Speech Production: Evidence from Language Switching

Mirjam Broersma, Diana Carter and Daniel J. Acheson

101 Effects of the Native Language on the Learning of Fundamental Frequency in Second-Language Speech Segmentation

Annie Tremblay, Mirjam Broersma, Caitlin E. Coughlin and Jiyoun Choi

Relevant dimensions in speech

116 Early Prosodic Acquisition in Bilingual Infants: The Case of the Perceptual Trochaic Bias

Ranka Bijeljic-Babic, Barbara Höhle and Thierry Nazzi

124 Limits on Monolingualism? A Comparison of Monolingual and Bilingual Infants' Abilities to Integrate Lexical Tone in Novel Word Learning

Leher Singh, Felicia L. S. Poh and Charlene S. L. Fu

140 Differences in the Association between Segment and Language: Early Bilinguals Pattern with Monolinguals and Are Less Accurate than Late Bilinguals

Cynthia P. Blanco, Colin Bannard and Rajka Smiljanic

157 Establishing New Mappings between Familiar Phones: Neural and Behavioral Evidence for Early Automatic Processing of Nonnative Contrasts

Shannon L. Barrios, Anna M. Namyst, Ellen F. Lau, Naomi H. Feldman and William J. Idsardi

173 More Limitations to Monolingualism: Bilinguals Outperform Monolinguals in Implicit Word Learning

Paola Escudero, Karen E. Mulak, Charlene S. L. Fu and Leher Singh



Editorial: Phonology in the Bilingual and Bidialectal Lexicon

Isabelle Darcy^{1*}, Annie Tremblay² and Miquel Simonet³

¹ Department of Second Language Studies, Indiana University, Bloomington, IN, USA, ² Department of Linguistics, University of Kansas, Lawrence, KS, USA, ³ Department of Spanish and Portuguese, University of Arizona, Tucson, AZ, USA

Keywords: second-language speech, bilingual and bidialectal lexicon, spoken word recognition, phonological knowledge, orthographic knowledge

Editorial on the Research Topic

Phonology in the Bilingual and Bidialectal Lexicon

One critical step when trying to comprehend a spoken message is to identify the words that the speaker intended. To recognize spoken words, listeners continuously attempt to map the incoming speech signal onto lexical representations stored in memory (McClelland and Elman, 1986; Norris, 1994): Words that partially overlap with the signal are activated until the lexical candidate that best matches the input wins over its competitors, a process known as *lexical competition*. Models of spoken-word recognition, most of which are based on native listener behavior, assume that lexical representations are stable, and contain at least the phonological form of words in citation. While lexical representations likely also contain other forms, for example the reduced forms found in conversational speech, it is a matter of debate whether native listeners encode spoken words exclusively as phonetically detailed exemplars (Johnson, 1997; Goldinger, 1998) or whether phonological abstraction also takes place (McQueen et al., 2006). Another assumption of models of native spoken-word recognition is that, under normal circumstances, listeners' perception of the input is optimal and faithful to the signal: Accurate lexical representations are easily contacted, and an optimal set of candidates is activated for quick lexical selection.

When applied to a later-learned second language (L2), two central premises of native spoken-word recognition models are compromised: (i) the premise that listeners' perception of the incoming speech signal is optimal; and (ii) the premise that listeners' lexical representations are accurate. L2 listeners are less successful at mapping the input to lexical representations, because they tend to perceive speech through their native-language (L1) phonetic categories and phonological representations. As a result, L2 listeners activate *more* and/or *different* lexical candidates than would native listeners (Broersma and Cutler, 2011). L2 listeners' knowledge of two languages further inhibits word recognition, as words from both lexicons are activated (Marian and Spivey, 2003). L2 listeners' perceptual difficulties in turn lead to the development of inaccurate or incomplete lexical representations. The fact that L2 listeners are often exposed to the orthographic form of words *before* they hear these words makes it difficult to determine the content of their lexical representations. Another relevant question is the potentially asymmetrical relationship between L2 listeners' lexical representations and their production of the same words. Thus, for L2 listeners, the links between perception, lexical representations, orthography, and production are all but clear. Even for simultaneous bilinguals, important questions remain about the specificity and interdependence of bilinguals' lexical representations and the factors influencing cross-language word activation.

This Frontiers Research Topic seeks to further our understanding of the factors that determine how bilinguals recognize and encode spoken words in the mental lexicon, with focus on the mapping between the input and lexical representations, and on the quality of lexical

OPEN ACCESS

Edited and reviewed by:

Manuel Carreiras,
Basque Center on Cognition, Brain
and Language, Spain

*Correspondence:

Isabelle Darcy
idarcy@indiana.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 20 February 2017

Accepted: 17 March 2017

Published: 04 April 2017

Citation:

Darcy I, Tremblay A and Simonet M
(2017) Editorial: Phonology in the
Bilingual and Bidialectal Lexicon.
Front. Psychol. 8:507.
doi: 10.3389/fpsyg.2017.00507

representations. Our call for papers resulted in 12 original contributions that represent a range of perspectives into the L2 mental lexicon and the interfaces between domains. The articles in this collection all present empirical research that revolves around three major themes.

The first theme targets interfaces and the multidirectional relationships between perception, lexical encoding, orthographic knowledge, and production. Four contributions fall under this theme. Amengual examines the interface between production and lexical encoding, and shows that even if a phonemic contrast is not part of the learner's lexical representations, it can be made in production. Cook et al. investigate the quality of phonological representations in the mental lexicon. They show that even when phonological contrasts can be perceived, learners' representations are less detailed than those of native speakers. The authors conclude that learners experience both fuzzy lexical representations and fuzzy form-to-meaning mappings. Choi et al. examine the interplay of information structure, meaning, and phonetics in the realization of word-final codas. They suggest that the L2 phonetic system can be better understood through an investigation of the phonetics-prosody interface that is further modulated by information structure and by the L2 speakers' L1 experience. Hayes-Harb and Cheng examine the interface between orthographic knowledge and the learning of new words. They show that, when establishing lexical representations, learners may need to suppress familiar orthographic information that can have interfering effects.

The second theme involves lexical access in the L1 and L2, and how it is impacted by the L1 and the developing L2 phonological systems. Three contributions fall under this theme. Freeman et al. show that L1 phonotactic knowledge impacts lexical searches during L2 word recognition. In that study, L1-Spanish L2-English bilinguals accessed their L1 Spanish phonotactic constraints during English comprehension, increasing lexical competition by activating both lexicons. A similar point is made in Broersma et al., who provide evidence for the occurrence of cross-language lexical competition in the speech of fluent Welsh-English early bilinguals: They report both facilitative and inhibitory effects in the production of cognates. The authors suggest that the shared phonological form of cognates may facilitate processing at the word-form level but result in lexical competition at the lexical-semantic level. Finally, Tremblay et al. demonstrate that L1-L2 similarities can *interfere* with segmentation processes during L2 word recognition: They show that the similarities between the prosodic systems of French and Korean make it more difficult for L1-Korean L2-French listeners to distinguish the two systems and *learn* to use the appropriate prosodic cues to word boundaries in French as compared to proficiency- and experienced-matched L1-English L2-French listeners.

The third theme deals with the speech dimensions that learners must learn to pay attention to, and how learners develop perceptual sensitivities to dimensions that matter for

the purpose of lexical acquisition. Five contributions fall under this theme. Bijeljac-Babic et al. present data about bilingual infants simultaneously acquiring German and French. They show that a trochaic bias found in monolingual German infants (but not in French monolingual infants) emerges at the same time in French-German infants, and that the amount of exposure to one or the other language has little impact on the emergence of the bias. Singh et al. examine phonological variation that is lexically relevant in one language but irrelevant in the other. They show that 12-to-13-month-old bilingual infants can bind tone to meaning in Mandarin words while disregarding tone variation in English words; in contrast, monolingual Mandarin learners did not integrate tones and word meanings at the same age. Their results suggest that, early on, infants selectively adjust which speech dimensions are relevant for lexical acquisition. Blanco et al. examine the possibility that adult bilinguals have more detailed phonological representations as a result of having to keep their two languages apart and having a more variable input on which to build these representations. Barrios et al. investigate how bilingual adults learn to reorganize their perceptual sensitivity to establish sound mappings that differ across their languages. They show that bilinguals are capable of establishing new mappings to phonemes for familiar phones. Finally, Escudero et al. deal with cross-situational novel-word learning in adults, also comparing monolinguals to bilinguals, and showing that bilinguals are more accurate than monolinguals at resolving conflicting information.

All the contributions focused on bilingual rather than bidialectal listeners, but similar issues could have been raised for bidialectal listeners. We might expect similarities between bilingual and bidialectal word recognition (e.g., cross-language activation), but also differences. For instance, bidialectal listeners may experience less difficulty than bilinguals in mapping the input to lexical representations and/or more phonetic interference across the two languages due to the greater phonetic similarity of the two dialects (relative to two languages). Future research should provide a detailed examination of bidialectal word recognition, which has received very limited attention.

The contributions in this Research Topic have provided diverse and broad-ranging insights from various perspectives into bilinguals' mapping of the speech signal onto lexical representations and the quality of their lexical representations. We hope that they will inspire much needed research in this exciting area.

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Broersma, M., and Cutler, A. (2011). Competition dynamics of second-language listening. *Q. J. Exp. Psychol.* 64, 74–95. doi: 10.1080/17470218.2010.499174

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251–279. doi: 10.1037/0033-295X.105.2.251

Johnson, K. (1997). "Speech perception without speaker normalization: an exemplar model," in *Talker Variability in Speech Processing*, eds

- K. Johnson and J. W. Mullennix (San Diego, CA: Academic Press), 145–165.
- Marian, V., and Spivey, M. (2003). Competing activation in bilingual language processing: within- and between-language competition. *Bilingual. Lang. Cogn.* 6, 97–115. doi: 10.1017/S1366728903001068
- McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0
- McQueen, J. M., Cutler, A., and Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cogn. Sci.* 30, 1113–1126. doi: 10.1207/s15516709cog0000_79
- Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52, 189–234. doi: 10.1016/0010-0277(94)90043-4

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Darcy, Tremblay and Simonet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Cross-Linguistic Influence in the Bilingual Mental Lexicon: Evidence of Cognate Effects in the Phonetic Production and Processing of a Vowel Contrast

Mark Amengual *

Bilingualism Research Laboratory, Department of Languages and Applied Linguistics, University of California, Santa Cruz, CA, USA

OPEN ACCESS

Edited by:

Annie Tremblay,
University of Kansas, USA

Reviewed by:

Joan Carles Mora,
Universitat de Barcelona, Spain
Melinda Fricke,
Pennsylvania State University, USA

*Correspondence:

Mark Amengual
amengual@ucsc.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 29 December 2015

Accepted: 13 April 2016

Published: 26 April 2016

Citation:

Amengual M (2016) Cross-Linguistic Influence in the Bilingual Mental Lexicon: Evidence of Cognate Effects in the Phonetic Production and Processing of a Vowel Contrast. *Front. Psychol.* 7:617. doi: 10.3389/fpsyg.2016.00617

The present study examines cognate effects in the phonetic production and processing of the Catalan back mid-vowel contrast (/o/-/ɔ/) by 24 early and highly proficient Spanish-Catalan bilinguals in Majorca (Spain). Participants completed a picture-naming task and a forced-choice lexical decision task in which they were presented with either words (e.g., /bosk/ “forest”) or non-words based on real words, but with the alternate mid-vowel pair in stressed position (*bosk/). The same cognate and non-cognate lexical items were included in the production and lexical decision experiments. The results indicate that even though these early bilinguals maintained the back mid-vowel contrast in their productions, they had great difficulties identifying non-words and real words based on the identity of the Catalan mid-vowel. The analyses revealed language dominance and cognate effects: Spanish-dominants exhibited higher error rates than Catalan-dominants, and production and lexical decision accuracy were also affected by cognate status. The present study contributes to the discussion of the organization of early bilinguals’ dominant and non-dominant sound systems, and proposes that exemplar theoretic approaches can be extended to include bilingual lexical connections that account for the interactions between the phonetic and lexical levels of early bilingual individuals.

Keywords: bilingualism, speech production, speech processing, cross-linguistic influence, mental lexicon, cognates, lexical storage

INTRODUCTION

A bilingual/multilingual individual must acquire two or more sound systems with differing sets of segments. Studies on the production and perception of language-specific phonological contrasts have examined early and late bilinguals differing in proficiency, age of acquisition, language dominance, amount of L2 input received, and other biographical non-linguistic variables in order to better understand cross-linguistic influence in bilingual speech (Flege, 1991, 2007; Flege et al., 1995, 1997, 1999; Guion, 2003; Flege and MacKay, 2004; Antoniou et al., 2011; Darcy and Krüger, 2012; Barlow, 2014; Simonet, 2014, 2015; Amengual and Chamorro, 2015, among others). In addition to producing and perceiving phonological categories specific to each of their languages, bilinguals

need to be able to establish lexical representations in their dominant and non-dominant language that encode language-specific phonemic contrasts. Following this assumption, recent studies have explored the dimension of the phonology/lexicon interface as opposed to experimental paradigms that focus exclusively on the categorization of phones without necessarily testing their linguistic function. This line of research seeks to determine how bilingual speakers encode words in their mental lexicon, how bilinguals resolve an increase in lexical competition due to having phonological representations of words in two different languages, and the impact of non-robust phonological representations with regard to bilingual lexical access (Weber and Cutler, 2004; Cutler et al., 2006; Escudero et al., 2008; Hayes-Harb and Masuda, 2008; Darcy et al., 2012; Amengual, 2015).

Prior research also suggests that not all lexical items are accessed and retrieved the same way, providing evidence of lexical effects in language acquisition and use. Some of these well-documented lexical effects include word frequency effects (Oldfield and Wingfield, 1965; Dell, 1990; Brysbaert et al., 2011), lexical neighborhood density effects (Baese-Berk and Goldrick, 2009; Peramunage et al., 2011; Scarborough, 2012), lexical bias effects (Vigliocco and Harsuiker, 2002; Nooteboom, 2005; Oppenheim and Dell, 2008), and cognate status effects (Dijkstra et al., 1999; Lemhöfer et al., 2004). Cognates, generally defined as lexical items with considerable phonological, semantic, and orthographic similarity (de Groot, 1995, p. 167), represent “the lexical overlap between languages” (Lemhöfer et al., 2004, p. 587). Given that many language pairs have lexical items that share form and meaning, these cognate words are likely to have a special status for bilinguals.

Facilitation effects with cognates have been widely studied in bilingual populations, particularly in psycholinguistic research. Word recognition and word naming experiments have shown that L2 cognate words are translated more rapidly and accurately than non-cognates (de Groot, 1992a,b), that there is faster (and more accurate) lexical access for cognate words compared to non-cognates in lexical decision tasks (Caramazza and Brunes, 1979; Dijkstra et al., 1998, 1999; de Groot et al., 2002), that cognates show greater repetition priming effects (Cristoffanini et al., 1986; Sánchez-Casas et al., 1992; de Bot et al., 1995), that cognates are easier to learn (de Groot et al., 2002), and that there are facilitatory effects of cognates in production (Costa et al., 2005), with cognates being named faster in word naming tasks (de Groot et al., 2002) and picture naming tasks (Costa et al., 2000; Hoshino and Kroll, 2008). Recent studies have also examined the effect of cognate status on the acoustic realization of phonetic segments, and the results support a cognate effect in bilingual speech production (Cochrane, 1980; Flege and Munro, 1994; Brown and Harper, 2009; Amengual, 2012; Mora and Nadeu, 2012; Goldrick et al., 2014; Brown and Amengual, 2015; Jacobs et al., 2016). These findings provide evidence of cross-language effects in the interface between the phonological and the lexical levels.

The phonetic variable under investigation in the present study is the Catalan-specific back mid-vowel contrast (/o/-ɔ/), which exists in Catalan but not in Spanish. Catalan stressed vowels have four degrees of height; with salient differentiation

in the mid-vowel area while the Spanish vowel system comprises the five cardinal vowels. There is a wealth of literature that has examined the production, perception, and processing of the Catalan mid-vowel contrasts showing that Spanish-Catalan bilinguals in Barcelona are merging /e/-ɛ/ to /e/ and /o/-ɔ/ to /o/ in their productions (i.e., producing Spanish-like mid-vowels) and they are reported to be failing to distinguish these Catalan-specific mid-vowel contrasts (Recasens, 1991; Pallier et al., 1997; Sebastián-Gallés and Soto-Faraco, 1999; Bosch et al., 2000). Furthermore, perception difficulties have been shown to also have consequences for lexical access. In a series of studies (Sebastián-Gallés and Baus, 2005; Sebastián-Gallés et al., 2005), Spanish-Catalan bilinguals in Barcelona participated in a lexical decision task involving Catalan words and non-words, in which non-words were based on real words but with the stressed vowel changed (i.e., the Catalan phoneme /e/ was substituted for /ɛ/, or vice versa). The results indicated that bilinguals in Barcelona had great difficulty distinguishing between words and non-words that differed by the Catalan front mid-vowel contrast (/e/-ɛ/), and Spanish-dominants overall exhibited a higher error rate than Catalan-dominants. These earlier findings in Barcelona may have been an artifact of the variety of Catalan being acquired. The study of the Catalan mid-vowels of early bilinguals in a different bilingual community, such as the one in Majorca, provides the opportunity to considerably reduce confounding factors that could have affected the previous results with Spanish-Catalan bilinguals in Barcelona. Due to differences in the historical evolution of the vowel systems in the dialects of Catalan, Majorcan Catalan has a vowel system and lexical distribution of these vowels that is distinct from the variety spoken in Barcelona. In addition, the Catalan mid-vowel contrasts in Majorca are more robustly maintained in the productions of these bilinguals in comparison to those in Barcelona (Herrick, 2003, 2006, 2007, 2008; Carrera-Sabaté and Fernández-Planas, 2005; Recasens and Espinosa, 2006, 2009; Amengual, 2011, 2013, 2016; Simonet, 2011, 2014). In a bilingual setting such as the one in Barcelona, Spanish-dominant speakers may receive highly variable and inconsistent Catalan input (i.e., Spanish-accented Catalan), which in terms of the Catalan mid-vowels lead to difficulties in the acquisition of the contrast (Bosch and Ramón-Casas, 2011).

Two recent studies examined the production, perception, and processing of the Catalan mid-vowels (/e/-ɛ/ and /o/-ɔ/) by early Spanish-Catalan bilinguals in Majorca. In Amengual (2016), 60 early Spanish-Catalan bilinguals in Majorca completed a categorical AXB discrimination task and picture-naming task to examine the perception and production of the Catalan front and back mid-vowel contrasts. The results showed that the Catalan-specific mid-vowels were more susceptible to discrimination difficulties than other vowel contrasts in the language. Even though these bilinguals were found to maintain robust mid-vowel contrasts in their productions, the degree of language dominance was found to have an effect on the acoustic distance maintained between the mid-vowels. Amengual (2015) explored the perception and processing of these mid-vowels by these same bilingual participants. Results from binary forced-choice identification, AX discrimination, and lexical decision tasks

indicated that even though these bilinguals demonstrated a high accuracy in perceptual identification and discrimination tasks, they had difficulties distinguishing between words and non-words in a lexical decision task, with Spanish-dominants exhibiting higher error rates than Catalan-dominants. If cognates are considered to be the crossroads of a bilingual's languages, these "special" lexical items may also be the locus where the bilingual phonologies are more likely to influence each other, affecting a bilingual individual's ability to produce, perceive, and process native-like targets, especially in their non-dominant language.

The present study examines the phonetic production and processing of the Catalan back mid-vowel contrast (/o/-/ɔ/) by 24 highly proficient early Spanish-Catalan bilinguals in Majorca (Spain) that are either Catalan-dominant or Spanish-dominant. Of central importance to this study, the production and lexical decision experiments investigate whether cognate lexical items increase phonetic interference in the acoustic realization and lexical representations of early and highly proficient bilinguals. The amount of overlap in the lexicon depends on the language pair of the bilingual. For instance, with closely related languages such as Spanish and Catalan, the lexicons share many words: between 60 and 85% of the words in the Catalan and Spanish lexicon are cognates (Lewis, 2009; Ramón-Casas et al., 2009). Although the phonological match between cognates in two languages is seldom perfect, correspondences noted between lexical items in two languages have been shown to more likely involve similarities at the phonological level rather than meaning or etymological history (Carroll, 1992). For this purpose, cognate items included as experimental stimuli in this study consist of words that are phonologically, orthographically and semantically similar. Examples of cognate lexical items are Catalan *boca* /bokə/ and Spanish *boca* /boka/ "mouth." Contrary to the cognate items, the Catalan non-cognate items included in this study are words that do not have an orthographically or phonologically similar translation equivalent in the other language (i.e., Catalan /pɔrk/ and Spanish /θerðo/ "pig"). This is not the first study to examine lexical effects in the production of a Catalan-specific mid-vowel contrast. For instance, cognate effects in the production of the Catalan front mid-vowel contrast (/e/-/ɛ/) were examined in Mora and Nadeu (2012). The study reports a cognate effect such that the group that used Spanish to a greater extent produced Catalan /ɛ/ significantly fronter (and thus with F2 values closer to /e/) in cognates than in non-cognates, and there were no significant differences between cognates and non-cognates in terms of vowel height (F1). Vowel height, however, is precisely the dimension that distinguishes Catalan /o/ and /ɔ/ (Recasens and Espinosa, 2006, 2009; Simonet, 2011, 2014).

The main questions that are explored in this study are the following: Is phonetic interference increased in the production of cognates? In other words, does cognate status have an impact on the acoustic realization of these mid-vowels? And also, does cognate status affect the lexical representations of Catalan words that include the Catalan-specific back mid-vowel contrast for these early Spanish-Catalan bilinguals? To the best of my knowledge there are no previous studies that have examined the phonetic production and processing of the same target cognate

and non-cognate lexical items in two groups of bilinguals that differ in language dominance. Because of the special status of cognates, it is reasonable to hypothesize that cognates will show different patterns of processing when compared to non-cognates. This cognate effect is expected to extend from the facilitation effects and processing advantages shown in previous psycholinguistic studies, demonstrating a cognate effect on the acoustic production and lexical representations of early bilinguals that affect the ability to maintain native-like contrasts in a language. The present study goes beyond Amengual (2015, 2016), Mora and Nadeu (2012) and Simonet (2011, 2014) in three ways: (i) in comparison to Mora and Nadeu (2012) it examines the phonetic production and processing of a Catalan-specific mid-vowel contrast in Majorcan Catalan, a dialect where the mid-vowels have a different distribution and where a robust contrast may be more available in the ambient input all bilinguals receive, (ii) it investigates the processing abilities of Catalan- and Spanish-dominant bilinguals involving the back mid-vowel contrast (/o/-/ɔ/), thus complementing the production and perception studies on these same back mid-vowels in Simonet (2011, 2014), and (iii) it adds the variable of cognate status to the analysis of these bilinguals' production and processing patterns, a factor that was not examined in Amengual (2015, 2016), in order to better understand the nature of Catalan-Spanish sound system interactions in this group of early and highly proficient bilinguals.

EXPERIMENT 1: PRODUCTION TASK

Method

Participants

A total of 24 male Spanish-Catalan bilinguals participated in the production experiment. All participants reported normal speech and hearing and normal or corrected to normal vision, and they all received monetary compensation for their participation in the study. Ages ranged from 18 to 35 ($M = 21.3$, $SD = 3.42$). All participants were born, raised, and educated in Majorca. They reported having extensive exposure to both languages on a daily basis, used Catalan and/or Spanish in the household, and were not native in any other language. This study focuses exclusively on male speakers due to the unbalanced number of male and female participants, which would make it impossible to consider "gender" as a factor if both were to be included.

In order to obtain information on the language dominance of the Spanish-Catalan bilingual participants, all participants completed the Bilingual Language Profile (BLP) questionnaire (Birdsong et al., 2012). The BLP is an instrument for assessing language dominance through self-reports and it produces a continuous dominance score and a general bilingual profile taking into account multiple dimensions: age of acquisition of the L1 and L2, frequency and contexts of use, competence in different skills, and attitudes toward each language (see Gertken et al., 2014 for more information). All of these factors are organized in four modules, which received equal weighting in the global language score (language history, language use, language proficiency, and language attitudes). The BLP was administered prior to the production and perception experiments, and was provided in Spanish or Catalan, depending on participant's preference. The

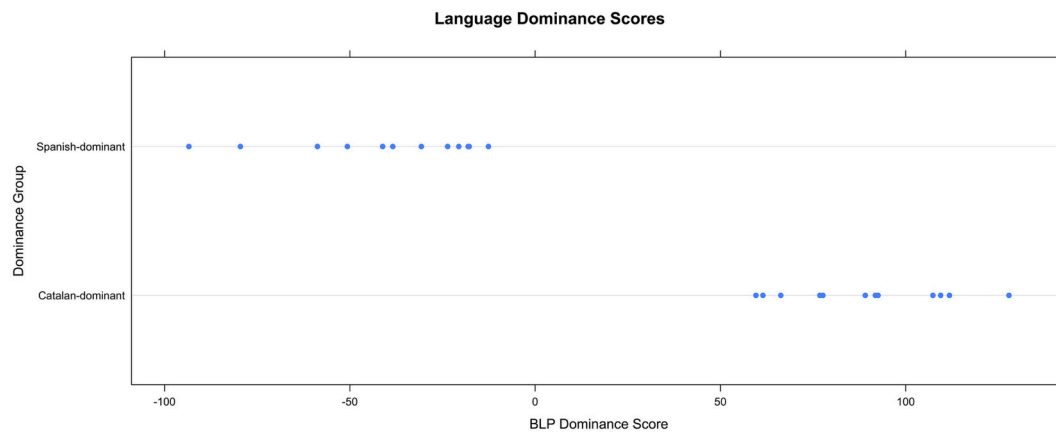


FIGURE 1 | Language dominance as a function of group according to the BLP (Birdsong et al., 2012).

TABLE 1 | Age, age of exposure, accent self-ratings, and typical daily use of both languages for each language dominance group.

	Catalan-dominant	Spanish-dominant
	<i>M (SD)</i>	<i>M (SD)</i>
Age	21.1 (1.6)	21.5 (4.6)
Age of exposure	CAT = 0 (0)	CAT = 1.8 (2.1)
	SPN = 1.2 (2.3)	SPN = 0 (0)
Self-reported accent (1 = strongly accented; 9 = native-like)	CAT = 8.3 (0.9)	CAT = 6.3 (2.4)
	SPN = 5.3 (1.9)	SPN = 8.5 (0.7)
Typical daily use (1 = only Spanish; 9 = only Catalan)	8.6 (0.8)	3.8 (2.5)

classification of participants as Spanish-dominant or Catalan-dominant was determined by the responses to the questionnaire, which generated a global score for each of the languages (Spanish and Catalan), a language particular score for each module, and a global score of dominance. The point system was converted to a scale score with the Catalan score subtracted from the Spanish score. Dominance scores ranged from -93.4 (strongly Spanish-dominant) to 127.8 (strongly Catalan-dominant). Participants with negative points were classified as Spanish-dominant while participants with positive points were classified as Catalan-dominant. **Figure 1** provides the distribution of the Spanish- and Catalan-dominant groups.

The main differences between the Catalan-dominant ($N = 12$) and Spanish-dominant ($N = 12$) groups were that Catalan-dominants were exposed earlier to Catalan than Spanish-dominants, the Catalan-dominant group reported a higher daily use of Catalan over Spanish, and also reported a more native-like accent in Catalan in comparison to the Spanish-dominant group. **Table 1** provides the language background for each language dominance group.

Materials

The production of the target Catalan mid-vowels /o/ and /ɔ/ in stressed position for cognate and non-cognate lexical items was elicited in a picture-naming task. The stimuli for this experiment consisted of illustrations representing non-ambiguous objects. Pictorial representations of lexical items were selected instead of the written form to avoid orthographic effects. In order to ensure that the Spanish-Catalan bilingual participants recognized the experimental items as cognates, 10 Spanish-dominant and 10 Catalan-dominant bilinguals that did not participate in the production or lexical decision experiments rated a list of Spanish-Catalan word pairs on a similarity scale (10 = “extremely similar,” 0 = “extremely different”). The ratings for the cross-language pairs were submitted to a one-way ANOVA to ensure that cognate and non-cognate items were rated differently. The analysis confirmed that the ratings for cognate pairs ($M = 9.25$, $SD = 0.38$) and non-cognate pairs ($M = 2.45$, $SD = 1.46$) were significantly different [$F_{(1, 18)} = 203.22$, $p < 0.001$]. The lexical conditions were also matched for word frequency, based on written word frequency in non-literary texts (Rafel i Fontanals, 1998). The lexical frequency of the cognate and non-cognate experimental items with /o/ and /ɔ/ were not significantly different [$F_{(1, 18)} = 1.99$, n.s]. The list of cognate and non-cognate stimuli is included in **Table 2**.

Procedure

The picture-naming task was conducted individually in a quiet room with participants comfortably seated in front of a computer display. Participants were told that the study involved naming pictures on a computer screen and that their speech would be recorded for subsequent acoustic analysis. All instructions and interactions between the participants and the researcher were in Spanish, independently of participants' language dominance. Spanish, instead of Catalan, was selected as the language to use when giving instructions and interacting with participants because Catalan-dominant bilinguals are generally more comfortable interacting in Spanish than Spanish-dominants are in Catalan. This decision was also made to

TABLE 2 | Stimuli included in the production and lexical decision tasks.

Catalan	Spanish	English	Target vowel	Cognate status
bota	bota	boot	/o/	Cognate
boca	boca	mouth	/o/	Cognate
ós	oso	bear	/o/	Cognate
copa	copa	glass	/o/	Cognate
doctor	doctor	doctor	/o/	Cognate
flor	flor	flower	/ɔ/	Cognate
escriptori	escritorio	desk	/ɔ/	Cognate
bosc	bosque	forest	/ɔ/	Cognate
sol	sol	sun	/ɔ/	Cognate
pilota	pelota	ball	/ɔ/	Cognate
poma	manzana	apple	/o/	Non-cognate
tassó*	vaso	glass	/o/	Non-cognate
tisores	tijeras	scissors	/o/	Non-cognate
papallona	mariposa	butterfly	/o/	Non-cognate
genoll	rodilla	knee	/o/	Non-cognate
porc	cerdo	pig	/ɔ/	Non-cognate
groc	amarillo	yellow	/ɔ/	Non-cognate
taronja	naranja	orange	/ɔ/	Non-cognate
foc	fuego	fire	/ɔ/	Non-cognate
oli	aceite	oil	/ɔ/	Non-cognate

*Most Catalan-Spanish bilinguals would consider Catalan "tassó" a cognate of Spanish "tazón" (Bowl/Mug) and the Catalan translation of Spanish "vaso" to be "got" (Glass). The translation of the Catalan word "tassó" into Spanish "vaso" is specific to Majorcan Catalan and it is expected that both Catalan- and Spanish- dominant bilingual participants in this study are familiar with this lexical pairing specific to the Majorcan dialect of Catalan.

minimize the potential impact of language mode on bilingual speech behavior, since language mode has been shown to influence the speech production and perception patterns of bilingual individuals (Soares and Grosjean, 1984; Grosjean, 1985, 1997, 1998, 2001, 2008).

Following the instructions in Spanish, participants were presented with the entire set of pictures in randomized order and each picture appeared together with the first letter of the target word. Each picture appeared on a computer screen for 5 s and participants were asked to name the experimental word in Catalan by embedding the target item in a carrier phrase, e.g., "Diuen TARGETWORD cada dia" "(They) say TARGETWORD every day," and to pronounce as clearly as possible and with a natural pace, speaking neither too quickly nor too slowly. Each session contained four randomized blocks. The Catalan block contained 20 experimental items eliciting the back mid-vowels in Catalan. Because each picture appeared four times (once in each block), each participant produced 80 tokens. A total of 1920 tokens were recorded from the productions of 24 Spanish-Catalan bilinguals. Because six tokens were excluded due to recording errors, or mispronunciations, the dataset comprised a total of 1914 measurements. The speech samples for all participants were recorded using a head-mounted microphone (Shure SM10A) and a solid-state digital recorder (Marantz PMD660), digitized (44 kHz, 16 bit quantization), and computer-edited for subsequent acoustic analysis.

Acoustic Analysis

Vowels were segmented with *Praat* (Boersma and Weenink, 2015) using synchronized waveform and spectrographic displays. *Praat* scripts were used to parse the recording of each participant into individual files for each target item. The boundaries of each vowel were determined by examining the waveform, spectrogram, and the intensity curve. Formant trajectories, especially the trajectory of the second formant (F2), as well as intensity displays were taken as indicators of vowel onsets and offsets. The onset of the vowel was marked as the beginning of the first voiced cycle where F2 was visible and/or the intensity was similar to that of the vowel's midpoint (for voiceless obstruents), after the release (for voiced stops), the beginning of the first cycle in which F2 was visible and darkened (for fricatives), and at the beginning of the increase in intensity (for nasals and laterals). The end of the vowel was marked by the disappearance of F2, on the last pitch period (before stops and voiceless fricatives), and the beginning of the decline in intensity and the lowering of F2 (before nasals and laterals). When the neighboring segment was an approximant, the onset and offset of the vowel was identified at the beginning of the transitional period between approximant and vowel. Finally for diphthongs, the formant values were calculated at the center-point of the steady-states (i.e., regions of stability with formant differences between time points close to zero) in the target vowel of the two adjacent vowels to avoid transitions. Vowel measurements (F1 and F2) were automatically extracted at the center of the steady-state period of the vowel, together with the duration of the vowel (in milliseconds) using a *Praat* script. Formant tracks were calculated with the Burg algorithm (Anderson, 1974) as built into the *Praat* program. The effective window length for the calculation was set at 25 ms, and was maintained across tokens and speakers. The maximum number of formants to be located by the formant tracker was always 5, and the ceiling was set at 5.0 kHz. Formant values were extracted in Hertz and were further converted to Bark, using the Hz-to-Bark function available in *Praat*. The bark scale is a logarithmic psychoacoustic scale that ranges from 1 to 24, and is a measure of frequency based on the critical bandwidths of hearing believed to reflect human perception (Zwicker, 1961; Traunmüller, 1990; Johnson, 2003). The effects of vocal tract-size differences caused by sex on the acoustics of vowels were minimized because the participant sample consisted exclusively of male speakers. This reduces the need for inter-speaker acoustic normalization procedures (Adank et al., 2004).

Results

In order to examine cognate effects in the productions of these bilinguals, datasets of by-subjects aggregates were created including the median F1 and F2 values over subjects as a condition of vowel and cognate status (four values per subject, two per vowel per cognate condition). The dataset was submitted to a mixed-model ANOVA with language dominance (Spanish-dominant, Catalan-dominant) as between-subjects factor, vowel (/o/, /ɔ/), and cognate status (cognate, non-cognate) as within-subjects factors, and subject as the random term. The results on the F1 and F2 data are reported separately below. **Figure 2** displays two contour maps plotting the distribution of the

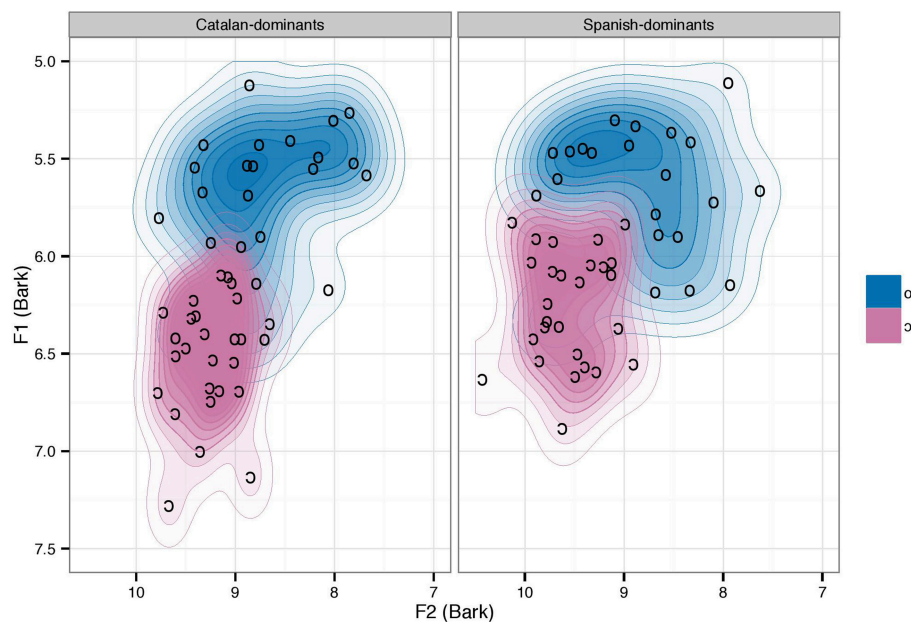


FIGURE 2 | Bark converted two-dimensional (F1 and F2) contour maps using kernel density estimation plotting the Catalan back mid-vowels as a function of language dominance.

Catalan back mid-vowels produced by male Catalan-dominant and Spanish-dominant bilinguals using kernel density estimation (KDE). Inspection of the two-dimensional contour maps shows that both groups maintain the Catalan-specific /o/-/ɔ/ contrast in their productions. This figure also suggests that the back mid-vowel contrast is more robust for Catalan-dominants than for Spanish-dominants, who show more overlap between the /o/ and /ɔ/ acoustic targets.

F1 (Vowel Height)

The mixed-design ANOVA yielded significant main effects of *vowel* [$F_{(1, 22)} = 110.97, p < 0.001$] and *cognate status* [$F_{(1, 22)} = 82.76, p < 0.001$], but not of *language dominance* [$F_{(1, 22)} = 2.69, n.s.$]. In addition, there was a significant interaction between *vowel* and *cognate status* [$F_{(1, 22)} = 39.44, p < 0.001$]. No other interactions were significant. The interaction was explored by analyzing the effects of cognate status and language dominance for each vowel separately. Therefore, the dataset was divided into two subsets as a function of vowel. For /o/, the model did not reveal any significant main effects or interactions. For /ɔ/, the analysis yielded a significant effect of *cognate status* [$F_{(1, 22)} = 145.18, p < 0.001$] and also an effect of *language dominance* [$F_{(1, 22)} = 24.39, p < 0.001$], but there was no significant interaction between *cognate status* and *language dominance* [$F_{(1, 22)} = 2.24, n.s.$]. These results indicate that both male Catalan-dominants and Spanish-dominants maintained robust height differences between /o/ and /ɔ/, in such a way that F1 varied as a function of the mid-vowel that was produced. Specifically, /o/ was significantly higher (lower F1 values) than /ɔ/ for both groups. Furthermore, /o/ and /ɔ/ were produced differently in terms of vowel height by each language dominance

group and cognate status was found to affect the F1 values of /ɔ/ but not /o/.

F2 (Vowel Fronting)

The analysis of F2 revealed a significant main effect of *vowel* [$F_{(1, 22)} = 85.88, p < 0.001$] and *cognate status* [$F_{(1, 22)} = 57.31, p < 0.001$], and an interaction between *vowel* and *cognate status* [$F_{(1, 22)} = 48.96, p < 0.001$], but no effect of *language dominance* [$F_{(1, 22)} = 3.35, n.s.$], and no other interactions. The interaction was explored by analyzing the effects of cognate status and language dominance for each vowel separately. Therefore, the dataset was divided into two subsets as a function of vowel. For /o/, the model revealed a significant effect of *cognate status* [$F_{(1, 22)} = 62.12, p < 0.001$], but no effect of *language dominance* [$F_{(1, 22)} = 1.12, n.s.$], or interaction [$F_{(1, 22)} = 3.34, n.s.$]. For /ɔ/, the analysis did not reveal any significant main effects or interactions. These results indicate that /o/ and /ɔ/ differed in F2, but there was no significant difference between the language dominance groups with respect to F2 (fronting). Finally, cognate status was found to affect the F2 values of /o/ but not /ɔ/. **Figure 3** displays two contour maps using kernel density estimation (KDE) to plot the Catalan back mid-vowels produced by male Catalan-dominant and Spanish-dominant speakers as a function of cognate status.

Because the investigation of group averages often obscures patterns of between-speaker variation, further analyses were carried out to investigate the extent to which the Catalan-specific /o/-/ɔ/ contrast is realized for each individual speaker. The Pillai score is a measure of the degree of merger (Hay et al., 2006; Hall-Lew, 2010; Sloos, 2013). The Pillai score is an output of a Multivariate Analysis of Variance (MANOVA) that represents

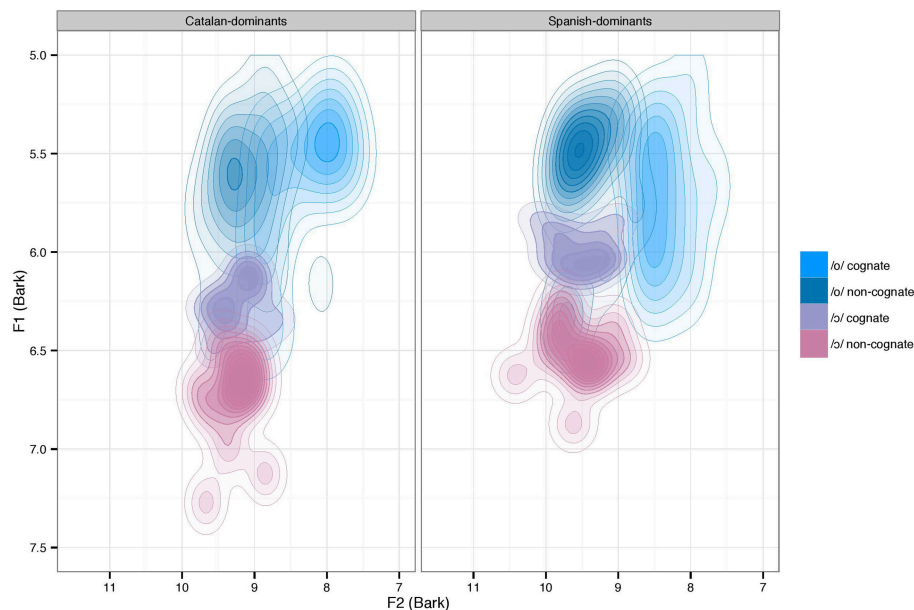


FIGURE 3 | Bark converted two-dimensional (F1 and F2) contour maps using kernel density estimation plotting the Catalan back mid-vowels as a function of cognate status and language dominance.

the degree of overlap between two vowel clusters. In addition to maintaining information about the vowel token cluster distribution, the Pillai score also accounts for phonological environment. The Pillai score representing the vowel cluster difference between /o/-/ɔ/ was calculated for each individual speaker, in which the higher the Pillai score, the lower the degree of overlap, and larger distinction, between the two vowel clusters. As **Figure 4** shows, the Pillai score is overall smaller for Spanish-dominant bilinguals (negative BLP score) than for Catalan-dominant bilinguals (positive BLP score), and every participant had a lower Pillai score for cognates (blue triangles) than for non-cognates (red circles). This indicates that each participant produced back mid-vowels with a higher degree of overlap in cognate lexical items. The Pillai value for cognate /o/ and /ɔ/ and for non-cognate /o/ and /ɔ/ in the productions of each individual speaker were correlated with that same speaker's language dominance score. The correlations between language dominance as reported in the BLP and Pillai score of the Spanish-dominant bilinguals showed that there was a significant correlation for cognates ($n = 12$, $df = 10$, $r = 0.70$, $R^2 = 0.49$, $p < 0.05$) and non-cognates ($n = 12$, $df = 10$, $r = 0.63$, $R^2 = 0.40$, $p < 0.05$). The analysis of the data from the Catalan-dominant group also revealed that there was a significant positive correlation between the /o/-/ɔ/ Pillai score and the BLP score in the production of cognates ($n = 12$, $df = 10$, $r = 0.62$, $R^2 = 0.39$, $p < 0.05$) as well as non-cognates ($n = 12$, $df = 10$, $r = 0.57$, $R^2 = 0.33$, $p < 0.05$). These results show that based on the information provided by the BLP, Spanish-dominants have a higher degree of overlap between these mid-vowels than Catalan-dominants. In addition, the language dominance continuum seems to be a strong predictor of the degree of overlap in the production of

the back mid-vowels, as the most Catalan-dominant bilinguals are the ones maintaining a more robust distinction between these mid-vowels.

EXPERIMENT 2: LEXICAL DECISION TASK

Method

Participants

Participants were the same Spanish-Catalan bilinguals that participated in Experiment 1.

Materials

The experimental stimuli for the lexical decision task consisted of the same list of 20 Catalan words used in the production experiment. The Catalan experimental items, which either contained the target mid-vowel /o/ or /ɔ/ in stressed position, were matched in word frequency and were further divided into cognate and non-cognate words according to similarity ratings (see Materials). The corresponding incorrectly pronounced words (i.e., non-words) were created by replacing the stressed mid-vowel with the other member of the contrast for each lexical item. For instance, the Catalan non-word */bosk/ was created from the real word /bɔsk/ "forest." Conversely, the correct pronunciation of /bokə/ "mouth" appeared alongside */bɔkə/ in the stimuli list. The complete list of experimental stimuli is presented in **Table 3**.

The auditory stimuli presented in the lexical decision task were obtained from the productions of three male native Majorcan Catalan speakers. The native speakers were asked to clearly enunciate the 40 experimental words (20 words and 20 non-words) providing 10 repetitions of each lexical item. The

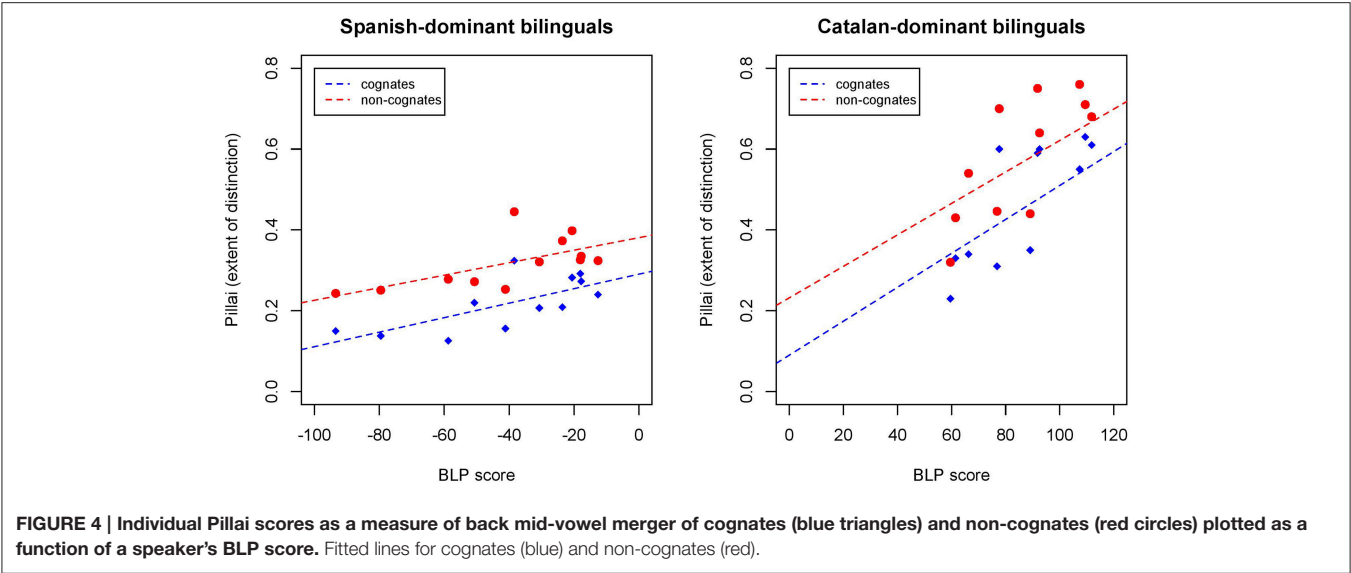


TABLE 3 | Experimental items used in the lexical decision task.

/o/-/ɔ/				Cognate status
/o/ word type		/ɔ/ word type		
Word	Non-word	Word	Non-word	
/o/→/o/	/o/ → */ɔ/	/ɔ/ → /ɔ/	/ɔ/ → */o/	
/botə/	*/bɔtə/	/flo/	*/flo/	Cognate
/bokə/	*/bɔkə/	/əskriptəri/	*/əskriptəri/	Cognate
/os/	*/ɔs/	/bosk/	*/bosk/	Cognate
/kopə/	*/kɔpə/	/sol/	*/sol/	Cognate
/dokto/	*/dɔktə/	/pilota/	*/pilota/	Cognate
/pomə/	*/pɔmə/	/pɜrk/	*/pork/	Non-cognate
/təso/	*/tɔsə/	/grɜk/	*/grok/	Non-cognate
/tizərəs/	*/tizɜrəs/	/təɾəndʒə/	*/təɾəndʒə/	Non-cognate
/pəpəˈɔnə/	*/pəpəˈɔnə/	/fɜk/	*/fok/	Non-cognate
/dʒənoːs/	*/dʒənoːs/	/ɔli/	*/oli/	Non-cognate

*indicates the incorrect mid-vowel (non-word).

recordings of the words and non-words were made using a Shure SM10A dynamic head-mounted microphone and a solid-state digital recorder (Marantz PMD660), and digitized at 44 KHz and 16 bits. In order to select the best “exemplars” for each word and non-word, three separate datasets (one for each speaker) were created including the median F1 and F2 values for each lexical item as a condition of vowel and vowel status (correct/incorrect). To ensure that there were only significant differences between /o/ and /ɔ/ productions independently of vowel status, each subset was submitted to a repeated measures ANOVA with F1 as the dependent variable, vowel (two levels: /o/ and /ɔ/) and vowel status (two levels: correct and incorrect). After confirming that the tokens selected based on the F1 median differed with respect to the vowel, but not because of vowel status (e.g., a mispronounced /ɔ/ vowel was not significantly different from a correctly pronounced /ɔ/ word), the same

dataset was submitted to a repeated measures ANOVA with F2 (Hz) as the dependent variable, and with vowel and vowel status as the independent variables. The statistical analyses again supported the initial selection of the median F1 as a measure to select the best exemplar of a word and non-word for each speaker. To summarize, the stimuli selected contained lexical items in which a properly pronounced /o/ was not different in height (F1) or fronting (F2) to a mispronounced target item produced with /o/ for any of the three speakers. The stimuli were normalized for peak intensity. If there was a DC offset, it was removed and the maximum amplitude was normalized to −0.5 dB at a project rate of 44 KHz. The picture stimuli that were presented together with the auditory stimuli consisted of the same pictorial representations employed in the picture-naming task.

Procedure

Participants completed the lexical decision task seated comfortably in front of a computer screen, and the stimulus presentation software *SuperLab 4.5* (Cedrus Corporation, USA) controlled the presentation of visual and auditory stimuli. Participants were told that the stimuli would consist of words and non-words, and that non-words were based on real words but with the stressed vowel changed (e.g., /o/ to /ɔ/, and vice versa). Participants were asked to classify each stimulus as being either a word or a non-word by pressing the right button on the USB Response Pad (RB-730) immediately after hearing a word stimulus, and the left button on hearing a non-word. The identity of the buttons was counterbalanced between subjects and the order of presentation was randomized for each participant. Participants responded to a total of 122 trials: 2 practice trials + 120 randomized test trials. Specifically, the experimental data consisted of 20 tokens × 2 type (correct/incorrect) × 3 voices = 120 responses per participant. As there were 24 participants, the dataset was comprised of 2880 data points.

Results

The lexical decision data were analyzed in a series of mixed-design ANOVAs, with language dominance (Spanish-dominant, Catalan-dominant) as between-subjects factor, vowel (/o/, /ɔ/) and cognate status (cognate, non-cognate) as within-subjects factor, and participant as the random term. The results for words and non-words are presented separately in order to analyze how Spanish-dominant and Catalan-dominant bilinguals differ in their categorization of mispronounced and properly pronounced words that vary exclusively in the Catalan back mid-vowel contrast. For this purpose, two datasets were created: the first one consisting of the responses to correctly produced real words, and the second one only including the responses to mispronounced words (i.e., non-words). The error rate (%) and response time data (ms) obtained from stimulus onset are presented for words and non-words.

Error Rate: Properly Pronounced Words (/o/→/o/ and /ɔ/→/ɔ/)

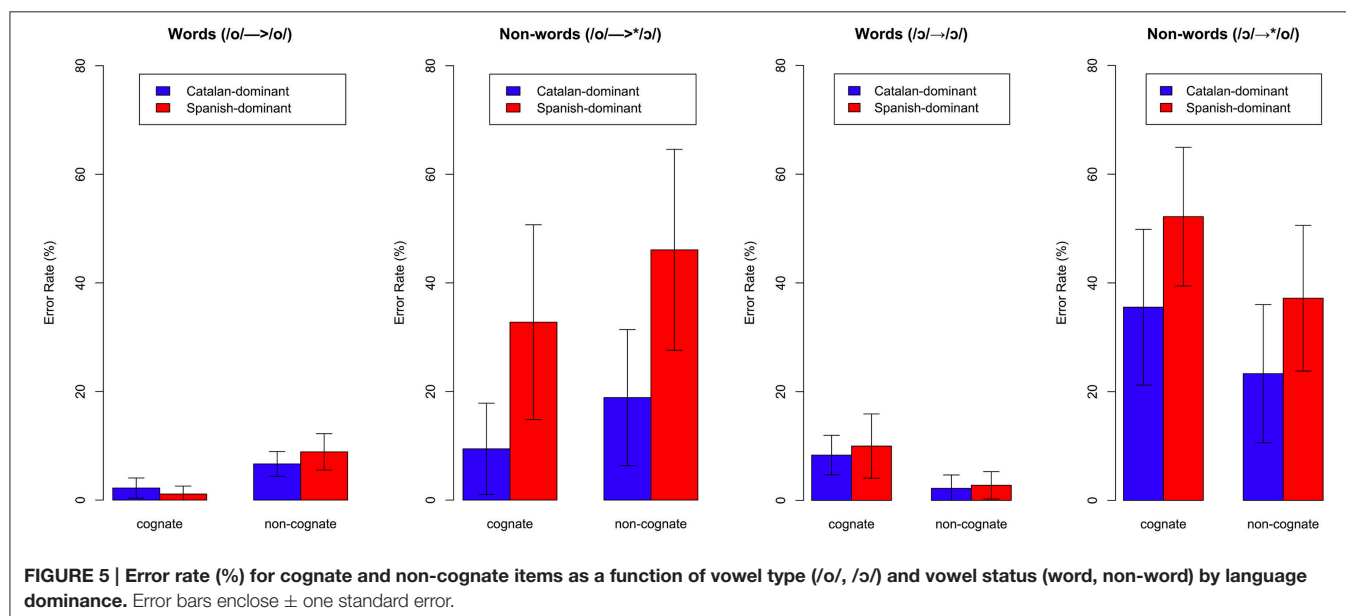
The analysis of the correctly produced /o/ and /ɔ/ stimuli did not yield significant main effects of *language dominance* [$F_{(1, 22)} = 0.29$, n.s.], *cognate status* [$F_{(1, 22)} = 0.14$, n.s.] or *vowel* [$F_{(1, 22)} = 0.70$, n.s.]. The model, however, did reveal a significant interaction between *vowel* and *cognate status* [$F_{(1, 22)} = 55.16$, $p < 0.001$]. The interaction between vowel and cognate status was explored by analyzing the effects of cognate separately for each vowel. Bonferroni-corrected paired *t*-tests confirmed that there were significant differences in the categorization accuracy of these bilinguals between cognates and non-cognates in /o/ type words [diff. = -6.10 , $t_{(23)} = -5.78$, $p < 0.001$], and also in /ɔ/ type words [diff. = 6.66 , $t_{(23)} = 5.53$, $p < 0.001$]. These results confirm that when responding to properly pronounced words these bilinguals made more mistakes in non-cognate than in cognate /o/ type words, but the effect was in the opposite direction in /ɔ/ type words: cognates elicited a higher error rate than non-cognates.

Error Rate: Non-Words (/o/→*/ɔ/ and /ɔ/→*/o/)

The analysis of the non-words revealed significant main effects of *language dominance* [$F_{(1, 22)} = 5.16$, $p < 0.05$] and *vowel* [$F_{(1, 22)} = 5.10$, $p < 0.05$], but the model did not yield a significant effect of *cognate status* [$F_{(1, 22)} = 1.59$, n.s.]. However, there was a significant interaction between *vowel* and *cognate status* [$F_{(1, 22)} = 49.92$, $p < 0.001$]. This interaction was explored by analyzing the effects of cognate separately for /o/→*/ɔ/ and /ɔ/→*/o/. Bonferroni-corrected paired *t*-tests confirmed that there were significant differences in the error rate between cognates and non-cognates in /ɔ/→*/o/ [diff. = -11.38 , $t_{(23)} = -5.05$, $p < 0.001$], and also in /o/→*/ɔ/ [diff. = 13.61 , $t_{(23)} = 8.35$, $p < 0.001$]. These results indicate that Spanish-dominant and Catalan-dominant bilinguals differed in their categorization of non-words in the lexical decision task. Spanish-dominant bilinguals in particular had great difficulties in recognizing mispronounced words that differed in the back mid-vowel contrast. Furthermore, cognate status was found to affect the categorization of /ɔ/ words incorrectly pronounced as /o/, and also /o/ words mispronounced as /ɔ/, but having an effect on the opposite direction. Cognates in /ɔ/ words incorrectly pronounced as /o/ showed a higher error rate than non-cognates indicating that having a cognate in Spanish with /o/ created more interference causing a higher proportion of non-words accepted as real words. In the case of /o/ words mispronounced as /ɔ/, the pattern showed that cognates elicited a lower error rate than non-cognates. **Figure 5** shows the error rate (%) in the categorization of words and non-words for each back mid-vowel as a function of cognate status, vowel status and language dominance.

Response Times

A dataset that included the median response times (ms) over subjects as a condition of vowel (/o/, /ɔ/) and word status (correct, incorrect) was created (four values per subject). The



median response times were calculated over accurate trials only, and a non-response was recorded if the participant did not press a key in the 2-s interval allowed. There were a total of 9 non-responses that were removed from the dataset. This dataset was submitted to a mixed-model ANOVA with language dominance (Catalan-dominant, Spanish-dominant) as between-subjects factor, vowel (/o/, /ɔ/), word status (correct, incorrect), cognate status (cognate, non-cognate) as within-subjects factors, and participant as the random term. The model yielded significant main effects of *language dominance* [$F_{(1, 22)} = 20.53, p < 0.001$], *vowel* [$F_{(1, 22)} = 8.60, p < 0.01$], *cognate status* [$F_{(1, 22)} = 38.30, p < 0.001$], and *word status* [$F_{(1, 22)} = 107.42, p < 0.001$]. In addition, there was a significant interactions between *vowel* and *cognate status* [$F_{(1, 22)} = 36.19, p < 0.001$]. The significant interaction was explored by analyzing the effects of cognate status for each vowel separately. For the /o/ type stimuli (/o/→/o/ and /o/→*/ɔ/), the model revealed a significant effect of *language dominance* [$F_{(1, 22)} = 11.58, p < 0.001$] and *word status* [$F_{(1, 44)} = 20.34, p < 0.001$], but no effects of *cognate status* [$F_{(1, 22)} = 2.10, n.s.$]. For the /ɔ/ type stimuli (/ɔ/→/ɔ/, /ɔ/→*/o/), there was a significant effect of *language dominance* [$F_{(1, 22)} = 28.22, p < 0.001$], *cognate status* [$F_{(1, 22)} = 130.6, p < 0.001$], *word status* [$F_{(1, 44)} = 32.22, p < 0.001$], and significant interactions between *language dominance* and *cognate status* [$F_{(1, 22)} = 11.7, p < 0.001$] and between *language dominance* and *word status* [$F_{(1, 44)} = 12.78, p < 0.001$]. These results show that Spanish-dominants took longer to respond to words and non-words that differed in the back mid-vowel contrast in comparison to Catalan-dominants. In addition, both groups had longer reaction times when responding to non-words than to real words. Finally, cognate effects were found in the response times of the /ɔ/ type stimuli, but these effects were not noticeable in the response times of the /o/ type words for both groups. **Figure 6** provides the response times (ms) as a function of vowel and cognate status for each language dominance group.

In order to investigate individual variation in the lexical decision task, the average error rate for words and non-words was calculated separately for each individual participant. The

individual error rate (%) in the lexical decision task was correlated with the participants' language dominance score as reported in the BLP. As **Figure 7** shows, the error rates are in general higher for Spanish-dominant bilinguals (negative BLP score) than for Catalan-dominant bilinguals (positive BLP score), and also both groups display a much higher error rate when responding to non-words than to correctly pronounced words. The correlations between BLP score and error rate for words and non-words as a function of language dominance are presented in **Table 4**.

The correlations between BLP score and error rate in the lexical decision task revealed that there was not a significant correlation for the Catalan-dominant or Spanish-dominant group in any of the stimuli, except for a significant correlation for the Spanish-dominants responding to both types of non-words (/o/→*/ɔ/ and /ɔ/→*/o/). These results show that there was a higher error rate in the lexical decision task as a function of being more Spanish-dominant, but this was only the case when responding to non-words. Further analyses also determined that there was not a significant correlation between the response time data with the error rate, that is, individuals who were faster at responding did not necessarily obtain lower or higher error rates.

The relationship between the speech production and perception of these early bilinguals was also examined. The Pillai scores of each individual speaker were compared to their error rates in the lexical decision task, collapsing words and non-words, for both cognates and non-cognates. The analyses revealed that there was a significant correlation between the Pillai score and accuracy in the lexical decision task for cognates ($n = 24, df = 22, r = -0.50, R^2 = 0.25, p < 0.05$) and non-cognates ($n = 24, df = 22, r = -0.51, R^2 = 0.26, p < 0.05$). **Figure 8** plots the accuracy rate in the lexical decision task and the individual speaker's Pillai score between /o/ and /ɔ/ as a function of cognate status. These results indicate that there is a correlation between the production of the back mid-vowel contrast and the ability to recognize properly pronounced and mispronounced words: bilinguals who produced the Catalan back mid-vowel contrast

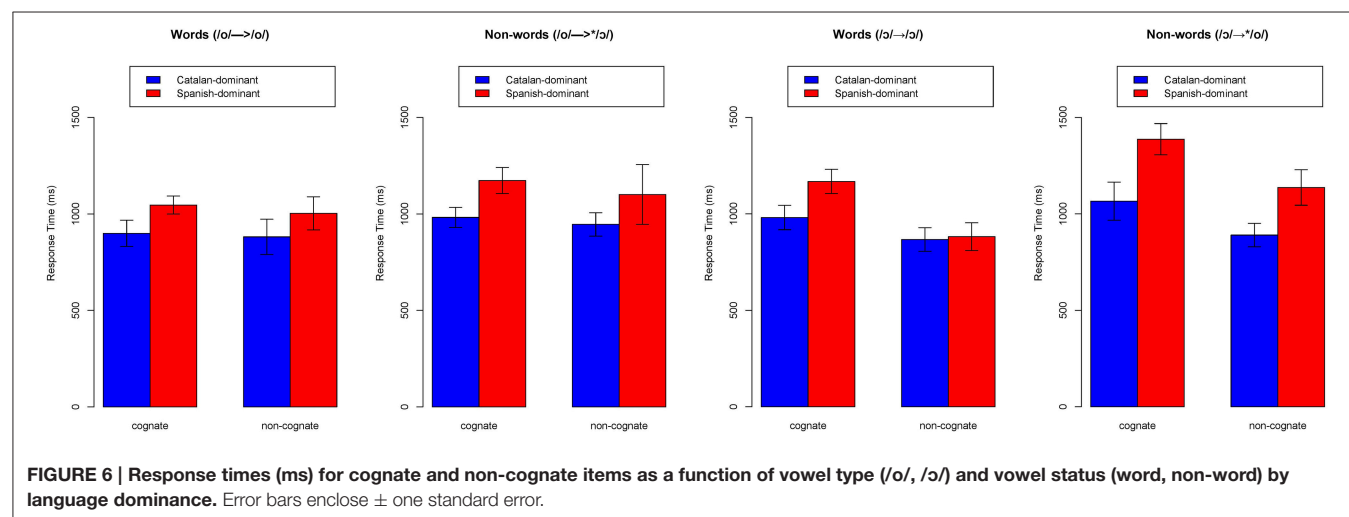


FIGURE 6 | Response times (ms) for cognate and non-cognate items as a function of vowel type (/o/, /ɔ/) and vowel status (word, non-word) by language dominance. Error bars enclose ± one standard error.

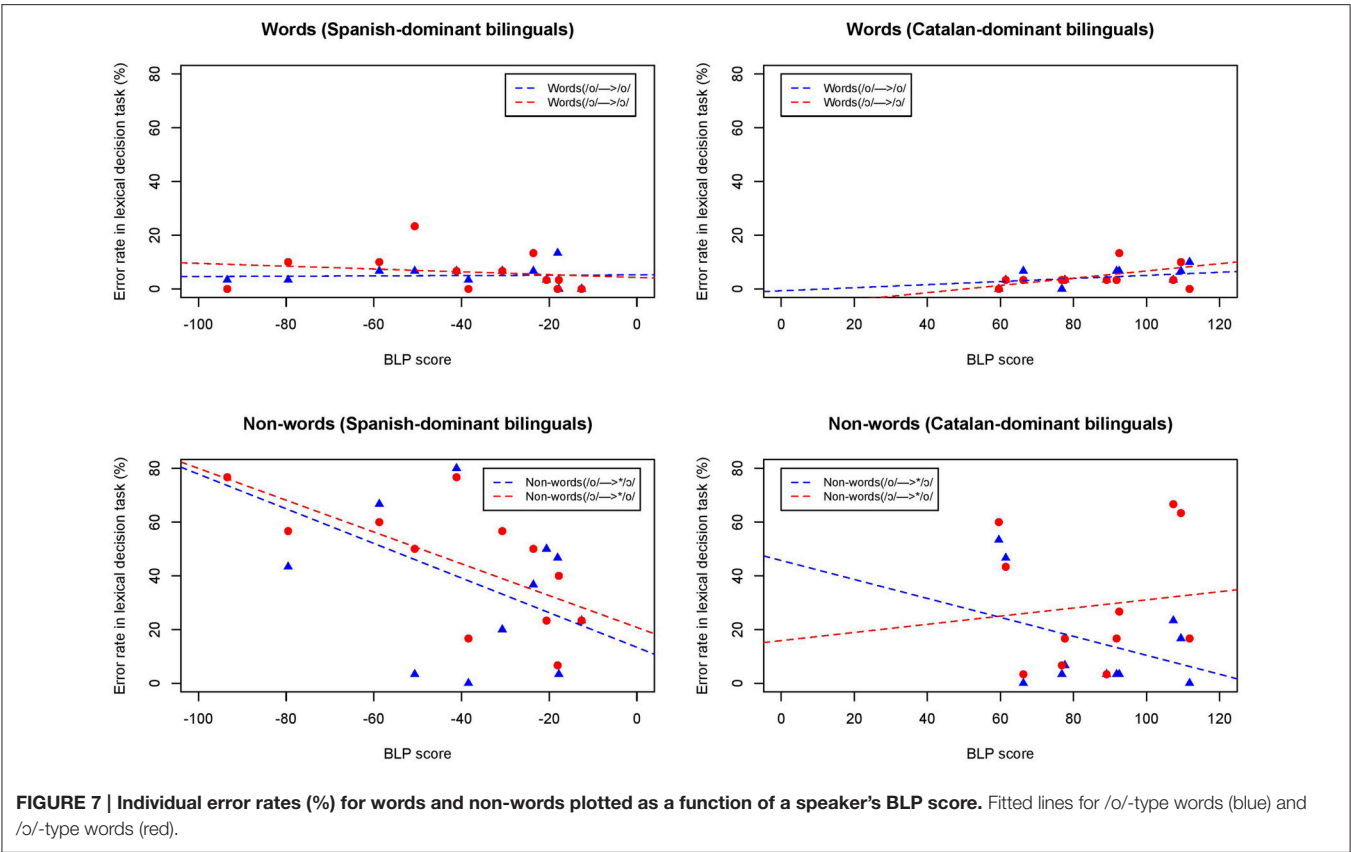


TABLE 4 | Results from the correlations between BLP score and error rate for words and non-words.

Stimulus	Spanish-dominant bilinguals	Catalan-dominant bilinguals
Words (/o/→/ɔ/)	$n = 12, df = 10, r = 0.04, R^2 = 0.002, n.s$	$n = 12, df = 10, r = 0.42, R^2 = 0.17, n.s$
Words (/ɔ/→/o/)	$n = 12, df = 10, r = -0.19, R^2 = 0.03, n.s$	$n = 12, df = 10, r = 0.56, R^2 = 0.25, n.s$
Non-words (/o/→*/ɔ/)	$n = 12, df = 10, r = -0.52, R^2 = 0.27, p < 0.05$	$n = 12, df = 10, r = -0.42, R^2 = 0.17, n.s$
/ɔ/-type (/ɔ/→*/o/)	$n = 12, df = 10, r = -0.66, R^2 = 0.44, p < 0.01$	$n = 12, df = 10, r = 0.14, R^2 = 0.01, n.s$

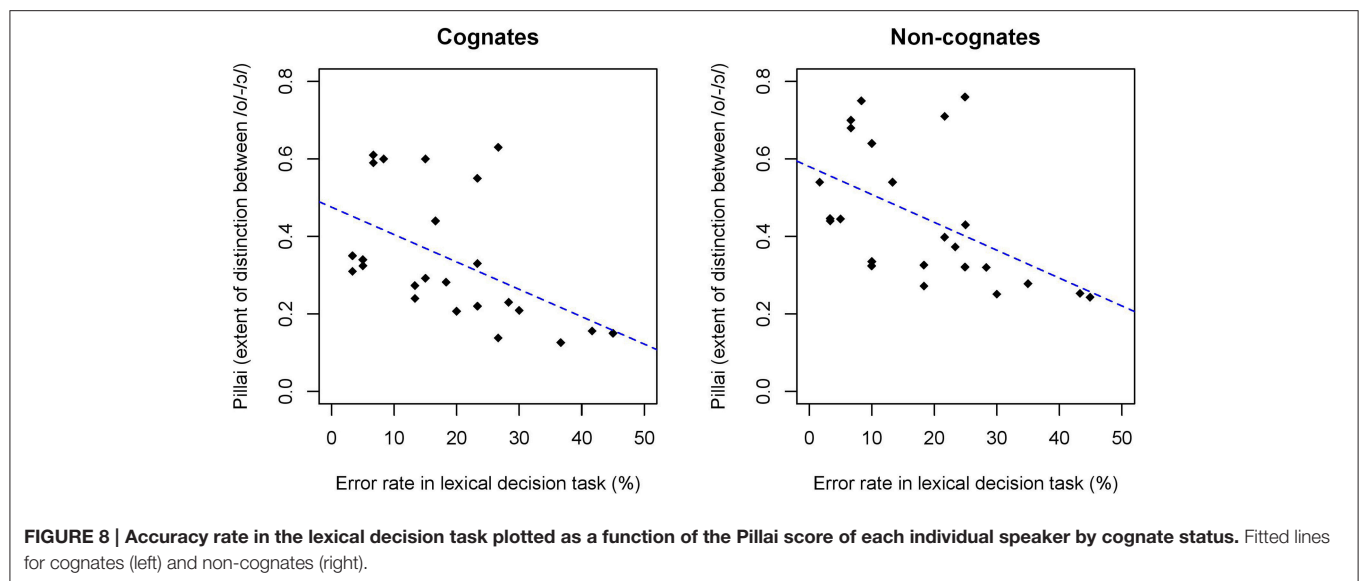
with a higher degree of overlap (i.e., smaller Pillai score) were more likely to have a higher error rate when responding to cognates and non-cognates in the lexical decision task.

DISCUSSION

Everyday linguistic performance involves much more than the ability to concentrate on isolated phonetic segments in speech perception and production experiments. In human communication, a combination of sounds are necessarily embedded in words, so beyond the ability to discriminate stimuli and produce acoustic targets, speakers must also encode these language-specific phonemes in the form of spoken words in their mental lexicon. Therefore, a language user must seamlessly learn which combination of vowel and consonant units are contained in a given word, and also be able to recognize which words include a specific phonemic category. Spanish-Catalan

bilinguals must acquire two vowel systems with a different set of segments, and crucially, they must learn to select the correct vowel depending on the lexical item that is going to be pronounced. This study probes if Spanish-Catalan bilinguals are able to produce and recognize the appropriate Catalan-specific mid-vowel in lexical items in general, and if cognates in particular enhance cross-linguistic influence.

The present study investigated cognate effects in a picture-naming and lexical decision task on the Catalan back mid-vowel contrast (/o/-/ɔ/) by 12 Spanish-dominant and 12 Catalan-dominant male Spanish-Catalan bilinguals from Majorca (Spain), complementing the findings from previous studies in the same bilingual setting (Amengual, 2015, 2016). These early and highly proficient bilinguals have been raised in a bilingual community where they have been exposed to both Catalan and Spanish before the age of 4. The results from recent studies in Majorca, and contrary to previous findings in Barcelona, indicate that both Spanish-dominants and Catalan-dominants maintain robust



mid-vowel contrasts in their productions and also demonstrate high perceptual accuracy when completing identification, AX discrimination, and AXB discrimination tasks. However, even though these bilinguals perform at ceiling in the perceptual tasks that consist of identifying and discriminating between isolated phonemes, their performance decreases in the lexical decision task. This is consistent with previous research showing that even high accuracy in phonetic categorization will not guarantee accurate lexical encoding of a difficult L2 contrast (Darcy et al., 2013). Adding to the previous literature, this study posed a different question regarding the phonetic production and processing abilities of these early bilinguals: Do cognates increase phonetic interference in the acoustic realization and lexical representations of these bilinguals? To answer this question, cognates and non-cognates were examined to detect cross-language influence. Non-cognates such as Catalan *poma* /pomə/ “apple” (Spanish *manzana* /manθana/) were investigated alongside cognates, such as *bosc* /bɔsk/ “forest” (Spanish *bosque* /boske/).

The results of the picture-naming and lexical decision tasks provide evidence of cognate effects in both the phonetic production and processing of the Catalan back mid-vowel contrast. This cross-linguistic influence was robust for both language dominance groups when selecting the appropriate phonetic representations of lexical items in order to produce the experimental stimuli as well as when identifying aurally presented stimuli either as a word or a non-word. Cognate status was found to influence both the vowel height and fronting for the Catalan back mid-vowels /o/ and /ɔ/ in the productions of both Spanish-dominant and Catalan-dominant bilinguals. The cognate status effect was especially robust in the production of the Catalan-specific /ɔ/. The production data showed that /ɔ/ in cognate lexical items were produced significantly higher than non-cognates, approximating the /o/ acoustic region. In other words, the cognate items were taking a different direction than

non-cognates, reducing the acoustic distance between /o/ and /ɔ/. Further evidence of phonetic interference at the lexical level was found in the lexical decision task. Results show that when responding to cognates in /ɔ/ words incorrectly pronounced as /o/ there was an increased cross-linguistic interference between the mid-vowel categories causing a higher error rate and longer response times. In this case there was a higher proportion of non-words accepted as real words. The opposite effect was found in the case of /o/ words mispronounced as /ɔ/. In this case, the pattern showed that cognates increased lexical decision accuracy in comparison with non-cognates. Taken together these results suggest that congruent cognates (cognates that contain a stressed mid-vowel in Spanish and a higher-mid vowel in Catalan, i.e., /o/-/o/) increased the lexical decision accuracy, facilitating lexical access, whereas incongruent cognates (cognates that contain a stressed mid-vowel in Spanish and a lower-mid vowel in Catalan, i.e., /o/-/ɔ/) increased cross-linguistic interference between the mid-vowel categories, causing a higher error rate in the lexical recognition process. The results from the reaction time data also show an effect of language dominance and word type: Spanish-dominant bilinguals took longer to respond to the stimuli than Catalan-dominants and both groups had a longer response latency with non-words (i.e., lexical items based on real words, but with the alternate mid-vowel pair) than real words. Finally, both groups took longer to respond to cognates in the /ɔ/ type stimuli, but these effects were not noticeable in the response times of the /o/ type words.

Analyses of individual data showed that the degree of language dominance as a function of a participant's BLP score had an effect on the error rate in the lexical decision task. Specifically, those participants that were more Spanish-dominant were the ones that were most likely to have a higher error rate when responding to non-words. Similarly, the degree of language dominance was a strong predictor of the acoustic distance and overlap maintained between both phonemes. The Pillai score, which measures

the degree of merger between two vowel clusters significantly correlated with the degree of language dominance. For Spanish-dominants there was a significant correlation between the degree of overlap of the /o/-/ɔ/ and the degree of Spanish dominance, as operationalized by the BLP. Similarly, for the Catalan-dominant group there was a more robust distinction between the back mid-vowels as a function of being more Catalan-dominant. Cognate effects were also evident in the individual data, as both Catalan-dominants and Spanish-dominants produced /o/ and /ɔ/ with a higher degree of overlap (i.e., lower Pillai score) in cognate than in non-cognate lexical items. Finally, the present study also examined the relationship between the phonetic production and perception abilities of each bilingual individual. The correlations between the production and lexical decision data indicate that there is a tight link between the production of the back mid-vowel contrast and the ability to recognize properly pronounced or mispronounced cognates and non-cognates in a lexical decision task. These findings provide evidence that cross-language phonetic interference occurs when early Spanish-Catalan bilinguals access their mental lexicon. The acoustic properties of cognate lexical items result in phonetic alterations in the lexical representations of these bilingual individuals.

Such an effect must be operationalized in a model of the bilingual lexicon that accounts for the variable production and lexical decision patterns linked to the bilinguals' lexical representations. The Perceptual Assimilation Model (PAM; Best, 1995), Perceptual Assimilation Model of Second Language Speech Learning (PAM-L2; Best and Tyler, 2007), and the Speech Learning Model (SLM; Flege, 1995) are models of cross-linguistic speech perception and production that assume that the learnability of new sounds in the L2 is perceptual in nature and depends on the perceived phonetic distance between the sounds in the L2 and the most similar segments in the L1 phonetic inventory. Despite these common assumptions, these models address different aspects of L2 phonological acquisition: the SLM focuses on individual phonetic categories whereas the PAM and PAM-L2 focus on pairwise phonological contrasts, and the SLM was primarily designed to address L2 production, whereas the PAM and PAM-L2 have a main focus on non-native speech perception and L2 perception respectively. The SLM, PAM, and PAM-L2 make straightforward predictions about the learnability of L2 sounds depending on the perceived similarity between the sounds of the L1 and L2. However, these models cannot account for an interaction between the phonological and lexical levels of representation across the two languages of a bilingual individual. In other words, these models cannot predict the phonetic interference found in the production and lexical decision of cognate lexical items, nor how the acoustic characteristics of the Catalan mid-vowels are related to the lexical representations stored in the bilingual mental lexicon. How can these results be theoretically interpreted?

Cognate facilitation effects in bilingual speech production have previously been explained with spreading activation models of speech production, such as cascaded activation models of lexical access (Dell, 1986; Goldrick and Blumenstein, 2006), in opposition to a strictly discrete activation model (Levelt, 1989; Levelt et al., 1999). Crucially, the differences between these

theoretical approaches are that the discrete models would not predict that lexical variables such as cognate status could affect its phonetic realization, because in this view, selection is made at the lexical level before articulation. As a result sublexical representations become active only after the target word has been selected. The cascaded activation models propose that processes at the lexical and phonological levels of planning can cascade down to affect the articulatory realization of acoustic targets. For instance, Jacobs et al. (2016) investigated effects of cross-language activation in the productions of L2 Spanish speakers of differing proficiencies (highly proficient speakers, intermediate learners in a domestic immersion program, and intermediate speakers in a classroom setting). Because the results from their study show effects of cognate status only in the articulation of the intermediate classroom learners of Spanish but not with the other groups, the authors argue that the speech production system of these bilinguals is cascaded, but that it exhibits “staged vs. cascading behavior as a function of task difficulty” (Jacobs et al., 2016, p. 25). A recent study, however, questions the cascading nature of the planning system. Buz and Jaeger (2016), using a picture-naming experiment, investigate the effects of phonological neighborhood density and provide evidence that the effect of phonological neighborhood density on word duration and vowel dispersion does not seem to be mediated through lexical planning (Buz and Jaeger, 2016), but admit that word-specific phonetic representations are compatible with their findings.

Assuming that lexicons in different languages are mentally interconnected (Costa et al., 2005; Jarvis and Pavlenko, 2008), lexical representations in one language are predicted to affect the lexical representations in the other. Exemplar models of lexical representation (Goldinger, 1997, 1998; Johnson, 1997a,b; Bybee, 2001; Pierrehumbert, 2001, 2003a,b; Hawkins, 2003) are theoretic approaches that are able to explore the lexical/phonetic interface in which the mental lexicon is represented phonetically. For the purpose of this study, the model is expanded to include bilingual data in order to analyze the interactions between the lexical representations of both languages in the bilingual lexicon. Adapting the exemplar model to bilingual lexicons can account for the interaction between the phonological and lexical levels of representation across a bilingual's languages and can explain the findings in the Majorcan bilingual phonetic production and processing of cognates and non-cognates.

Exemplar models assume that speech perception and production are closely linked. Clusters of similar experiences—that is, “exemplars” of the same word—are formed including productions that share a particular acoustic property. These exemplars are categorized by their similarity to extant stored exemplars so that clouds of memory traces group similar exemplars close to each other while dissimilar ones are more distant. The exemplars themselves include much more than just purely phonetic information: the representation of a specific word includes its meaning(s) and all the acoustic, lexical, social, and contextual information from the perceptual event (Ettlinger and Johnson, 2009). Exemplar models assume that when a new stimulus is presented, the memory traces (i.e., exemplars) are activated in proportion to their similarity to the stimulus, and

the pattern of activation is used to determine the category membership of the exemplar. This automatically eliminates a separation between pre-lexical and lexical phonological processing abilities (Mehler, 1981; McClelland and Elman, 1986; Pisoni and Luce, 1987; Norris, 1994; Gaskell and Marslen-Wilson, 1997). Such a model accounts for how speakers might possess fine-grained, detailed, and word-specific knowledge about the sounds and words of their language and require no phonological abstraction prior to lexical access (Pierrehumbert, 2001; Coleman, 2002; Johnson, 2007).

The application of an exemplar-based approach to the production and perception of early Spanish-Catalan bilinguals might assume mostly distinct exemplar clouds representing Catalan and Spanish. However, since these clouds are organized by the phonetic similarity of the exemplars and also include semantics, there is likely to be an overlap between the two otherwise independent language systems with respect to cognates. Since cognates by their very nature have the same meaning and similar phonetic forms in the two languages, the exemplar clouds for such cross-linguistic pairs (e.g., Catalan /sol/ “sun” and Spanish /sol/ “sun”) may in fact overlap, such that exemplars from both languages exist in the same perceptual space. Thus, bilingual production and lexical decision of cognates potentially draws from both Catalan and Spanish exemplars instead of restricting the possible targets to the language-specific exemplars available for each language separately.

The results reported in this study indicate that the cognate status of a lexical item influences the production targets and the selection of the correct phonetic category in a lexical decision task. In the picture-naming task, the phonetic output of a specific lexical item of a Spanish-Catalan bilingual is the average over the set of exemplars in the vicinity of a randomly selected exemplar. Therefore, cognate effects would result from the selection of a region in the exemplar space, and specifically the average over this region containing overlapping acoustic properties. For example, the acoustic properties of the target word /sol/ “sun” might be influenced by the average over the region in the exemplar space that contains instances of /sol/ exemplars from Spanish, as opposed to the Catalan word /oli/ “oil,” where the average from the exemplar space would not be affected by the acoustic properties of Spanish exemplars in the cloud of memory traces containing a back mid-vowel (Spanish *aceite* /aθeite/). In other words, a cognate effect in production is expected if the average over a cloud of memory traces in the exemplar space includes instances of Spanish-influenced exemplars (i.e., Spanish words or Spanish-accented Catalan words) instead of native-like Catalan exemplars, ultimately having an impact on the acoustic realization of this Catalan-specific vowel contrast. The average over a region in the exemplar space can also account for the gradient that has typically been observed in studies of cross-linguistic phonetic influence. By taking into account the distribution of vowels in the production study, exemplar models are also able to account for why the lexical decision results show the asymmetry in error rates between /ɔ/ words and /o/ words. The production data shows that for both groups of speakers (but especially for the Spanish-dominant bilinguals),

the production of /o/ in non-cognates is likely to overlap in acoustic space with the production of /ɔ/ in cognates. This pattern in the production data explains the asymmetry in the perception results: when /o/ words are mispronounced with /ɔ/, most of the errors are on non-cognates, because in general, the vowel space for non-cognate /o/ tends to overlap with the vowel space for /ɔ/. Conversely, when /ɔ/ words are mispronounced with /o/, most of the errors are on cognates because the vowel space for /ɔ/ in cognates is much closer to the vowel space for /o/. Exemplar models would assume that past experience with cognate and non-cognate words creates lexically-specific expectations for where these words might fall in the acoustic space, and the results from the lexical decision task reflect that.

CONCLUSION

The results of this study indicate that cognate status has an effect on both the phonetic production and processing of the Catalan back mid-vowel contrast by early Spanish-Catalan bilinguals. This cross-linguistic influence was robust for both language dominance groups when producing the experimental stimuli as well as when identifying aurally presented stimuli either as a word or a non-word. Interference at the lexical/phonetic interface has been accounted for in previous studies (Brown and Harper, 2009; Amengual, 2012; Mora and Nadeu, 2012; Brown and Amengual, 2015; Jacobs et al., 2016), but this acoustic interference must be operationalized in a theoretical model that accounts for the observed alterations in the lexical representations of bilingual individuals. This study argues that an exemplar model of lexical representation can be applied to bilingual data to explain cognate effects in which bilinguals do not separate “clouds of memory traces” in each language—they are in fact interconnected—and that the phonetic features of cognate lexical items form a stronger link than non-cognates, thus enhancing cross-language influence. The assumption that the bilingual individual has a single lexicon where lexical elements in different languages are stored together and interconnected has already been proposed in previous bilingual production models (de Bot, 1992). For instance, Hartsuiker et al. (2004) in a study of syntactic priming in bilingual individuals also adopt an integrated view of the bilingual lexicon and make the case for language-specific lexical-syntactic representations, which are then connected to lemma-level representations that are shared between both languages.

While the episodic account provided by exemplar theoretic approaches is reasonable, it is acknowledged that the interpretations provided necessitate further research and support. The extension of this model to include bilingual or multilingual data is intended to open a debate on how the lexical representations and the phonetic abilities of bilinguals interact and how the exemplar model can be extended to include bilingual lexical connections through which cognates facilitate phonetic interference. The study of the mental lexicon either as containing multiple episodes (Goldinger, 1997, 1998; Johnson, 1997a,b; Bybee, 2001; Pierrehumbert, 2001, 2003a,b; Hawkins, 2003) or abstract prototypes (Mehler, 1981; McClelland and

Elman, 1986; Pisoni and Luce, 1987; Norris, 1994; Gaskell and Marslen-Wilson, 1997), or a combination of both in a hybrid model holds considerable promise (McQueen et al., 2010). A challenge for future research is to specify which components of the mental lexicon are episodic and which are abstract.

AUTHOR CONTRIBUTIONS

The author (MA) states that he is solely responsible for the conception or design of the work, and the acquisition, analysis, interpretation of the data, and the drafting of the manuscript.

REFERENCES

- Adank, P., Smits, R., and van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *J. Acoust. Soci. Am.* 116, 3099–3107. doi: 10.1121/1.1795335
- Amengual, M. (2011). “Spanish and catalan in majorca: are there contact-induced changes in the majorcan catalan vowel system?,” in *Selected Proceedings of the 13th Hispanic Linguistics Symposium*, ed L. Ortiz (Somerville, MA: Cascadilla Proceedings Project), 214–223.
- Amengual, M. (2012). Interlingual influence in bilingual speech: cognate status effects in a continuum of bilingualism. *Bilingualism* 15, 517–530. doi: 10.1017/S1366728911000460
- Amengual, M. (2013). *An Experimental Approach to Phonetic Transfer in the Production and Perception of Early Spanish-Catalan Bilinguals*. Unpublished doctoral dissertation, University of Texas at Austin.
- Amengual, M. (2015). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Appl. Psycholinguistics*. doi: 10.1017/S0142716415000557. [Epub ahead of print].
- Amengual, M. (2016). The perception and production of language-specific mid-vowel contrasts: shifting the focus to the bilingual individual in early language input conditions. *Int. J. Bilingualism* 20, 133–152. doi: 10.1177/1367006914544988
- Amengual, M., and Chamorro, P. (2015). The effects of language dominance in the perception and production of the Galician mid-vowel contrasts. *Phonetica* 72, 207–236. doi: 10.1159/000439406
- Anderson, N. (1974). On the calculation of filter coefficients for maximum entropy spectral analysis. *Geophysics* 39, 69–72. doi: 10.1190/1.1440413
- Antoniou, M., Best, C. T., Tyler, M. D., and Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: asymmetries in phonetic code-switching. *J. Phon.* 39, 558–570. doi: 10.1016/j.wocn.2011.03.001
- Baese-Berk, M. M., and Goldrick, M. (2009). Mechanisms of interaction in speech production. *Lang. Cogn. Process.* 24, 527–554. doi: 10.1080/01690960802299378
- Barlow, J. A. (2014). Age of acquisition and allophony in spanish-english bilinguals. *Front. Psychol.* 5:288. doi: 10.3389/fpsyg.2014.00288
- Best, C. T. (1995). “A direct realist perspective on cross-language speech perception,” in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed W. Strange (Timonium, MD: York Press), 171–204.
- Best, C. T., and Tyler, M. (2007). “Nonnative and second-language speech perception: commonalities and complementarities,” in *Language Experience in Second-Language Speech Learning: The Role of Language Experience in Speech Perception and Production: A Festschrift in Honour of James Emil Flege*, eds O.-S. Bohn and M. J. Munro (Amsterdam: John Benjamins), 13–34.
- Birdsong, D., Gertken, L. M., and Amengual, M. (2012). *Bilingual Language Profile: An Easy-To-Use Instrument to Assess Bilingualism*. Austin, TX: COERLL; University of Texas at Austin.
- Boersma, P., and Weenink, D. (2015). *Praat: Doing Phonetics by Computer (5.4.02)*. Available online at: <http://www.praat.org>
- Bosch, L., Costa, A., and Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *Eur. J. Cognit. Psychol.* 12, 189–221. doi: 10.1080/09541446.2000.10590222
- Bosch, L., and Ramón-Casas, M. (2011). Variability in vowel production by bilingual speakers: can input properties hinder the early stabilization of contrastive categories? *J. Phon.* 39, 514–526. doi: 10.1016/j.wocn.2011.02.001
- Brown, E. L., and Amengual, M. (2015). Fine-grained and probabilistic cross-linguistic influence in the pronunciation of cognates: evidence from corpus-based spontaneous conversation and experimentally elicited data. *Stud. Hisp. Lusophone Linguist.* 8, 59–83. doi: 10.1515/shll-2015-0003
- Brown, E. L., and Harper, D. (2009). Phonological evidence of crosslinguistic exemplar connections. *Stud. Hisp. Lusophone Linguist.* 2, 257–274. doi: 10.1515/shll-2009-1052
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., and Böhl, A. (2011). The word frequency effect: a review of recent developments and implications for the choice of word frequency estimates in German. *Exp. Psychol.* 58, 412–424. doi: 10.1027/1618-3169/a000123
- Buz, E., and Jaeger, F. T. (2016). The (in)dependence of articulation and lexical planning during isolated word production. *Lang. Cognit. Neurosci.* 31, 404–424. doi: 10.1080/23273798.2015.1105984
- Bybee, J. (2001). *Phonology and Language Use*. Cambridge, UK: Cambridge University Press.
- Caramazza, A., and Brones, I. (1979). Lexical access in bilinguals. *Bull. Psychon. Soc.* 13, 212–214. doi: 10.3758/BF03335062
- Carrera-Sabaté, J., and Fernández-Planas, A. M. (2005). *Vocals Mitjanes Tòniques del Català. Estudi contrastiu interdialectal*. Barcelona: Horsori.
- Carroll, S. (1992). On cognates. *Second Lang. Res.* 8, 93–119. doi: 10.1177/026765839200800201
- Cochrane, R. (1980). The acquisition of /r/ and /l/ by Japanese children and adults learning english as a second language. *J. Multilingual Multicult. Dev.* 1, 331–360. doi: 10.1080/01434632.1980.9994031
- Coleman, J. (2002). “Phonetic representations in the mental lexicon,” in *Phonetics, Phonology and Cognition*, eds J. Durand and B. Laks (Oxford: Oxford University Press). 96–130.
- Costa, A., Caramazza, A., and Sebastián-Gallés, N. (2000). The cognate facilitation effect: implications for models of lexical access. *J. Exp. Psychol.* 38A, 367–393. doi: 10.1037/0278-7393.26.5.1283
- Costa, A., Santesteban, M., and Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain Lang.* 94, 94–103. doi: 10.1016/j.bandl.2004.12.002
- Cristoffanini, P., Kirsner, K., and Milech, D. (1986). Bilingual lexical representation: the status of Spanish-english cognates. *Q. J. Exp. Psychol.* 38, 367–393. doi: 10.1080/14640748608401604
- Cutler, A., Weber, A., and Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *J. Phon.* 34, 269–284. doi: 10.1016/j.wocn.2005.06.002

FUNDING

This work was supported by National Science Foundation DDIG # 1226964.

ACKNOWLEDGMENTS

I would like to take the opportunity to thank Miquel Simonet, Barbara E. Bullock, Almeida Jacqueline Toribio, and David Birdsong for their feedback on this project. I would also like to thank Annie Tremblay, Stephanie Lain, Eva Bosch Roura, and the reviewers for comments and help with several aspects of this manuscript. I am solely responsible for any remaining errors.

- Darcy, I., Daidone, D., and Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *Ment. Lexicon* 8, 372–420. doi: 10.1075/ml.8.3.06dar
- Darcy, I., Dekydtspotter, L., Sprouse, R., Glover, J., Kaden, C., McGuire, M., et al. (2012). Direct mapping of acoustics to phonology: on the lexical encoding of front rounded vowels in L1 english-L2 French acquisition. *Second Lang. Res.* 28, 5–40. doi: 10.1177/0267658311423455
- Darcy, I., and Krüger, F. (2012). Vowel perception and production in Turkish children acquiring L2 German. *J. Phon.* 40, 568–581. doi: 10.1016/j.wocn.2012.05.001
- de Bot, K. (1992). A bilingual production model: levelt's speaking model adapted. *Appl. Linguist.* 13, 1–24. doi: 10.1093/applin/13.1.1
- de Bot, K., Cox, A., Ralston, S., Schaufeli, A., and Weltens, B. (1995). Lexical processing in bilinguals. *Second Lang. Res.* 11, 1–19. doi: 10.1177/026765839501100101
- de Groot, A. M. B. (1992a). "Bilingual lexical representations: a closer look at conceptual representations," in *Orthography, Phonology, Morphology, and Meaning*, eds R. Frost and L. Katz (Amsterdam: Elsevier), 389–412.
- de Groot, A. M. B. (1992b). Determinants of word translation. *J. Exp. Psychol. Learn. Mem. Cognit.* 18, 1001–1018. doi: 10.1037/0278-7393.18.5.1001
- de Groot, A. M. B. (1995). Determinants of bilingual lexicosemantic organisation. *Comput. Assist. Lang. Learn.* 8, 151–181. doi: 10.1080/0958822940080204
- de Groot, A. M. B., Borgwaldt, S., Bos, M., and van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *J. Mem. Lang.* 47, 91–124. doi: 10.1006/jmla.2001.2840
- Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychol. Rev.* 93, 282–321. doi: 10.1037/0033-295X.93.3.283
- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Lang. Cogn. Process.* 5, 313–349. doi: 10.1080/01690969008407066
- Dijkstra, T., Grainger, J., and van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: the neglected role of phonology. *J. Mem. Lang.* 41, 496–518. doi: 10.1006/jmla.1999.2654
- Dijkstra, T., van Jaarsveld, H., and ten Brinke, S. (1998). Interlingual homograph recognition: effects of task demands and language intermixing. *Bilingualism* 1, 51–66. doi: 10.1017/S1366728998000121
- Escudero, P., Hayes-Harb, R., and Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *J. Phon.* 36, 345–360. doi: 10.1016/j.wocn.2007.11.002
- Ettlinger, M., and Johnson, K. (2009). Vowel discrimination by English, French and Turkish speakers: an exemplar-based approach to speech perception. *Phonetica* 66, 222–242. doi: 10.1159/000298584
- Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *J. Acoust. Soc. Am.* 89, 395–441. doi: 10.1121/1.400473
- Flege, J. E. (1995). "Second language speech learning: theory, findings and problems in speech perception and linguistic experience," in *Issues in Cross-Language Research*, ed W. Strange (Timonium, MD: York Press), 171–204.
- Flege, J. E. (2007). "Language contact in bilingualism: phonetic system interactions," in *Laboratory Phonology Vol. 9*, eds J. Cole and J. Hualde (Berlin: Mouton de Gruyter), 353–382.
- Flege, J. E., Bohn, O. S., and Jang, S. (1997). Effects of experience on non-native speakers' production and perception of english vowels. *J. Phon.* 25, 437–470. doi: 10.1006/jpho.1997.0052
- Flege, J. E., and MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Stud. Lang. Acquis.* 26, 1–34. doi: 10.1017/S0272263104261010
- Flege, J. E., MacKay, I. R. A., and Meador, D. (1999). Native Italian speakers' perception and production of english vowels. *J. Acoust. Soc. Am.* 106, 2973–2987. doi: 10.1121/1.428116
- Flege, J. E., Munro, M. J., and MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *J. Acoust. Soc. Am.* 97, 3125–3134. doi: 10.1121/1.413041
- Flege, J., and Munro, M. J. (1994). The word unit in L2 speech production and perception. *Stud. Lang. Acquis.* 16, 381–411. doi: 10.1017/S0272263100013437
- Gaskell, M. G., and Marslen-Wilson, W. D. (1997). Integrating form and meaning: a distributed model of speech perception. *Lang. Cogn. Process.* 12, 613–656. doi: 10.1080/016909697386646
- Gertken, L. M., Amengual, M., and Birdsong, D. (2014). "Assessing language dominance with the Bilingual Language Profile," in *Measuring L2 Proficiency: Perspectives from SLA*, eds P. Leclercq, A. Edmonds, and H. Hilton (Bristol, UK: Multilingual Matters), 208–225.
- Goldinger, S. D. (1997). "Words and voices: perception and production in an episodic lexicon," in *Talker Variability in Speech Processing*, eds K. Johnson and J. Mullennix (San Diego, CA: Academic Press), 33–66.
- Goldinger, S. D. (1998). Echoes of echoes?: an episodic theory of lexical access. *Psychol. Rev.* 105, 251–279. doi: 10.1037/0033-295X.105.2.251
- Goldrick, M., and Blumenstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: evidence from tongue twisters. *Lang. Cogn. Process.* 21, 649–683. doi: 10.1080/01690960500181332
- Goldrick, M., Runnqvist, E., and Costa, A. (2014). Language switching makes pronunciation less nativelike. *Psychol. Sci.* 25, 1031–1036. doi: 10.1177/0956797613520014
- Grosjean, F. (1985). The bilingual as a competent but specific speaker-hearer. *J. Multilingual Multicult. Dev.* 6, 467–477. doi: 10.1080/01434632.1985.9994221
- Grosjean, F. (1997). "Processing mixed language: issues, findings, and models," in *Tutorials in Bilingualism: Psycholinguistic Perspectives*, eds A. De Groot and J. Kroll (New York, NY: Erlbaum), 225–254.
- Grosjean, F. (1998). Studying bilinguals: methodological and conceptual issues. *Bilingualism* 1, 131–149. doi: 10.1017/S136672899800025X
- Grosjean, F. (2001). "The bilingual's language modes," in *One Mind, Two Languages: Bilingual Language Processing*, ed J. L. Nicol (Malden, MA: Blackwell), 1–22.
- Grosjean, F. (2008). *Studying Bilinguals*. Oxford: Oxford University Press.
- Guion, S. (2003). The vowel systems of Quichua-Spanish bilinguals: an investigation into age of acquisition effects on the mutual influence of the first and second languages. *Phonetica* 60, 98–128. doi: 10.1159/000071449
- Hall-Lew, L. (2010). "Improved representation of variance in measures of vowel merger," in *Proceedings of Meetings on Acoustics*, Vol. 127 (Baltimore, MD), 060002–060012.
- Hartsuiker, R. J., Pickering, M. J., and Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychol. Sci.* 15, 409–414. doi: 10.1111/j.0956-7976.2004.00693.x
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *J. Phon.* 31, 373–405. doi: 10.1016/j.wocn.2003.09.006
- Hayes-Harb, R., and Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Lang. Res.* 24, 5–33. doi: 10.1177/0267658307082980
- Hay, J., Warren, P., and Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *J. Phon.* 34, 458–484. doi: 10.1016/j.wocn.2005.10.001
- Herrick, D. (2003). *An Acoustic Analysis of Phonological Vowel Reduction in Six Varieties of Catalan*. Unpublished doctoral dissertation, UC Santa Cruz.
- Herrick, D. (2006). "Mid vowels and Schwa in Eastern Catalan: five non-Barcelona Dialects," in *New Perspectives on Romance Linguistics*, ed J.-P. Y. Montreuil (Amsterdam: John Benjamins), 113–126. doi: 10.1075/cilt.276.09her
- Herrick, D. (2007). Eastern Catalan vowel reduction is characterized by raising – not centralization. *Stud. Lang. Sci.* 6, 227–241.
- Herrick, D. (2008). An acoustic description of Central Catalan vowels based on real and nonsense word data. *Catalan Rev.* 21, 231–256.
- Hoshino, N., and Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition* 106, 501–511. doi: 10.1016/j.cognition.2007.02.001
- Jacobs, A., Fricke, M., and Kroll, J. (2016). Cross-language activation begins during speech planning and extends into second language speech. *Lang. Learn.* 66, 324–353. doi: 10.1111/lang.12148
- Jarvis, S., and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. New York, NY: Routledge.
- Johnson, K. (1997a). "Speech perception without speaker normalization: an exemplar model," in *Talker Variability in Speech Processing*, eds K. Johnson and J. Mullennix (San Diego, CA: Academic Press), 14–166.
- Johnson, K. (1997b). "The auditory/perceptual basis for speech segmentation," in *Ohio State University Working Papers in Linguistics*, Vol. 50 (Columbus, OH), 101–113.
- Johnson, K. (2003). *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell.
- Johnson, K. (2007). "Decisions and mechanisms in Exemplar-based phonology," in *Experimental Approaches to Phonology. In Honor of John Ohala*, eds

- M. J. Solé, P. Beddor, and M. Ohala (Oxford: Oxford University Press), 25–40.
- Lemhöfer, K., Dijkstra, A., and Michel, M. (2004). Three languages, one ECHO: cognate effects in trilingual word recognition. *Lang. Cogn. Process.* 19, 585–611. doi: 10.1080/01690960444000007
- Levelt, W. J. M. (1989). *Speaking: from Intention to Articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–75. doi: 10.1017/S0140525X99001776
- Lewis, M. P. (2009). *Ethnologue: Languages of the World, 16th Edn.*, Dallas, TX: SIL International. Available online at: <http://www.ethnologue.com/>
- McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0
- McQueen, J. M., Cutler, A., and Norris, D. (2010). Phonological abstraction in the mental lexicon. *Cogn. Sci.* 30, 1113–1126. doi: 10.1207/s15516709cog0000_79
- Mehler, J. (1981). The role of syllables in speech processing: infant and adult data. *Philos. Trans. R. Soc.* 295, 333–352. doi: 10.1098/rstb.1981.0144
- Mora, J. C., and Nadeu, M. (2012). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *Int. J. Bilingualism* 16, 484–499. doi: 10.1177/1367006911429518
- Nooteboom, S. G. (2005). “Listening to oneself: Monitoring speech production,” in *Phonological Encoding and Monitoring in Normal and Pathological Speech*, eds R. J. Harsuiker, R. Bastiaanse, A. Postma, and W. Wijnen (Hove: Psychology Press), 167–186.
- Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52, 189–234. doi: 10.1016/0010-0277(94)90043-4
- Oldfield, R. C., and Wingfield, A. (1965). Response latencies in naming objects. *Q. J. Exp. Psychol.* 4, 272–281. doi: 10.1080/17470216508416445
- Oppenheim, G. M., and Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition* 106, 528–537. doi: 10.1016/j.cognition.2007.02.006
- Pallier, C., Bosch, L., and Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition* 64, B9–B17. doi: 10.1016/S0010-0277(97)00030-9
- Peramunage, D., Blumstein, S., Myers, E., Goldrick, M., and Baese-Berk, M. M. (2011). Phonological neighborhood effects in spoken word production: an fMRI study. *J. Cogn. Neurosci.* 23, 593–603. doi: 10.1162/jocn.2010.21489
- Pierrehumbert, J. B. (2001). “Exemplar dynamics: word frequency, lenition, and contrast,” in *Frequency and the Emergence of Linguistic Structure*, eds J. Bybee and P. Hopper (Philadelphia: John Benjamins), 137–157.
- Pierrehumbert, J. B. (2003a). Phonetic diversity, statistical learning, and acquisition of phonology. *Lang. Speech* 46, 115–154. doi: 10.1177/00238309030460020501
- Pierrehumbert, J. B. (2003b). “Probabilistic phonology: discrimination and robustness,” in *Probabilistic Linguistics*, eds R. Bod, J. Hay, and S. Jannedy (Cambridge: MIT Press), 177–228.
- Pisoni, D., and Luce, P. (1987). Acoustic-phonetic representations in word recognition. *Cognition* 25, 21–25. doi: 10.1016/0010-0277(87)90003-5
- Rafel i Fontanals, J. (1998). *Diccionari de freqüències*. Barcelona: Institut d’Estudis Catalans.
- Ramón-Casas, M., Swingle, D., and Sebastián-Gallés, N. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cogn. Psychol.* 59, 96–121. doi: 10.1016/j.cogpsych.2009.02.002
- Recasens, D. (1991). *Fonètica Descriptiva del Català*. Barcelona: Institut d’Estudis Catalans.
- Recasens, D., and Espinosa, A. (2006). Dispersion and variability in Catalan vowels. *Speech Commun.* 48, 645–666. doi: 10.1016/j.specom.2005.09.011
- Recasens, D., and Espinosa, A. (2009). Dispersion and variability in Catalan five and six peripheral vowel systems. *Speech Commun.* 51, 240–258. doi: 10.1016/j.specom.2008.09.002
- Sánchez-Casas, R., Davis, C. W., and García-Albea, J. E. (1992). Bilingual lexical processing: exploring the cognate/noncognate distinction. *Eur. J. Cogn. Psychol.* 4, 293–310. doi: 10.1080/09541449208406189
- Scarborough, R. (2012). Lexical similarity and speech production: neighborhoods for nonwords. *Lingua* 112, 164–176. doi: 10.1016/j.lingua.2011.06.006
- Sebastián-Gallés, N., and Baus, C. (2005). “On the relation between perception and production in L2 categories,” in *Twenty-first Century Psycholinguistics: Four Cornerstones*, ed A. Cutler (New York, NY: Erlbaum), 279–292.
- Sebastián-Gallés, N., Echeverría, S., and Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *J. Mem. Lang.* 52, 240–255. doi: 10.1016/j.jml.2004.11.001
- Sebastián-Gallés, N., and Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition* 72, 111–123. doi: 10.1016/S0010-0277(99)00024-4
- Simonet, M. (2011). Production of a Catalan-specific vowel contrast by early Spanish-Catalan bilinguals. *Phonetica* 68, 88–110. doi: 10.1159/000328847
- Simonet, M. (2014). Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *J. Phon.* 43, 26–37. doi: 10.1016/j.wocn.2014.01.004
- Simonet, M. (2015). An acoustic study of coarticulatory resistance in “dark” and “light” alveolar laterals. *J. Phon.* 52, 138–151. doi: 10.1016/j.wocn.2015.07.006
- Sloos, M. (2013). The reversal of the Bären-Beeren merger in Austrian Standard German. *Ment. Lexicon* 8, 353–371. doi: 10.1075/ml.8.3.05slo
- Soares, C., and Grosjean, F. (1984). Bilinguals in a monolingual and a bilingual speech mode: the effect on lexical access. *Mem. Cogn.* 12, 380–386. doi: 10.3758/BF03198298
- Trautmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *J. Acous. Soc. Am.* 88, 97–100. doi: 10.1121/1.399849
- Vigliocco, G., and Harsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychol. Bull.* 128, 442–472. doi: 10.1037/0033-2909.128.3.442
- Weber, A., and Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *J. Mem. Lang.* 50, 1–25. doi: 10.1016/S0749-596X(03)00105-0
- Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands. *J. Acous. Soc. Am.* 33, 248. doi: 10.1121/1.1908630

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Amengual. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Fuzzy Nonnative Phonolexical Representations Lead to Fuzzy Form-to-Meaning Mappings

Svetlana V. Cook^{1*}, Nick B. Pandža^{2,3}, Alia K. Lancaster^{2,3} and Kira Gor²

¹ National Foreign Language Center, University of Maryland, College Park, College Park, MD, USA, ² Graduate Program in Second Language Acquisition, School of Languages, Literatures, and Cultures, University of Maryland, College Park, College Park, MD, USA, ³ Center for Advanced Study of Language, University of Maryland, College Park, College Park, MD, USA

OPEN ACCESS

Edited by:

Isabelle Darcy,
Indiana University Bloomington, USA

Reviewed by:

Jianfeng Yang,
Shaanxi Normal University, China
Rebecca Foote,
University of Illinois at
Urbana-Champaign, USA

*Correspondence:

Svetlana V. Cook
svcook@umd.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 02 February 2016

Accepted: 23 August 2016

Published: 21 September 2016

Citation:

Cook SV, Pandža NB, Lancaster AK
and Gor K (2016) Fuzzy Nonnative
Phonolexical Representations Lead to
Fuzzy Form-to-Meaning Mappings.
Front. Psychol. 7:1345.
doi: 10.3389/fpsyg.2016.01345

The present paper explores nonnative (L2) phonological encoding of lexical entries and dissociates the difficulties associated with L2 phonological and phonolexical encoding by focusing on similarly sounding L2 words that are not differentiated by difficult phonological contrasts. We test two main claims of the fuzzy lexicon hypothesis: (1) L2 fuzzy phonolexical representations are not fully specified and lack details at both phonological and phonolexical levels of representation (Experiment 1); and (2) fuzzy phonolexical representations can lead to establishing incorrect form-to-meaning mappings (Experiment 2). The Russian-English Translation Judgment Task (Experiment 1, TJT) explores how the degree of phonolexical similarity between a word and its lexical competitor affects lexical access of Russian words. Words with smaller phonolexical distance (e.g., *parent-parrot*) show longer reaction times and lower accuracy compared to words with a larger phonolexical distance (e.g., *parent-parchment*) in lower-proficiency nonnative speakers, and, to a lesser degree, higher-proficiency speakers. This points to a lack of detail in nonnative phonolexical representations necessary for efficient lexical access. The Russian Pseudo-Semantic Priming task (Experiment 2, PSP) addresses the vulnerability of form-to-meaning mappings as a consequence of fuzzy phonolexical representations in L2. We primed the target with a word semantically related to its phonological competitor, or a potentially confusable word. The findings of Experiment 2 extend the results of Experiment 1 that, unlike native speakers, nonnative speakers do not properly encode phonolexical information. As a result, they are prone to access an incorrect lexical representation of a competitor word, as indicated by a slowdown in the judgments to confusable words. The study provides evidence that fuzzy phonolexical representations result in unfaithful form-to-meaning mappings, which lead to retrieval of incorrect semantic content. The results of the study are in line with existing research in support of less detailed L2 phonolexical representations, and extend the findings to show that the fuzziness of phonolexical representations can arise even when confusable words are not differentiated by difficult phonological contrasts.

Keywords: lexical access, phonological representations, form-to-meaning mapping, nonnative auditory perception, Russian

INTRODUCTION

Current research suggests that second language (L2) learners experience persistent difficulties in auditory perception of non-native speech (for a review, see Gor, 2015). Comprehension of speech by nonnative speakers is typically characterized by a higher propensity for errors and communication breakdowns than in native speakers, and is believed to be more cognitively demanding. The challenges in comprehension are traditionally associated with a difficulty in identifying phonemes that do not exist in the native language (Polivanov, 1932; Sheldon and Strange, 1982; Best, 1994, 1995; Flege et al., 1994, 1996; Flege, 1995; Kuhl and Iverson, 1995; Strange, 1995; Ingram and Park, 1998; Best et al., 2001). A common example is the inability of Japanese learners of English to distinguish between the English /r/ and /l/ phonemes, which are both conflated into a single Japanese phoneme /r/ (Goto, 1971; McClelland et al., 1999; for a recent review, see Cutler, 2015).

Indeed, a reduced ability for phonological categorization of nonnative sounds coupled with unfaithful nonnative *phonological* representations can lead to a breakdown in lexical access (Pallier et al., 1997, 2001; Cutler and Otake, 2004; Weber and Cutler, 2004; Cutler et al., 2006; Broersma, 2012; Diaz et al., 2012). Crucially, a nonnative deficit at the level of phonological representation is only part of the difficulty. Word recognition hinges upon a successful match between the auditory signal and the existing phonological representation of the stored word (Pisoni and Luce, 1987). Therefore, chances of a match are contingent, on the one hand, upon the listener's ability to decode the auditory signal, and, on the other, upon the quality of the *phonolexical* representation, or the phonological representation of the word as a whole (Luce et al., 2000; Chrabaszcz and Gor, 2014). Late L2 learners typically experience deficits in both aspects.

There are two implicit assumptions with respect to the existing relationship between phonology and the lexicon that have recently been subject to scrutiny (see Gor, 2015 for a review). The first assumption is that the acquisition of accurate phonological representations precedes accurate acquisition of lexical knowledge. That is, without establishing distinct phonological representations (or representations of phonemes) it is impossible to establish accurate representations of words containing those phonemes (or phonolexical representations). This is a bottom-up view of lexical acquisition. The second assumption, which is an extension of the first, is that once the nonnative phonological representations are acquired, they can easily transfer into phonolexical representations and contribute to lexical knowledge. This view, and the first assumption in particular, has been challenged by some empirical evidence demonstrating that L2 learners are capable of distinguishing between minimal pairs of words that differ by one phoneme while maintaining unreliable performance on phonological discrimination of those same contrasting phonemes (Weber and Cutler, 2004; Cutler et al., 2006; Escudero et al., 2008; Hayes-Harb and Masuda, 2008). Conversely, as shown by Darcy et al. (2013) in a study of the acquisition of Japanese geminates and German front-rounded and back-rounded vowels

by L1 speakers of English, even if the contrast is acquired at a phonological or phonetic level, this knowledge does not necessarily transfer to phonolexical representations. In this study, nonnative participants performed less accurately on nonwords than on words containing the same minimal pair distinction in a lexical decision task. Poor performance on nonwords suggests that L2 speakers were reluctant to reject the nonwords because they were unsure about their phonological composition. This result is in contrast to the results of the phonetic discrimination task, where accuracy was, for the most part, at ceiling. Furthermore, Darcy et al. (2013) extend the original findings of Weber and Cutler (2004) and Cutler et al. (2006), who proposed that the lack of accurate phonological perception does not mandate the lack of a distinction at a phonolexical level; on the contrary, the distinction between two forms can be maintained in the absence of the phonological contrast. It should be noted that if this is the case, the phonological contrast at the phonolexical level is still not target like, and the unfamiliar category tends to be interpreted not as a distinct category in its own right, but rather as a poor exemplar of the familiar (or native) category (see also Escudero et al., 2008; Hayes-Harb and Masuda, 2008). The fact that the distinction between the two phonolexical forms is evident at a lexical, but not at a phonetic level indicates that while there is a phonetic divergence of the two entries from one homophonous form during lexical access, there is no need to postulate a prerequisite of stable phonological representations. More generally, these findings present additional evidence against the “phonology-first,” or bottom-up, approach to acquisition of L2 phonology and suggest an alternative possibility, such that phonological representations evolve together with lexical knowledge, and do not necessarily precede it (Davidson et al., 2007; Dufour et al., 2010; Reinisch et al., 2013).

To complicate the matter of auditory speech perception further, several other factors not directly connected to phonological encoding can interfere with accurate speech perception by L2 speakers, namely those related to lexical knowledge and how this knowledge is represented in the mental lexicon. Some existing research suggests that the organization of the L2 lexicon is qualitatively different from the L1 lexicon in several respects. For example, compared to native speakers, phonological links among words tend to play a much more prominent role in organizing the L2 mental lexicon than semantic links do (Stolz and Tiffany, 1972; Meara, 1978, 1983, 1984; Wolter, 2001; Fitzpatrick, 2006). The main conclusion stemming from these studies is that semantic links between words in the learner's mental lexicon are “fairly tenuous ones, easily overridden by phonological similarities, in a way that is very uncharacteristic of native speakers” (Meara, 1983, p. 31). It is conceivable that L2 learners rely on phonological similarity to make sense of an unknown word. For example, if learners of English hear an unfamiliar word *coffin* without any context to help them figure out the meaning of the word, they may decide that the word is related to the word *cough* or *coffee*.

The representational deficit in lexical knowledge can be quite detrimental to nonnative speech comprehension, especially if it leads to the retrieval of an incorrect word. Because L2

lexical representations are unreliable, and persist into advanced levels of proficiency, L2 learners use additional strategies to resolve phonolexical ambiguities, such as the use of context (Rüschmeyer et al., 2008; Gor, 2014) and morphosyntactic cues (Conrad, 1983; Chrabaszcz and Gor, 2014). When additional cues are unavailable, lexical access can result in retrieval of a non-target word due to increased activation of phonological neighbors and spurious competition (Weber and Cutler, 2004; Broersma and Cutler, 2008, 2011; Broersma, 2012). Words that are less known to the learners are usually associated with greater phonolexical ambiguity and tend to cause error-prone translations in favor of a word phonologically related to the target (Cook and Gor, 2015). This suggests that an unintended word can be accessed as a result of an error in matching auditory input to an existing phonolexical representation, leading to an erroneous form-to-meaning mapping.

While a number of studies have found evidence of separation of pseudo-homophonous phonolexical forms in L2 into distinct, but not necessarily target like, representations (Cutler and Otake, 2004; Broersma and Cutler, 2008, 2011; Broersma, 2012; Darcy et al., 2012, 2013), there is little discussion of the consequences of incorrect lexical access of phonolexically ambiguous words, such as *lock* and *rock*. Indeed, the words used in these studies (except Darcy et al., 2012, 2013) had a clear point of disambiguation, and the ambiguity at the phonetic level created only temporary uncertainty, which was later resolved without affecting the outcome of lexical access. In these studies, the experimental items had little potential for being ambiguous for comprehension beyond the first syllable overlap. For example, for Dutch learners of English, *panda* is only confusable with *pencil* through the presentation of the first syllable, but once the second syllable is reached, the word can be uniquely identified as *panda*, and not *pencil* (Weber and Cutler, 2004).

There is only one study that we are aware of that looked at phonolexical ambiguity in L2 access of the lexical semantics, and form-to-meaning mappings of phonolexically ambiguous words, in particular. Ota et al. (2009) found similar semantic relatedness judgments effects for visually presented pairs of pure homophones (*rock-hard* and *beach-ocean*) and for pairs with pseudo-homophones for English learners of L1 Japanese (*lock-hard*) and L1 Arabic (*peach-ocean*). Despite the fact that the study assessed phonolexical ambiguity in words that differed by a phoneme that was perceptually difficult for a particular L1 population (/l/ and /r/ for Japanese, and /p/ and /b/ for Arabic), the study is particularly relevant for the current investigation, since it provides evidence for possible confusion of words' meanings stemming from the lack of detail in their form. The study by Ota and colleagues empirically tested the possibility of erroneous lexical access due to phonolexical ambiguity, and validated the claim that lexical entries for these types of words are resistant to complete separation and are potentially mutually confusable.

The present paper extends the findings of the previous body of research, and tests the *fuzzy lexicon* hypothesis (Cook, 2012; Cook and Gor, 2015). The fuzzy lexicon hypothesis claims that L2 learners operate with *fuzzy*, or low-resolution, phonolexical representations. A fuzzy representation of a word

is a mental representation of phonolexical form that does not represent the word as a fixed phonological sequence. Such a representation may leave some phonemes underspecified (e.g., either a final /d/ or /t/) or contain some uncertainty (and ensuing optionality) regarding the exact phonemes and their sequence. Similar to child L1 acquisition, when L2 learners first acquire a word, they represent the word with a purpose to differentiate it from other, similar-sounding words in their lexicon. As in child L1 acquisition, at the initial stages, the representations are approximate, but as the lexicon expands, these fuzzy representations need to be revised. (Note, however, that in both child L1 and adult L2 acquisition, underdifferentiated words may or may not include underdifferentiated phonemes. Such phoneme underdifferentiation would constitute a purely phonological problem, although it will impact word recognition as well.) Crucially for the construct of fuzzy phonolexical representations, lexical underdifferentiation may take place even if there are no phonological problems associated with the word *per se*. As a result, two words may become confusable if they overlap in their form, and their representations are not robust. For example, at the initial stage of acquisition, the word *parent* can be represented as [ˈpɛrə(n)t] with [n] being optional. This type of representation is unstable, because the phonological details of the representation are not fully spelled out. A fuzzy representation has a certain degree of non-targetlike flexibility making it possible to accommodate the input that has contrasting features (both [ˈpɛrənt] and [ˈpɛrət]). The learner can successfully operate with this fuzzy representation for the word *parent*, because it is sufficiently detailed to differentiate this word from other words in the mental lexicon. As soon as *parrot* starts to appear in the input, most learners unknowingly continue to map both words *parent* and *parrot* to the same fuzzy representation [ˈpɛrə(n)t] due to fuzziness of the phonolexical representation, which is a match to both words in the input, and to uncertainty about form-to-meaning mappings. With greater vocabulary knowledge and differentiation, at some point learners start to realize that the words *parent* and *parrot* are different both semantically and phonologically. They are forced to revise the existing fuzzy phonolexical representation, and split it into two separate more detailed representations, even if they are still not entirely targetlike.

We make two main assumptions about how fuzzy phonolexical representations function during language use and whether they are of consequence for processing. First and foremost, fuzzy phonolexical representations are not fully specified and lack detail at both the phonological and semantic levels of representation. As a number of studies have successfully demonstrated, including the ones discussed earlier, the presence of an adequate phonological category does not necessarily result in the target like representations of the phoneme at the phonolexical level. To extend this finding further, the fuzzy lexicon hypothesis suggests that in many cases, phonological difficulty associated with the acquisition of L2 phonemes is only one of the factors contributing to the difficulties associated with L2 lexical access. The lack of fidelity in phonolexical representations may or may not have to do with the encoding of a difficult L2 phonological contrast. The main cause of non

target like performance is that L2 learners do not know the exact phonological composition of the word that they are trying to access. Fuzzy representations are in many cases episodic, prototype-based, or Gestalt-type representations that allow for mostly reliable access to the correct meaning. Fuzzy phonolexical representations that have a phonological form that is not sufficiently detailed and, consequently, not target like, still make it possible to access the correct meaning.

The second assumption made by the fuzzy lexicon hypothesis concerns incorrect form-to-meaning mappings. The fact that for some time the two representations were merged into a fuzzy representation leads to continued difficulties in correct mapping of the input to the corresponding phonolexical form, and [ˈperət] can still end up being comprehended as *parent*. Thus, one of the consequences of fuzzy representations is potential confusion in form-to-meaning mappings even when phonolexical fuzziness is partially resolved. This makes the representation unstable, such that the form-to-meaning mapping is still variable and inconsistent, in some cases still resulting in an erroneous match between the auditory input and the phonolexical representation.

The incorrect form-to-meaning mappings are, however, not permanent. As shown by Darcy et al. (2013), L2 proficiency plays an important role in the discrimination of potentially homophonous forms (as was the case with German participants in Experiments 3 and 4), while some phonological contrasts are more resistant to separation (in Experiments 1 and 2 with Japanese participants). The reported lack of separation of potentially homophonous forms (*rock* and *lock*) can result in erratic access to the two potential meanings (*lock* or *rock*). We address these possible scenarios in two auditory experiments with L1 English learners and native speakers of Russian.

The Present Study

This study aims to further explore how confusability at the phonolexical level affects nonnative lexical access. To our knowledge, the present study is among the first to extend the focus of the research beyond the lexical encoding of difficult phonological contrasts and to deal with phonolexical representations in terms of their global similarity in phonolexical form (as in *parent-parrot*, for example). While difficult phonological contrasts and global similarity in phonolexical form both influence lexical processing, we seek to assess the quality of phonolexical representations on their own right by eliminating the need to encode problematic L2 phonemes.

In Experiment 1, we explore the first assumption made by the fuzzy lexicon hypothesis, which claims that at early stages of L2 acquisition, the phonolexical form of a new word is stored without a detailed specification. As a result, it is confusable with similar sounding words. The present experiment makes an assumption that the likelihood that a word will be confused with another word is determined by the degree of phonological overlap between the two words. In a Translation Judgment Task (TJT) with aurally presented Russian words, we explore how varying degrees of similarity affect lexical access. We operationalize this similarity between competing phonolexical forms as phonological Levenshtein Distance (LD), with higher LD indicating less phonological overlap. Three

groups of participants completed the task: Advanced and Superior L2 learners of Russian, and a group of native Russian speaker controls. In critical trials, actual Russian word primes were replaced by a competitor Russian word with a similar phonolexical form. We predict that L2 learners will be less accurate than Native speakers in judging the translation of words that differ from the competitor in only one or two phonemes. Since the phonolexical form of L2 words is represented coarsely and without fine detail, during the matching procedure, fine differences between the auditory stimuli and the available representations stored in the mental lexicon will be overlooked. Alternatively, they may be discounted as allowable variation due to speaker or pronunciation differences. Further, the lower-proficiency Advanced L2 group will show less sensitivity to the differences between the target word and its competitor, because lower-proficiency L2 speakers represent words more “holistically” than the higher-proficiency Superior L2 speakers. As a consequence, for the lower-proficiency L2 group, the confusability effect for a mismatch in one phoneme will be similar to the effect for a mismatch in two phonemes. Unlike the Advanced group, the Superior group will have a greater sensitivity to the increase in the LD between the competitors, because they operate with a higher-definition phonolexical representation, which gives them more chances to detect the mismatch.

Experiment 2 addresses the second assumption made by the fuzzy lexicon hypothesis and assesses the vulnerability of form-to-meaning mappings as a consequence of fuzzy phonolexical representations in L2. In a modification of a semantic priming experiment (a Pseudo-Semantic Priming task, PSP) we primed the target with a word semantically related to its competitor, but not to the target itself. Learners heard the word *КОРОВА* /karova/ “cow” as the prime and then *МОЛОТОК* /malatok/ “hammer” as the target. We hypothesized that they would be biased to think that they had heard a word they knew and expected—*МОЛОКО* /malako/ “milk.” Indeed, the two words sound very much alike, and the L2 learner temporarily identifies the word *МОЛОТОК* /malatok/ “hammer” as the closest phonological match *МОЛОКО* /malako/ “milk,” which is semantically related to “cow.” Thus, the target word is confused with another one based on the similarity of their phonological forms. The predictions are that both a native and nonnative listener will expect to hear “milk” after they hear “cow.” At a point in time when they realize that they hear “hammer” instead of the onset-matched “milk,” a native speaker will quickly recover from the unmet expectation, while a nonnative speaker will be slower in recovering from the semantic “garden path” created by the prime and the target with a highly expected onset. If L2 learners show an increase in the processing time for a pseudo-target (“hammer”), this will be an indication that some confusion at the level of phonolexical representations has taken place. This scenario is only possible if neither of the words has a phonological representation that is detailed enough, or if there is an imbalance between them in terms of frequency and, thus, availability for efficient L2 lexical access. Sekine (2006) reports that lower-frequency L2 words have a tendency to be identified as similar-sounding higher-frequency words during auditory perception. While these results can be

also explained by the acquisitional sequence, where words of higher frequency are learned before words of lower frequency, the critical difference between the two confusable words remains—one has a more detailed representation than the other. The pattern of substitution will be to replace a lesser-known word with a better-known word. It is also possible that if the learner is not able to make a distinction between the forms of two similar-sounding words in the mental lexicon, both phonological forms will be loosely linked to the respective meanings, and can be swapped in lexical access. This assumption aligns with the fuzzy lexicon hypothesis: under certain circumstances, fuzzy phonological representations can activate the lexical meaning of the competitor, and as a result the wrong lexical meaning could be accessed. This is exactly the effect that the pseudo-priming experiment is designed to produce. If our assumption is true, then learners will tend to confuse the pseudo-related target with the actual semantically related word, resulting in less accurate judgment of lexical acceptability and slower reaction time in making the judgment.

EXPERIMENT 1: TRANSLATION JUDGMENT TASK (TJT)

Method

Participants

Thirty-two native speakers of Russian (22 female) and 52 adult American learners of Russian (33 female) participated in Experiment 1. **Table 1** displays the language background and demographic information of the speaker groups. Native speakers of Russian on average spent 18.9 years in the classroom learning English ($SD = 4.1$) and began learning English at an average age of 8.3 ($SD = 3.0$). Their self-rated English proficiency was on average 8.13 ($SD = 1.8$) for grammar, 8.34 ($SD = 1.43$) for speaking, 8.97 ($SD = 1.12$) for listening, and 9.13 ($SD = 1.07$) for reading on a ten-point Likert scale (0—“no proficiency” to 10—“native-like command”).

All nonnative participants prior to participation were pre-tested with a standard test of oral proficiency, a formal Oral Proficiency Interview (OPI), which assigned them a proficiency level in Russian on the Interagency Language Roundtable (ILR) scale widely used in the USA for government testing. Based on the OPI scores, L2 participants were subdivided into two proficiency

groups matched to the ILR levels 2 and 2+ ($n = 21$), and 3 and 3+ ($n = 31$), with higher scores indicating higher proficiency levels. Respectively, these group levels correspond to Advanced and Superior oral proficiency on the American Council on the Teaching of Foreign Languages (ACTFL) academic scale. All participants completed a language background questionnaire.

Materials

Russian words were selected from two frequency ranges—high (HF, ~130–500 instances per million) and low (LF, ~30–100 instances per million). The experimental set included words from different grammatical categories, but the majority belonged to the noun, verb, and adjective classes. The stimuli varied in phonological length (4–10 phonemes) and syllabic length (1–4 syllables). The experimental trials were counterbalanced across the two presentation lists. Since the same target appeared in matched and unmatched conditions on different presentation lists, lexical parameters of the items were naturally balanced across lists.

Each participant completed a total of 162 trials, each with an auditorily presented Russian word followed by a visually presented English word. In half of the trials, the Russian and English words matched (i.e., the Russian and English words were translations of one another), while in the other half of the trials the words mismatched (i.e., the Russian and English words were not translations of one another). For instance, one matched trial began with the auditory presentation of the Russian word *МОЛОТОК* /malatok/ “hammer,” followed by the visual presentation of the English word HAMMER. Presentation lists were balanced along the matching condition, such that if one target word appeared in a matched trial in List A, the same target appeared in a mismatched trial in List B.

In addition to the matching manipulation, the Russian words in the mismatch trials were manipulated using Levenshtein Distance (LD). LD is the measure used to calculate the degree of overlap between two phonological forms. It represents the “distance” between two word forms as measured by the number of replacements, additions, and deletions needed to generate one from the other (Levenshtein, 1966). For example, the Russian words *МОЛОТОК* /malatok/ “hammer,” and *МОЛОКО* /malako/ “milk” have an LD of two. By two changes, replacing the /t/ in /malatok/ with a /k/ and removing the final /k/, /malatok/ becomes /malako/. The psychological reality of LD and similar metrics has been demonstrated in various psycholinguistic tasks (e.g., Beijering et al., 2008; Yarkoni et al., 2008). In the mismatch trials, the Russian word was similar in form to the actual Russian translation of the English word presented. Thus, in these trials, the Russian word acted as a competitor to the actual translation. For instance, in one competitor mismatch trial /malako/ “milk” was followed by HAMMER. Since /malako/ “milk” and the actual Russian translation of “hammer,” /malatok/, have an LD of two, we expected participants to respond differently to these trials than to non-competitor mismatch trials (e.g., /zvezda/ “star”—BASEMENT). The experiment contained 54 non-competitor mismatch trials in which the Russian word heard and the actual Russian translation were not phonolexically similar. Within the competitor mismatch trials, the LD between

TABLE 1 | Experiment 1 (TJT) language background and demographic information by participant group.

Group	Advanced ($n = 21$)	Superior ($n = 31$)	Native ($n = 32$)
	Mean (SD)	Mean (SD)	Mean (SD)
Age	27.5 (4.1)	29.5 (8.0)	30.7 (9.0)
Age of acquisition	18.1 (2.8)	19.0 (3.1)	—
Classroom instruction	3.3 (2.1)	3.1 (2.0)	—
Immersion experience	1.6 (1.8)	2.8 (1.6)	—

All variables are measured in years.

the Russian word presented and the actual Russian translation ranged from 1 to 5.

Procedure

After completing a prescreening, which included the language history questionnaire, potential participants were invited to participate in the experiment. Participants completed the study remotely using DMDX testing software (Forster and Forster, 2003). Consent form and procedures were approved by the University of Maryland Institutional Review Board. The participants were instructed to take the test individually on a computer with headphones in a quiet room. The TJT was a part of a larger set of tasks not reported here. This test took ~20 min to complete, and all participants were paid upon completion of the study.

The materials for both experiments were digitally recorded by the same female native speaker of Russian in a sound-attenuated booth. Recordings were broadcast wave files (16 bit/48 kHz), made on a *Zoom H4n* digital audio recorder. The speaker read the items one by one in a clear citation style. Three or more recordings of each item were made, and the best-sounding token was chosen and included in the test stimuli. The sounds were digitally processed in Praat (Boersma and Weenink, 2013). Each individual token was extracted from the original recording at a zero-crossing boundary. Upon extraction, all stimuli were normalized for intensity.

A single trial consisted of a fixation cross, presented in the center of the screen for 250 ms, followed by a blank screen for 50 ms; then the auditory Russian prime was presented (the screen remained blank throughout the presentation of the audio file). At the offset of the prime the screen continued to remain blank for the duration of the inter-stimulus interval (ISI) for 1500 ms; a visually presented English target word immediately followed (centered on the screen, typeset Calibri, size 12, in bold, all upper-case letters). The target remained on the screen until the response was made or until the trial timed out (4000 ms from the onset of the visually-presented target). If no response was given when the timeout was reached, the next trial was advanced without a button press. Each trial was followed by a 1000 ms inter-trial interval (ITI). Participants were instructed to decide whether the two words were translation equivalents or not by pressing the appropriate button on the computer keyboard (right Control key for “YES” and left Control key for “NO”). Accuracy and reaction time (RT) from the onset of the visual target were digitally recorded. Participants completed nine practice trials before beginning the experimental trials. All stimuli were presented in 5 blocks (4 blocks with 35 and 1 block with 22 trials each), with opportunities for the participants to take self-paced breaks between the experimental blocks. Except for the practice trials, there was no feedback on accuracy provided to the participants.

Results

To model our data, we employed multilevel modeling (MLMs, or mixed-effects models) because of several advantages the method yields over traditional multiple regression or ANOVA methods: (1) by-subject and by-item analyses can be done simultaneously,

so as to generalize across people and items within a single analysis; (2) each individual trial is included in the analysis rather than averaging across multiple trials to obtain a single value for each participant; and (3) it properly models the multilevel structure of the data (e.g., trial-level variables such as word frequency vs. subject-level variables such as language proficiency) and is therefore not subject to the assumption of independence of observations as are multiple regression or ANOVA (Baayen et al., 2008; Linck and Cunnings, 2015).

The multilevel models we report here were conducted with the *lme4* package version 1.1-9 (Bates et al., 2015) in R version 3.2.0 (R Core Team, 2015) for logistic and linear multilevel modeling. Logistic MLMs for accuracy analyses were run using the “bobyqa” optimizer. In the RT analyses, correct responses were trimmed to exclude RTs lower than 300 ms because these reflect RTs that are too fast for normal processing, after which responses with long RTs were excluded if they exceeded a three standard deviation by-participant cutoff. Linear MLMs for RTs were reported using restricted maximum likelihood estimation, as full maximum likelihood underestimates the standard errors of the estimates. Due to the ongoing debate in calculating *p*-values for linear MLMs, only *t*-values are provided in *lme4* output, so $|t| > 1.65$ is considered marginal ($p < 0.10$), and $|t| > 2.00$ is considered significant at $p < 0.05$ (Gelman and Hill, 2007). All models were run as forced entry models for fixed effects and cross-classified subject and item random intercepts, and random slopes were tested one-by-one via likelihood ratio tests; only random slopes that significantly improved model fit and resulted in converging models were retained (Baayen, 2008; Baayen et al., 2008).

Accuracy

Due to the low number of stimuli at LD 5, those items were excluded from further analysis (1.5% of observations). Two more items were excluded from further analysis (1.5% of observations) due to technical issues. Accuracy results were then submitted to a logistic multilevel model (Table 2). The dependent variable was accuracy (0, 1); fixed effects included Condition (dummy-coded: Match, Competitor Mismatch, Non-Competitor Mismatch), Russian auditory prime match Frequency (log-transformed and *z*-scored), phonological LD between the competitor and the auditory match (LD 1–4; centered on LD 1, which is an LD of 1 phoneme), and Proficiency (dummy-coded as Advanced: ILR scale 2 and 2+, Superior: ILR 3 and 3+, and Native), as well as all two- and three-way interactions except those involving LD and Condition, as LD is only relevant to the Competitor Mismatch condition. Native speakers and the Competitor Mismatch condition were baseline; thus all significant effects in the model are interpreted with respect to this baseline (e.g., a significant effect for Advanced signifies the group is significantly different than the Native group). Note that a logistic MLM is not modeling mean accuracy but the probability of a correct or an incorrect response on an item given the predictors in the model.

The model intercept indicates that Competitor Mismatch trials are more likely than not to be correctly identified as incorrect translations by Native speakers, although Advanced ($b = -1.59$, $SE = 0.35$, $p < 0.001$) and Superior ($b = -1.73$, $SE = 0.33$, $p < 0.001$) are both about five times less likely to correctly

TABLE 2 | Experiment 1 (TJT) results of logistic multilevel modeling for Accuracy.

Fixed effects	<i>b</i>	exp(<i>b</i>)	<i>SE</i>	<i>p</i> value
Intercept (Native/Competitors/LD 1)	3.85	46.99	0.29	< 0.001*
Group:				
Advanced	−1.59	0.20	0.35	< 0.001*
Superior	−1.73	0.18	0.33	< 0.001*
Russian target freq (Native/LD 1)	−0.69	0.50	0.22	< 0.01*
Freq × Advanced	0.39	1.48	0.17	< 0.01*
Freq × Superior	0.29	1.34	0.17	0.08^
Levenshtein distance (Native)	0.38	1.46	0.24	0.11
LD × Advanced	−0.12	0.89	0.28	0.67
LD × Superior	0.39	1.48	0.28	0.17
LD × Freq (Native)	0.25	1.28	0.17	0.15
LD × Freq × Advanced	0.34	1.40	0.14	0.02*
LD × Freq × Superior	0.27	1.31	0.15	0.07^
Condition:				
Match (Native)	0.31	1.36	0.47	0.50
Match × Advanced	0.85	2.34	0.56	0.13
Match × Superior	2.09	8.08	0.54	< 0.001*
Match × Freq	1.57	4.81	0.29	< 0.001*
Non-competitor Mismatch (Native)	2.87	17.64	0.55	< 0.001*
NCM × Advanced	0.97	2.64	1.42	0.16
NCM × Superior	2.08	8.00	0.67	< 0.01*
NCM × Freq	1.17	3.22	0.32	< 0.001*
Random effects	Variance	SD	Correlation	
Intercepts Subject	0.18	0.43		
Intercepts Item	2.31	1.52		
Advanced Item	1.21	1.10	−0.32	
Superior Item	1.04	1.02	−0.25	0.97

*Significant at $p < 0.05$; ^Marginal at $p < 0.10$. Covariates are shaded in gray.

respond to LD 1 competitors (but note high accuracy overall in **Table 3**).

All three groups show an inverse effect of frequency on LD 1 trials, such that performance is worse as competitor frequency increases. Specifically, the Native group shows the strongest disadvantage to frequency at LD 1 ($b = -0.69$, $SE = 0.22$, $p < 0.01$), while the Superior group shows a trend for a weaker effect ($b = 0.29$, $SE = 0.17$, $p = 0.08$) and the Advanced group shows a significantly weaker effect ($b = 0.39$, $SE = 0.17$, $p < 0.01$). **Table 3** lists the average accuracy as LD increases by speaker group, with LD clearly affecting nonnative speakers but not native speakers. Covariate interactions indicate that for Match and Non-Competitor Mismatch conditions, there is a strong canonical frequency effect such that, as frequency of the Russian word they heard increases, participants are more likely to correctly respond to the English translation.

No group shows a significant effect of LD independent of frequency (all $ps > 0.10$), and the Native group does not show any effect of LD with increasing frequency. However, the Advanced group does show a positive effect of increasing LD with increasing frequency ($b = 0.34$, $SE = 0.14$, $p = 0.02$) such that, as the

frequency of the competitor word increases, participants are more accurate the less phonological overlap the Russian word has with the correct Russian translation. The Superior group also shows a similar positive trend ($b = 0.27$, $SE = 0.15$, $p = 0.07$). Taken together with the patterns observed in **Table 3** these effects suggest that the high frequency trials are driving the LD effect in L2 learners, and that accuracy increases with the increase in the LD between the incorrect Russian competitor and the correct Russian translation of the target.

Reaction Time

RT results for correct responses were trimmed as described above (eliminating 0.7% of observations) and submitted to a linear multilevel model (**Table 4**). All fixed effects, including interactions and baselines were identical to those in the logistic MLM for the accuracy data above. The random effects structure differed in that additional random slopes significantly improved the fit of this model, likely due to the large variability in RT whereas accuracy was largely high and near-ceiling for some conditions.

As **Figure 1** suggests (also see Supplementary Material for raw values), the Competitor Mismatch trials at LD 1 are more slowly responded to than match and Non-Competitor Mismatch trials to varying degrees for all groups.

As in the accuracy data, all three groups show an inverse effect of frequency on LD 1 trials, such that performance slows as competitor frequency increases; however, unlike the accuracy data, Native ($b = 0.03$, $SE = 0.01$, $t = 2.44$), Superior ($b = 0.002$, $SE = 0.01$, $t = 0.21$), and Advanced groups ($b = -0.003$, $SE = 0.01$, $t = -0.26$) all show an effect in RT as the effect for Natives was significant and the interaction terms for Superior and Advanced are not statistically different from that effect. Covariate interactions indicate (as they did for the accuracy data) that for Match and Non-Competitor Mismatch conditions, there is a strong canonical frequency effect such that, as frequency of the Russian word they heard increases, participants are faster to respond correctly to the English translation.

None of the groups shows a significant effect of LD independent of frequency; however, the Advanced group shows a marginal effect ($b = -0.03$, $SE = 0.02$, $t = -1.68$) for faster responses as LD increases (i.e., phonological overlap decreases). Unlike in the accuracy data, the Native group here does show a strong effect of LD with increasing frequency ($b = -0.03$, $SE = 0.01$, $t = -3.83$), as do the Superior ($b = -0.001$, $SE = 0.01$, $t = 0.00$) and the Advanced groups ($b = -0.002$, $SE = 0.01$, $t = -0.31$), whose performance is not significantly different from the Native group. Thus, as the frequency of the competitor word increases, participants respond more quickly the less phonological overlap the Russian word has with the correct Russian translation. Taken together with the patterns observed in **Figure 1**, these effects suggest that the high frequency trials are driving the LD effect for all groups (with the possible exception of the Advanced group), and that speed increases with the increase in the LD value, or as the phonological similarity between the words decreases.

In order to evaluate the strength of the confusability effect as a function of phonolexical distance independently of the

TABLE 3 | Mean accuracy to Russian match trials, non-competitor mismatch trials, and competitor trials of different Levenshtein distance split by frequency for both native and nonnative speakers in Experiment 1 (TJT).

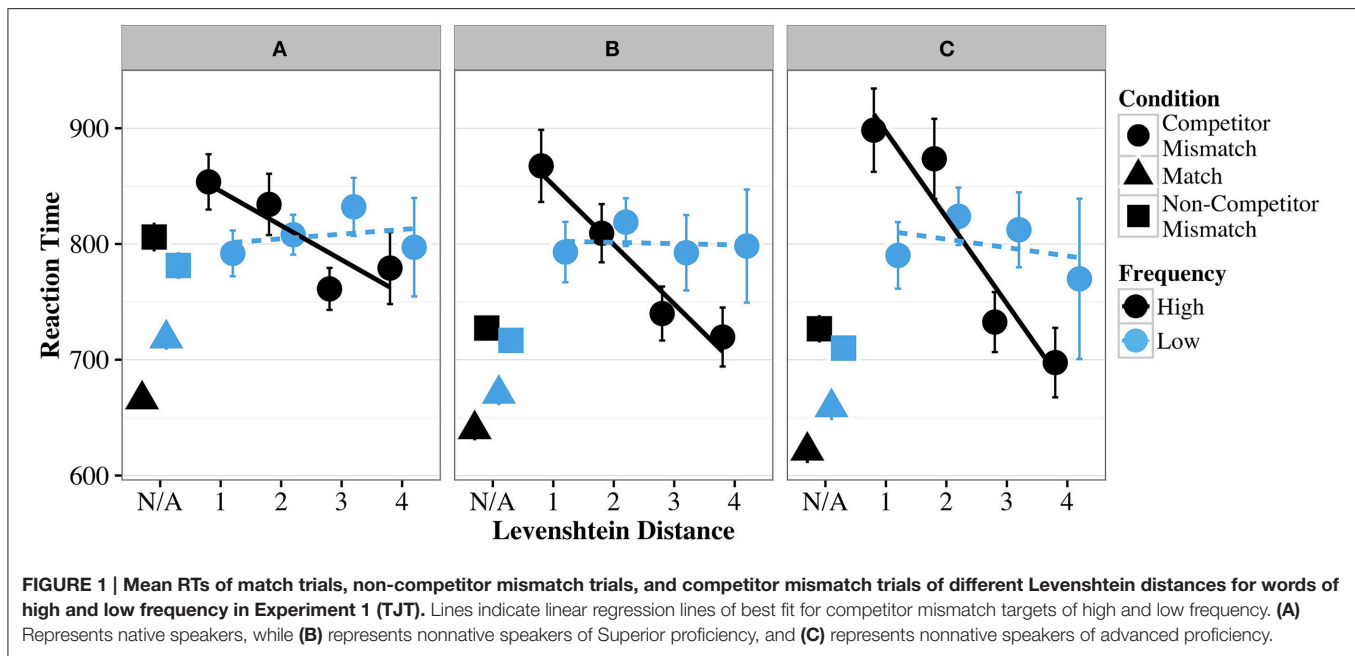
Group	Frequency	Match	Non-competitor mismatch	Competitor mismatch (levenshtein distance)			
				1	2	3	4
NATIVE							
	High	0.96 (0.19)	0.99 (0.09)	0.96 (0.20)	0.92 (0.20)	0.98 (0.14)	0.99 (0.11)
	Low	0.91 (0.29)	0.99 (0.09)	0.97 (0.16)	0.98 (0.15)	0.99 (0.12)	1.00 (0.00)
SUPERIOR							
	High	0.95 (0.21)	0.99 (0.09)	0.77 (0.42)	0.87 (0.34)	0.96 (0.19)	0.99 (0.10)
	Low	0.91 (0.28)	0.99 (0.09)	0.89 (0.31)	0.90 (0.30)	0.97 (0.17)	1.00 (0.00)
ADVANCED							
	High	0.93 (0.25)	0.99 (0.10)	0.83 (0.38)	0.93 (0.26)	0.91 (0.29)	1.00 (0.00)
	Low	0.86 (0.34)	0.98 (0.13)	0.87 (0.33)	0.88 (0.32)	0.91 (0.29)	0.78 (0.43)

Smaller Levenshtein distance indicates greater phonological similarity.

TABLE 4 | Experiment 1 (TJT) results of linear multilevel modeling for RT.

Fixed effects		<i>b</i>	<i>SE</i>	<i>t</i> value
Intercept (Native/Competitors/LD 1)		6.66	0.04	167.91*
Group:				
Advanced		0.04	0.06	0.71
Superior		−0.001	0.06	−0.01
Russian target frequency (Native/LD 1)		0.03	0.01	2.44*
Freq × Advanced		−0.003	0.01	−0.26
Freq × Superior		0.002	0.01	0.21
Levenshtein distance (Native)		0.002	0.01	0.16
LD × Advanced		−0.03	0.02	−1.68^
LD × Superior		−0.01	0.02	−0.96
LD × Freq (Native)		−0.03	0.01	−3.83*
LD × Freq × Advanced		−0.002	0.01	−0.31
LD × Freq × Superior		−0.00001	0.01	0.00
Condition:				
Match (Native)		−0.19	0.03	−6.60*
Match × Advanced		−0.16	0.05	−3.48*
Match × Superior		−0.10	0.04	−2.50*
Match × Freq		−0.10	0.02	−5.89*
Non-competitor mismatch (Native)		−0.07	0.03	−2.58*
NCM × Advanced		−0.16	0.04	−3.92*
NCM × Superior		−0.11	0.04	−3.12*
NCM × Freq		−0.07	0.02	−4.21*
Random effects		Variance	<i>SD</i>	Correlation
Intercept	Subject	0.04	0.21	
LD	Subject	< 0.001	0.02	−0.41
Frequency	Subject	< 0.001	0.01	−0.51 −0.42
Match	Subject	0.01	0.10	−0.49 0.51 0.35
NCM	Subject	< 0.01	0.06	−0.67 0.87 0.08 0.78
Intercept	Item	0.01	0.11	
Advanced	Item	0.01	0.09	−0.20
Superior	Item	0.01	0.08	−0.09 0.65
Residual		0.09	0.30	

*Significant at $p < 0.05$; ^Marginal at $p < 0.10$. Covariates are shaded in gray.



frequency manipulation, we fitted an additional model to the high frequency RT data alone. Only the Advanced group show an independent effect of phonological distance ($b = -0.05$, $SE = 0.02$, $t = -2.28$), evidenced by a greater slowdown in lexical access with confusable words that differed from the competitor in one or two phonemes (LD 1 and LD 2) compared to the words that differed in two or three phonemes (LD 3 and LD 4). In contrast, neither Native nor Superior group demonstrates a significant effect of phonological distance ($b = -0.02$, $SE = 0.01$, $t = -1.82$ and $b = -0.02$, $SE = 0.02$, $t = -1.17$ for the Native and Superior groups, respectively).

Discussion

The results of Experiment 1 revealed that the nonnative ability to access the target word reliably is associated with the degree of phonological similarity (more dissimilar words are easier to tell apart), and improves with higher proficiency. A reduced accuracy rate in the Advanced group in rejecting the incorrect translations with smaller phonological differences (LD 1 and LD 2) indicates that the Advanced participants are willing to accept auditory forms with a much greater phonological variability than the Superior or Native participants. Lower-proficiency L2 learners require a greater degree of difference between the words with potentially confusable phonolexical representations to differentiate among them and efficiently establish a correct match. As proficiency increases, this constraint is no longer at play and even small deviations from the target form are detected. Thus, we have found support that the effect in the Advanced group is driven by the reduced ability of the learner to match the auditory stimulus to the available phonolexical representations.

The RT data further confirm the picture. All groups have demonstrated an effect of phonological distance, which interacted with frequency, such that an increase in competitor frequency

resulted in a slowdown as the number of differentiating phonemes decreased. Words with competitors only minimally different from the intended target incurred the greatest processing costs.

In sum, the results of Experiment 1 suggest that unlike Native speakers, lower-proficiency learners do not properly encode the phonolexical information, and are thereby prone to access the incorrect lexical representation of a lexical competitor.

EXPERIMENT 2: PSEUDO-SEMANTIC PRIMING (PSP)

Method

Participants

Forty-seven adult American learners of Russian (9 female) and 20 adult native Russian controls (11 female) participated in Experiment 2. All nonnative participants were assigned to one of two proficiency levels: Intermediate or Advanced. Table 5 displays the language background and demographic information of each speaker group. Participants were recruited throughout the United States at universities with Russian Language Programs. As seen in Table 5, the Advanced learner group had spent more time in Russian-speaking countries than the Intermediate group [$t_{(43)} = 9.73$, $p < 0.001$]. However, the L2 groups are similar in duration of classroom instruction due to the structure of the program in which the Advanced students were enrolled [$t_{(31.1)} = 0.17$, $p = 0.566$]. That is, many Advanced learners were in a program which did not require extensive classroom instruction before immersion in a Russian-speaking country.

The determination of proficiency assignment was predominantly based on the results of a C-test (see Section Materials for details); however, other background information

TABLE 5 | Experiment 2 (PSP) language background and demographic information by participant group.

Group	Intermediate (<i>n</i> = 20)	Advanced (<i>n</i> = 27)	Native (<i>n</i> = 20)
	Mean (SD)	Mean (SD)	Mean (SD)
Age	24.3 (5.2)	23.9 (3.6)	23.0 (4.2)
Age of acquisition	21.0 (5.9)	20.2 (4.5)	–
Classroom instruction	2.8 (2.0)	2.4 (1.5)	–
Immersion experience	0.4 (0.4)	1.7 (0.5)	–

All variables are measured in years.

was also an important factor, such as length of study and length of immersion. In addition, the participants provided self-assessment data on their abilities in Speaking, Writing, Pronunciation, and their estimate of their L2 lexicon size. All these factors were taken into account for the group assignment. For example, if prospective participants had a significant immersion experience and high self-assessment scores, but had a borderline score on the C-test, they were assigned to the Advanced group.

Materials

In order to prescreen the participants in terms of their level of Russian language proficiency, a C-test was constructed based on the story “Modern day Mowgli,” which was adopted for testing purposes from a Russian language textbook (Niznik et al., 2009). A C-test is assumed to be a reliable measure of global language proficiency (Eckes and Grotjahn, 2006) and can also be successfully used in vocabulary research as a measure of vocabulary size (Singleton and Little, 1991; Singleton and Singleton, 1998; Singleton, 1999). According to the specification of the test, the first sentence of the text remained unchanged, and starting with the second sentence, every other word was partially deleted. The deletion was done according to the prescribed methodology: if the word has an even number of letters, the split is done in the middle, and the beginning half of the word is presented to the test-taker; if the word has an uneven number of letters, then the beginning half is preserved plus one additional letter, and this combination is presented to the test-taker. This process led to 40 partially deleted words. The scoring was done on a 3-point scoring scale for each testing item. Three points were assigned for a correct answer; two points were assigned for a correct vocabulary item, but in an incorrect form, resulting from an incorrect inflection (number, person, gender, tense, and mood errors); one point was assigned for a correct vocabulary item in a default form, i.e., uninflected; and zero points were assigned for an incomplete or incorrect vocabulary item. The ceiling accuracy score was 120 points (40×3 points per item). All of the Advanced participants scored above a 100 point mark on the C-test ($M = 107.14$, $SD = 4.11$), while the participants in the Intermediate group showed much greater variability, with scores distributed over a larger range ($M = 76.79$, $SD = 17.14$).

In Experiment 2 participants were required to listen to two Russian words and indicate if the second word (target) was a real Russian word. There were 320 trials in this experiment,

half of which (160) included real words and the other half included nonword targets. Nonwords were created from real Russian words by manipulating the first syllable; primes were always real words. Frequency was matched within prime-target pairs (high, low frequency); other parameters (e.g., number of syllables, length in phonemes) were balanced across conditions. Due to proficiency limitations, the Intermediate group was only exposed to a subset of the experimental materials, those in the high-frequency condition, therefore, the number of experimental trials for this group was reduced (160 overall instead of 320).

For the experiment we created 40 pairs that were related semantically, 40 pairs for the pseudo-semantic condition, 20 pairs for the unrelated (control) condition, and 100 distractor trials. Words in the unrelated trials were matched to the words in the critical conditions in frequency and length in phonemes. Real word prime-target pairs were created for the semantic priming condition (e.g., *КОРОВА* /karova/ “cow”—*МОЛОКО* /malako/ “milk”). Then a matching pseudo-semantic target was selected for each prime, appearing in the semantic condition (*КОРОВА* /karova/ “cow”—*МОЛОТОК* /malatok/ “hammer”). The words for the pseudo-semantic condition were selected based on their phonological similarity to the semantically related target and were always lower in frequency, but still within the targeted frequency band (high or low). Keeping in mind the frequency split (high, low), the materials were also constructed by using words that were moderately known to the participants. That is, known well enough for the lexical meaning to be accessed, but not well enough to accurately access the correct phonological representation in the mental lexicon. Pseudo-semantic pairs were pilot-tested on two Russian language learners prior to the study. The pilot-testers were not participants of the present study. Items that performed the best were retained for the use in the experiment.

The semantic and pseudo-semantic trials were balanced across two presentation lists, and during the experiment, each participant heard each prime only once, either in the semantic or in the pseudo-semantic condition. For instance, /malatok/ was heard if the priming pair is a pseudo-semantically related pair, or /malako/ if the target was a true semantically related word. Among the 20 native and 20 intermediate speakers, presentation list type (A, B) was evenly split. Due to the uneven number of advanced speakers, 14 heard list A while 13 heard list B.

Procedure

After completing a prescreening, which included the language history questionnaire and C-test, potential participants were invited to participate in the experiment as part of the Intermediate or the Advanced group. There were two ways that participants could complete the study—in person or remotely. The same testing software, DMDX (Forster and Forster, 2003), and delivery sequence was used in both methods. Consent form and procedures were approved by the University of Maryland Institutional Review Board. Each participant took the test individually on a computer with headphones in a quiet room. The results of the PSP presented here were part of a larger set of tasks not reported here. The experiment took ~30 min to complete. All participants were paid upon completion of the study.

A single trial consisted of a sequence of two aurally presented lexical items. Each trial started with a 300 ms pre-stimulus interval, then the audio prime was played in its entirety. The prime was followed by an ISI of 300 ms, after which the audio target was presented. Auditory stimuli were always played in their entirety, and subjects were given 4000 ms from the onset of target presentation to respond. Participants were instructed to decide whether the second word (the target) is a real Russian word or not by pressing the appropriate button on the computer keyboard (right Control key for “YES” and left Control key for “NO”). Each trial was followed by a 600 ms ITI. Accuracy and RT from the onset of the auditory target were digitally recorded. If no response was given after 4000 ms, the next trial was advanced without a button press. RT and accuracy were digitally recorded. Participants completed 10 practice trials before beginning the experimental trials. All experimental stimuli were presented in 8 blocks of 40 trials each with opportunities for participants to take self-paced breaks between the experimental blocks. Throughout the experiment, participants received feedback on the accuracy and speed of their responses to motivate optimal performance.

Results

The following logistic and linear multilevel models were conducted in the same manner as described in Experiment 1, with the exception of the cross-classified subject and item structure. The random effects structure for both models consisted of random intercepts by subject crossed by random intercepts by prime word nested within random intercepts by unique prime-target item pair, due to the nature of how the stimuli were constructed.

Accuracy

Accuracy results were submitted to a logistic multilevel model (Table 6). The dependent variable was accuracy (0, 1); fixed effects included Condition (dummy-coded: Control, Semantic Priming, and Pseudo Priming), Word Pair Frequency (HF, LF), and Group (dummy-coded: Intermediate, Advanced, and Native), and all two- and three-way interactions thereof. The model baseline was high frequency control trials for the Native group, and so all effects are to be interpreted with respect to this baseline. To help visualize the effects presented in the model, a simplified characterization of the data as cell means is presented in Table 7.

On HF Control trials, the Native and Advanced group (the latter not statistically different from the Native group) perform more accurately than the Intermediate group ($b = -1.23$, $SE = 0.35$, $p < 0.001$).

For HF semantic priming trials, the Native and Advanced groups are significantly more accurate compared to control trials, showing a strong semantic priming effect. However, the effect for the Intermediate group ($b = -4.45$, $SE = 1.07$, $p < 0.001$) is twice the size of the Native group effect, in the opposite direction, meaning that Intermediate participants are significantly less accurate on semantically primed words vs. control trials.

In the pseudo priming trials at HF, the Native group ($b = -1.17$, $SE = 0.45$, $p < 0.01$) is significantly less accurate compared

TABLE 6 | Experiment 2 (PSP) results of logistic multilevel modeling for Accuracy.

Fixed effects	<i>b</i>	exp(<i>b</i>)	<i>SE</i>	<i>p</i> value
Intercept (Native/Control/HF)	3.51	33.45	0.34	< 0.001*
Group:				
Intermediate	-1.23	0.29	0.35	< 0.001*
Advanced	0.35	1.42	0.38	0.51
Condition:				
Semantic Priming (Native/HF)	2.56	12.94	1.08	0.02*
Semantic × Intermediate	-4.45	0.01	1.07	< 0.001*
Semantic × Advanced	-1.85	0.16	1.14	0.10
Pseudo Priming (Native/HF)	-1.17	0.31	0.45	< 0.01*
Pseudo × Intermediate	-0.72	0.49	0.41	0.08^
Pseudo × Advanced	-0.80	0.45	0.44	0.07^
Low frequency (Native)	-0.02	0.98	0.46	0.96
LF × Advanced	-1.05	0.35	0.49	0.03*
LF × Semantic (Native)	-0.15	0.86	1.53	0.92
LF × Semantic × Advanced	-0.18	0.84	1.56	0.91
LF × Pseudo (Native)	-0.79	0.45	0.62	0.20
LF × Pseudo × Advanced	1.18	3.25	0.58	0.04*
Random effects	Variance	SD	Correlation	
Intercept Subject	0.31	0.56		
Low frequency Subject	0.40	0.64	-0.78	
Intercept Prime	0.56	0.75		
Intercept Prime/Item Pair	0.56	0.75		

*Significant at $p < 0.05$; ^Marginal at $p < 0.10$.

TABLE 7 | Mean accuracy to Russian pseudo-semantic priming trials, semantic trials, and control trials split by frequency for both native and nonnative speakers in Experiment 2 (PSP).

Group	Frequency	Pseudo-Semantic Priming	Semantic Priming	Unrelated Control
NATIVE				
	High	0.87 (0.34)	1.00 (0.07)	0.95 (0.21)
	Low	0.78 (0.42)	1.00 (0.07)	0.95 (0.22)
ADVANCED				
	High	0.80 (0.40)	0.98 (0.15)	0.96 (0.19)
	Low	0.71 (0.45)	0.93 (0.26)	0.90 (0.30)
INTERMEDIATE				
	High	0.56 (0.50)	0.57 (0.50)	0.86 (0.34)

The Intermediate group was not exposed to low frequency trials.

to control trials, and the Intermediate ($b = -0.72$, $SE = 0.41$, $p = 0.08$) and Advanced ($b = -0.80$, $SE = 0.44$, $p = 0.07$) groups are marginally showing an even stronger pseudo priming effect suggesting they are even less likely to answer those trials correctly.

For LF Control trials, Natives perform just as well as on HF trials. The Advanced group shows a frequency effect ($b = -1.05$, $SE = 0.49$, $p = 0.03$) in that they perform less well on LF control trials compared to the Native group. The Intermediate group was not exposed to LF trials due to

proficiency limitations. The size of the semantic priming effect for LF trials is just as strong for Natives and Advanced as it is for HF trials, as the interaction terms indicate the semantic effect for LF trials is not significantly different than the effect for HF trials.

For pseudo priming LF trials, the effect for the Native group is the same as for HF trials, and the three-way positive interaction for the Advanced group ($b = 1.18$, $SE = 0.58$, $p = 0.04$) essentially negates the frequency effect. Put another way, the Advanced group does not show a pseudo priming effect for LF trials, but still performs marginally worse on pseudo prime trials compared to NSs as they did on HF trials. Interestingly, the Advanced and Native groups, despite obvious descriptive trends toward a frequency effect on pseudo priming trials, statistically show no frequency effect.

Reaction Time

RT results for correct responses were trimmed as described previously (eliminating 1.8% of observations) and submitted to a linear multilevel model (Table 8). All fixed effects, including interactions and baselines were identical to those in the logistic MLM for the PSP accuracy data above. The random effects structure for the linear model differed again compared to the logistic MLM, again likely due to larger variability in RT compared to accuracy. To help visualize the effects presented in the model, a simplified characterization of the data as cell means is presented in Figure 2 (also see Supplemental Material for raw values).

On HF Control trials, the Native and Advanced group (the latter not statistically different from the Native group) make correct judgments more quickly than the Intermediate group ($b = 0.18$, $SE = 0.04$, $t = 4.29$).

For HF semantic priming trials, the Native ($b = -0.12$, $SE = 0.02$, $t = -5.33$) and Advanced groups ($b = 0.06$, $SE = 0.02$, $t = 2.38$) show a significant semantic priming effect in that they are faster on semantic trials compared to controls; however, note the significant effect for the Advanced group indicates that the semantic priming effect for the Advanced group is half as strong as for the Native group. Finally, for the Intermediate group, the semantic priming effect is no longer observed ($b = 0.14$, $SE = 0.03$, $t = 4.28$), meaning the group shows no speedup or slowdown on semantically primed trials.

In the pseudo priming trials for HF, the Native group and the Intermediate group show no effect of pseudo priming compared to control trials. However, the Advanced group ($b = 0.09$, $SE = 0.03$, $t = 3.30$) responds significantly more slowly on pseudo priming trials.

For low frequency (LF) Control trials, Natives perform just as well as on HF trials. The Advanced group shows a frequency effect ($b = 0.05$, $SE = 0.02$, $t = 2.40$) in that they perform more slowly on LF control trials compared to the Native group. The Intermediate group was not exposed to LF trials. Compared to HF trials, the size of the semantic priming effect for LF trials shows a marginally larger effect for Natives ($b = -0.06$, $SE = 0.03$, $t = -1.82$) and Advanced ($b = -0.02$, $SE = 0.03$, $t = -0.70$), since the three-way interaction of LF \times Semantic \times Advanced

TABLE 8 | Experiment 2 (PSP) results of linear multilevel modeling for RT.

Fixed effects	<i>b</i>	<i>SE</i>	<i>t</i> value
Intercept (Native/Control/HF)	6.79	0.03	228.64
Group:			
Intermediate	0.18	0.04	4.29*
Advanced	0.03	0.04	0.93
Condition:			
Semantic Priming (Native)	-0.12	0.02	-5.33*
Semantic \times Intermediate	0.14	0.03	4.28*
Semantic \times Advanced	0.06	0.02	2.38*
Pseudo Priming (Native)	0.01	0.02	0.57
Pseudo \times Intermediate	0.04	0.03	1.25
Pseudo \times Advanced	0.09	0.03	3.30*
Low frequency (Native)	0.004	0.02	0.20
LF \times Advanced	0.05	0.02	2.40*
LF \times Semantic (Native)	-0.06	0.03	-1.82 [^]
LF \times Semantic \times Advanced	-0.02	0.03	-0.70
LF \times Pseudo (Native)	-0.04	0.03	-1.32
LF \times Pseudo \times Advanced	-0.06	0.04	-1.68 [^]
Random effects	Variance	<i>SD</i>	Correlation
Intercept Subject	0.01	0.12	
Semantic Subject	< 0.001	0.02	0.29
Pseudo Subject	< 0.01	0.03	0.93 -0.04
Intercept Prime	< 0.01	0.05	
Intermediate Prime	< 0.01	0.05	-0.33
Advanced Prime	< 0.001	0.02	-0.14 -0.88
Intercept Prime/Item Pair	< 0.01	0.05	
Intermediate Prime/Item Pair	< 0.01	0.06	-0.59
Advanced Prime/Item Pair	< 0.01	0.05	-0.40 0.97
Residual	0.02	0.15	

*Significant at $p < 0.05$; [^]Marginal at $p < 0.10$.

was not significantly different from the marginal effect for Natives.

For pseudo priming LF trials, the Native group still shows no effect as for HF trials ($b = -0.04$, $SE = 0.03$, $t = -1.32$). The three-way marginal interaction for the Advanced group ($b = -0.06$, $SE = 0.04$, $t = -1.68$) indicates that (on top of the estimate and t -value for Natives) the Advanced group has no pseudo priming effect for LF trials (releveling the model with Advanced as baseline does exhibit this as a significant effect). Put another way, the Advanced group appears to only respond more slowly to HF pseudo priming trials and treats LF pseudo priming trials no differently than LF control trials.

Discussion

Results of Experiment 2 primarily indicate that both Advanced learners and Native speakers were more likely to make an erroneous judgment on the target when it was primed by a word prime semantically-related to the competitor, such that when they heard КОРОВА /karova/ "cow" they were more likely to judge МОЛОТОК /malatok/ "hammer" as a nonword compared to a similar sounding semantically related target МОЛОКО /malako/ "milk." Consistent with our predictions, the Advanced learners

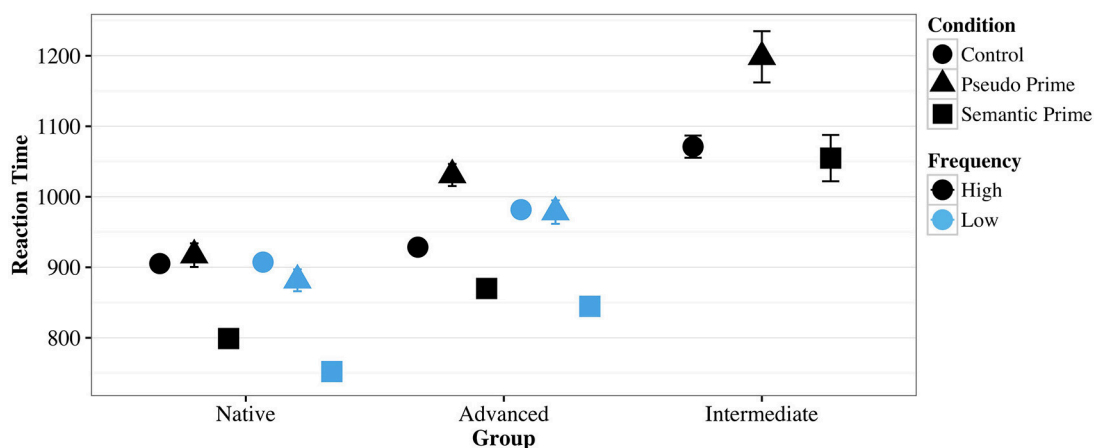


FIGURE 2 | Mean RTs of pseudo-semantic priming trials, semantic trials, and unrelated control trials split by language group for words of high and low frequency in Experiment 2.

show a processing delay with pseudo-semantic targets, albeit only in the high-frequency condition (we will come back to this point in the Section General Discussion). Unlike Advanced L2 learners, Native participants show no evidence of processing delays in either of the frequency conditions: they are as efficient in accessing a pseudo-semantic target as they are in accessing an unmatched control target. With the evidence of a robust semantic priming effect, we can conclude that no semantic priming effects guided their performance on pseudo-semantic targets.

The Intermediate group also shows no pseudo-semantic priming effect, but most likely for a different reason. We see here a similar trend as in the Advanced participants, but the slowdown is not supported statistically. The variability in the responses of the Intermediate participants is an indication that their lexical representations are unstable and are probably not yet sufficiently integrated into their mental lexicon. The conclusion is also supported by the lack of a semantic priming effect in the Intermediate group, which suggests that semantic associations among words in the developing L2 lexicon are not yet sufficiently entrenched to produce a native-like semantic priming effect (for an entrenchment account, see Gollan et al., 2008; Diependaele et al., 2013; Cook and Gor, submitted).

Overall, the experiment has succeeded in demonstrating that even Advanced learners operate with fuzzy phonolexical representations, which do not ensure reliable access to the intended meaning of the input word. In the pseudo-semantic priming manipulation, the Advanced learners were biased toward a semantic associate of the prime. While Native speakers showed no evidence of engaging in the processing of a similar-sounding, but semantically unrelated competitor, the Advanced group did.

GENERAL DISCUSSION

The results of the study are in line with the existing research in support of unfaithful L2 phonolexical representations (Pallier et al., 2001; Weber and Cutler, 2004; Darcy et al., 2012, 2013).

Crucially, the present study extends the research agenda to demonstrate that the fuzziness of phonolexical representations can arise even when confusable words do not contain difficult phonological contrasts.

There is a current understanding that the acquisition of phonological categories and lexical representations, while being closely interrelated, still shows some autonomy in nonnative learners. This relative autonomy may lead to asymmetries between L2 efficiency in phonological encoding and lexical (or phonolexical) encoding (Weber and Cutler, 2004; Darcy et al., 2013). The present paper takes the next step in the direction of exploring the nature of L2 lexical encoding. It attempts to dissociate the L2 phonological and phonolexical encoding difficulties. It does so by focusing on similarly sounding L2 words that are not differentiated by difficult phonological contrasts (e.g., the hard-soft consonant contrast in Russian, as in Chrabaszc and Gor, 2014).

In Experiment 1—the TJT—we looked at how Levenshtein Distance, which operationalizes the degree of phonolexical similarity between the words that are potential lexical competitors, affects native and nonnative lexical access. Native speakers do not show statistically significant sensitivity to phonolexical similarity between the target word and its implied competitor. This suggests that they have access to fully-specified, detailed phonolexical representations, which allow them to reliably reject words that are not a complete match to the stored representation and which is independent of the degree of phonological overlap. At the same time, as predicted, only the lower-proficiency speakers (Advanced group) show the effect of phonolexical similarity, which interacts with lexical frequency, and is much weakened in the higher-proficiency speakers (Superior group).

To challenge the processing delay interpretation that we are proposing in Experiment 1, one can hypothesize that the effect of phonolexical distance is due to the speedup in the words with greater phonolexical distance between the competitors instead. In following with this argument, the effect is driven

not by the slowdown in the lexical access of words with a smaller phonological distance, but rather by faster access to the words with greater phonological distance. One can reject this interpretation based on the inspection of the results visually represented in **Figure 1**. When we compare the RTs in the LD and Unmatched conditions in the Advanced and Superior groups, it is clear that LD 1 and LD 2 items incur additional processing costs compared to the unmatched items, while LD 3 and LD 4 items do not. Therefore, low-similarity items in the Confusable condition were treated in the same way as the unrelated ones (this conclusion is also supported by the statistical analyses). In terms of the speed of access, the results for the Advanced and Superior groups in the control conditions, or for words without phonological overlap, are not significantly different from each other. These results provide us with reasonable grounds to claim that longer latencies in the processing of words with smaller phonological differences from the competitors are due to poorer quality of phonolexical representations, which entails a less efficient matching mechanism, and causes processing delays during lexical access in lower-proficiency L2 learners.

The performance of the Native participants in Experiment 1 also warrants additional discussion. It is typically assumed that native speakers of the language are more efficient and more rapid in performing lexical access than nonnative speakers (e.g., Gollan and Kroll, 2001; Michael and Gollan, 2005). The results of the present experiment do not challenge this observation, despite the fact that the RTs in the Native group are slower than in the two learner groups across all conditions. The experiment was designed as such that only the Russian primes can lead to a native processing advantage. However, the reaction times are measured on the responses to the English targets, which are the English translations of the Russian primes. It is quite reasonable to expect overall processing delays in the performance of the Native Russian group in processing of the English stimuli. The study explores how the relative difference in the processing speed of words with competitors varying in the degree of phonolexical similarity manifests itself in each individual group; therefore, the slowness in processing English stimuli of the native speakers does not interfere with the findings.

The result of Experiment 2—the Pseudo-Semantic Priming task—extends the finding of Experiment 1. Unlike native speakers, nonnative speakers do not properly encode phonolexical information, and, as indicated by the slowdown in the judgments on the confusable words, are thereby prone to access the incorrect lexical representation of a lexical competitor. Experiment 2 also succeeded in demonstrating that learners are unable to reliably access the word that they have heard because its phonolexical representation is not detailed enough, and are attempting to access the confusable word semantically related to the target instead.

This study provides evidence that nonnative ability to differentiate two lexical entries is not only affected by a perceptual inability to reliably identify L2 phonemes (as shown in other studies, e.g., Weber and Cutler, 2004; Cutler et al., 2006; Escudero et al., 2008; Hayes-Harb and Masuda, 2008), but is also associated with how well the word is known, or its degree of entrenchment in the mental lexicon (Diependaele et al., 2013; Veivo and

Järvikivi, 2013; Cook and Gor, submitted). This conclusion is supported by the role of L2 learners' proficiency and lexical frequency during lexical access. The results of Experiment 1 demonstrate that at some point in the development of the L2 lexicon, learners operate with fuzzy phonolexical representations that lack detailed phonological encoding. With increasing proficiency, phonolexical representations acquire greater detail and become less fuzzy. This progression from fuzzy to fully detailed phonolexical representations is evidenced in how the sensitivity to the degree of mismatch between the auditory input and the existing representations affects lexical access in two nonnative groups with different proficiency levels and the native group. As the results of the experiment demonstrate, there is no delay in lexical access of the words that are different from their competitor only in one or two phonemes in native speakers. The higher-proficiency Superior group shows a tendency to some delay that does not reach statistical significance. Conversely, fuzzy representations preclude an effective match between the auditory input and the stored phonolexical representations, and thereby cause a significant slowdown in processing observed in the lower-proficiency nonnative group.

The effect of lexical frequency is observed in the results of Experiment 2, where only in the high-frequency condition did the Advanced learners show a processing delay in accessing the pseudo-semantic target. At Advanced proficiency, phonolexical representations of high-frequency words are sufficiently detailed, and the mismatch between the stimulus and the representation is readily detected; however, the ability to efficiently discount the competitor in favor of the correct target is not yet nativelike. The lack of a pseudo-semantic effect in the low-frequency condition suggests that the phonolexical form of these words in L2 does not have sufficient phonolexical detail to detect the mismatch, and thereby trigger a slowdown. Our findings are in full agreement with the entrenchment proposal based on a computer simulation that showed how lower levels of subjective familiarity lead to increased activation of such lexical entries and to a reduced ability to inhibit other activated candidates (Diependaele et al., 2010; see also Cook and Gor, submitted). These modeling results parallel recent empirical findings. Veivo and Järvikivi (2013) explored the role of L1 orthography in L2 lexical access of Finnish-French bilinguals and found that when a word did not have a stable L2 representation (as evidenced by subjective familiarity with the word), priming by an interlingual orthographic homophone resulted in a processing benefit attributable to prelexical facilitation. A similar conclusion in relation to phonolexical, rather than orthographic, representations was reached by Broersma (2012), who hypothesized that the lack of inhibition effect from error-induced homophones in a priming lexical decision for L2 learners (e.g., *flesh–flash*) should be taken as evidence for the reduced ability to mediate competition between the coactivated words.

Finally, our study provides evidence that fuzzy phonolexical representations result in unfaithful form-to-meaning mappings that lead to retrieval of incorrect semantic content. Experiment 2 has succeeded in establishing the effect of fuzzy form-meaning mappings on the activation of semantic networks during priming with a prime that was phonologically similar, but semantically

unrelated to the target. The observed inhibition was due to spurious activation of a semantically plausible phonological neighbor. The results primarily suggest that the meaning of the competing words is not only activated, but considered as a possible meaning of the target. The involvement of the semantic level provides evidence in support of occurrence of erroneous form-to-meaning mappings in a developing L2 lexicon. As suggested by the *weaker links* hypothesis (Gollan et al., 2008) bilinguals split their language experience between two (or more) languages, and are, therefore, disadvantaged compared to monolingual speakers in terms of establishing lexical representations. Indeed, reduced exposure to nonnative words does not sufficiently strengthen the links between semantics and phonolexical representations. While our study is not designed to support or falsify the weaker links hypothesis, our results are compatible with it.

The present study provides further empirical evidence in support of the *fuzzy lexicon* hypothesis (Cook, 2012; Cook and Gor, 2015), which suggests that speed and accuracy of lexical access is mediated by the degree of detail in L2 phonolexical representations and by the strength of form-to-meaning mappings. In Experiment 2 we show that Advanced learners are sensitive to the semantic as opposed to pseudo-semantic priming manipulation. While they are able to detect the difference between the intended and the actual target, their difficulty lies in the ability to overcome the initial bias toward the semantic target. The processing slowdown indicates that the separation of phonolexical representations in the semantic and pseudo-semantic targets is not fully resolved, and lexical access of the correct meaning incurs additional processing costs. At the same time, the two representations are to a certain degree distinct from each other—a result also reported by Darcy et al. (2013). Had they been completely merged together, we would have observed a facilitation effect associated with semantic priming as an indication that the mismatching phonolexical form had activated the competitor's semantic meaning. This is not the mechanism that we observed. Instead, pseudo-semantic priming creates a semantic garden path that sets up a strong prediction, which is further confirmed by the target onset. This garden path effect is initially the same for both native and nonnative groups. While native participants quickly recover from this competition, with no additional processing costs observed, nonnative participants are not as efficient. On the one hand, they do not have sufficiently detailed representations to be certain about the match to the word they hear. On the other hand, they need to break the semantic connection from the prime to the pseudo-semantic target, to which they were guided by the prime, and further, by the initial phonological overlap of the actual target with the virtual semantic target. As both phonolexical and semantic representations are weak and generate uncertainty, the

step of rejecting one lexical entry and reaccessing a different word (the pseudo-semantic target) incurs significant processing costs. The competition of form contributes to the processing costs, but it is mediated by a semantic link, and in this sense, the ambiguity resolution takes place at both levels—phonolexical and semantic.

Overall, the study takes the next step in identifying the locus of nonnative difficulties in lexical access and investigates challenges in phonological and lexical representations that go beyond discriminating difficult nonnative contrasts. It provides further evidence for the *fuzzy lexicon* hypothesis and empirically demonstrates that speed and accuracy of lexical access is mediated by the degree of detail in L2 phonolexical representations, which, in turn, is constrained by subjective familiarity with lexical items and L2 proficiency.

AUTHOR CONTRIBUTIONS

SC and KG developed the theoretical framework, and designed the study. SC developed the materials and was in charge of data acquisition and writing the draft of the manuscript. NP contributed to the analysis and interpretation of the data, and drafting the Results Section. AL participated in data collection and drafting the Methods Sections of the paper. KG contributed to writing the Introduction, Present Study, and the Discussion Sections. SC, KG, NP, and AL reviewed the data, and contributed to certain aspects of the discussion. All authors are responsible for final approval of the manuscript, and endorse all aspects of its design, and the interpretation of the results.

ACKNOWLEDGMENTS

The authors would like to thank Cathy Doughty and Scott Jackson for their help in the development and administration of the study. The study was supported in part by the School of Languages, Literatures and Cultures, the Center for Advanced Study of Language, and the College of Arts and Humanities at the University of Maryland. Some of the material reported in the paper is based upon work supported, in whole or in part, with funding from the United States Government. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the University of Maryland, College Park and/or any agency or entity of the United States Government. This material is being made available for personal or academic research use.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01345>

REFERENCES

- Baayen, H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge, UK: Cambridge University Press.
- Baayen, H., Davidson, D., and Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). *lme4: Linear Mixed-Effects Models using Eigen and S4*. R Package Version 1.1-9. Available online at: <http://CRAN.Rproject.org/package=lme4>
- Beijering, K., Gooskens, C., and Heeringa, W. (2008). "Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm," in *Linguistics in the Netherlands*, eds M. van Koppen and B. Botma (Amsterdam: John Benjamins), 13–24.

- Best, C. T. (1994). "The emergence of native-language influence in infants: a perceptual assimilation model," in *The Transition from Speech Sounds to Spoken Words: The Development of Speech Perception*, eds H. Nusbaum, J. Goodman, and C. Howard (Cambridge, MA: MIT Press), 167–224.
- Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross Language Research*, ed W. Strange (Timonium, MD: York Press), 171–203.
- Best, C. T., McRoberts, G. W., and Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *J. Acoust. Soc. Am.* 109, 775–794. doi: 10.1121/1.1332378
- Boersma, P., and Weenink, D. (2013). *Praat: Doing Phonetics by Computer (version 5.3.56)[computer program]*. Amsterdam.
- Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Lang. Cogn. Process.* 27, 1205–1224. doi: 10.1080/01690965.2012.660170
- Broersma, M., and Cutler, A. (2008). Phantom word activation in L2. *System* 36, 22–34. doi: 10.1016/j.system.2007.11.003
- Broersma, M., and Cutler, A. (2011). Competition dynamics of second-language listening. *Q. J. Exp. Psychol.* 64, 74–95. doi: 10.1080/17470218.2010.499174
- Chrabaszcz, A., and Gor, K. (2014). Context effects in the processing of phonolexical ambiguity in L2. *Lang. Learn.* 64, 415–455. doi: 10.1111/lang.12063
- Conrad, L. R. (1983). *Listening Comprehension Strategies in Native and Second Language. Unpublished Doctoral dissertation*, Michigan State University.
- Cook, S. (2012). *Phonological Form in L2 Lexical Access: Friend or Foe? Unpublished Doctoral dissertation*, University of Maryland.
- Cook, S. V., and Gor, K. (2015). Lexical access in L2: representational deficit or processing constraint? *Ment. Lex.* 10, 247–270. doi: 10.1075/ml.10.2.04coo
- Cutler, A. (2015). Representation of second language phonology. *Appl. Psycholinguist.* 36, 115–128. doi: 10.1017/S0142716414000459
- Cutler, A., and Otake, T. (2004). Pseudo-homophony in non-native listening. *J. Acoust. Soc. Am.* 115, 2392. doi: 10.1121/1.4780547
- Cutler, A., Weber, A., and Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *J. Phon.* 34, 269–284. doi: 10.1016/j.wocn.2005.06.002
- Darcy, I., Daidone, D., and Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *Ment. Lex.* 8, 372–420. doi: 10.1075/bct.80.06dar
- Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., et al. (2012). Direct mapping of acoustics to phonology: on the lexical encoding of front rounded vowels in L1 English-L2 French acquisition. *Second Lang. Res.* 28, 5–40. doi: 10.1177/0267658311423455
- Davidson, L., Shaw, J., and Adams, T. (2007). The effect of word learning on the perception of non-native consonant sequences. *J. Acoust. Soc. Am.* 122, 3697–3709. doi: 10.1121/1.2801548
- Diaz, B., Mitterer, H., Broersma, M., and Sebastian-Galles, N. (2012). Individual differences in late bilinguals' L2 phonological processes: from acoustic-phonetic analysis to lexical access. *Learn. Individ. Differ.* 22, 680–689. doi: 10.1016/j.lindif.2012.05.005
- Diependaele, K., Lemhöfer, K., and Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: a lexical entrenchment account. *Q. J. Exp. Psychol.* 66, 843–863. doi: 10.1080/17470218.2012.720994
- Diependaele, K., Ziegler, J. C., and Grainger, J. (2010). Fast phonology and the bimodal interactive activation model. *Eur. J. Cogn. Psychol.* 22, 764–778. doi: 10.1080/09541440902834782
- Dufour, S., Nguyen, N., and Frauenfelder, U. H. (2010). Does training on a phonemic contrast absent in the listener's dialect influence word recognition? *J. Acoust. Soc. Am.* 128, EL43–EL48. doi: 10.1121/1.3431102
- Eckes, T., and Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Lang. Test.* 23, 290–325. doi: 10.1191/0265532206lt330oa
- Escudero, P., Hayes-Harb, R., and Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *J. Phon.* 36, 345–360. doi: 10.1016/j.wocn.2007.11.002
- Fitzpatrick, T. (2006). Habits and rabbits: words associations and the L2 lexicon. *EUROSLA Yearbook* 6, 121–145. doi: 10.1075/eurosla.6.09fit
- Flege, J. E. (1995). Two procedures for training a novel second language phonetic contrast. *Appl. Psycholinguist.* 16, 425–442.
- Flege, J. E., Munro, M. J., and Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *J. Acoust. Soc. Am.* 95, 3623–3641. doi: 10.1121/1.409931
- Flege, J. E., Takagi, N., and Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *J. Acoust. Soc. Am.* 99, 1161–1173. doi: 10.1121/1.414884
- Forster, K. I., and Forster, J. C. (2003). DMDX: a Windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35, 116–124. doi: 10.3758/BF03195503
- Gelman, A., and Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press.
- Gollan, T. H., and Kroll, J. F. (2001). "Bilingual lexical access," in *The Handbook of Cognitive Neuropsychology: What Deficits Reveal About the Human Mind*, ed B. Rapp (New York, NY, US: Psychology Press), 321–345.
- Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001
- Gor, K. (2014). Raspberry, not a car: context predictability and a phonological advantage in early and late learners' processing of speech in noise. *Front. Psychol.* 5:1449. doi: 10.3389/fpsyg.2014.01449
- Gor, K. (2015). "Phonology and morphology in lexical processing," in *The Cambridge Handbook of Bilingual Processing*, ed J. Schwieter (Cambridge: Cambridge University Press), 173–199.
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia* 9, 317–323. doi: 10.1017/cbo9781107447257.007
- Hayes-Harb, R., and Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Lang. Res.* 24, 5–33. doi: 10.1177/0267658307082980
- Ingram, J. C., and Park, S. G. (1998). Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *J. Acoust. Soc. Am.* 103, 1161–1174. doi: 10.1121/1.421225
- Kuhl, P. K., and Iverson, P. (1995). "Linguistic experience and the perceptual magnet effect," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed W. Strange (Baltimore, MD: York Press), 121–154.
- Levenshtein, V. I. (1966, February). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10, 707–710.
- Linck, J. A., and Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Lang. Learn.* 65, 185–207. doi: 10.1111/lang.12117
- Luce, P. A., Goldinger, S. D., Auer, E. T., and Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Percept. Psychophys.* 62, 615–625. doi: 10.3758/BF03212113
- McClelland, J. L., Thomas, A. G., McCandliss, B. D., and Fiez, J. A. (1999). Understanding failures of learning: hebbian learning, competition for representational space, and some preliminary experimental data. *Prog. Brain Res.* 121, 75–80. doi: 10.1016/S0079-6123(08)63068-X
- Meara, P. (1983). Word associations in a foreign language: a report on the Birkbeck vocabulary project. *Nottingham Linguist. Circ.* 11, 29–38.
- Meara, P. (1984). "The study of lexis in interlanguage," in *Interlanguage*, eds A. Davies, C. Cripe, and A. Howatt (Edinburgh: Edinburgh University Press), 225–235.
- Meara, P. M. (1978). Learners' word associations in French. *Interlang. Stud. Bull.* 3, 192–211.
- Michael, E. B., and Gollan, T. H. (2005). "Being and becoming bilingual: Individual differences and consequences for language production," *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. F. Kroll and A. M. B. de Groot (New York, NY: Oxford University Press), 389–407.
- Niznik, M., Vinokurova, A., Voroncova, I., Kagan, O., and Cherp, A. (2009). *Русский без границ* [Russian without Borders]. Moscow: Russkij Mir.
- Ota, M., Hartsuiker, R. J., and Haywood, S. L. (2009). The KEY to the ROCK: near-homophony in nonnative visual word recognition. *Cognition* 111, 263–269. doi: 10.1016/j.cognition.2008.12.007
- Pallier, C., Bosch, L., and Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition* 64, B9–B17. doi: 10.1016/S0010-0277(97)00030-9

- Pallier, C., Colomé, A., and Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: exemplar-based versus abstract lexical entries. *Psychol. Sci.* 12, 445–449. doi: 10.1111/1467-9280.00383
- Pisoni, D. B., and Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition* 25, 21–52. doi: 10.1016/0010-0277(87)90003-5
- Polivanov, E. (1932). The subjective nature of the perception of sounds. *Polivanov* 223–237.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <http://www.R-project.org/>
- Reinisch, E., Weber, A., and Mitterer, H. (2013). Listeners retune phoneme categories across languages. *J. Exp. Psychol.* 39, 75–86. doi: 10.1037/a0027979
- Rüschmeyer, S. A., Nojack, A., and Limbach, M. (2008). A mouse with a roof? Effects of phonological neighbors on processing of words in sentences in a non-native language. *Brain Lang.* 104, 132–144. doi: 10.1016/j.bandl.2007.01.004
- Sekine, S. (2006). *The Effects of Phonological and Lexical Factors on L2 Word Recognition: A Cross-linguistic Study*. Unpublished Doctoral dissertation, University of California Los Angeles, Los Angeles, CA.
- Sheldon, A., and Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede speech perception. *Appl. Psycholinguist.* 3, 243–261. doi: 10.1017/S0142716400001417
- Singleton, D. (1999). *Exploring the Second Language Mental Lexicon*. Cambridge: Cambridge University Press.
- Singleton, D., and Little, D. (1991). The second language lexicon: some evidence from university-level learners of French and German. *Second Lang. Res.* 7, 62–81. doi: 10.1177/026765839100700103
- Singleton, D., and Singleton, E. (1998). “The C-test and L2 acquisition/processing research,” in *University Language Testing and the C-Test*, ed J. A. Coleman (Portsmouth: University of Portsmouth), 150–178.
- Stolz, W. S., and Tiffany, J. (1972). The production of ‘child-like’ word associations by adults to unfamiliar adjectives. *J. Verb. Learn. Learn. Behav.* 11, 38–46. doi: 10.1016/S0022-5371(72)80057-4
- Strange, W. (1995). “Cross-language studies of speech perception: a historical review,” in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed W. Strange (Baltimore, MD: York), 3–48.
- Veivo, O., and Järvisikivi, J. (2013). Proficiency modulates early orthographic and phonological processing in L2 spoken word recognition. *Bilingualism* 16, 864–883. doi: 10.1017/S136672912000600
- Weber, A., and Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *J. Mem. Lang.* 50, 1–25. doi: 10.1016/S0749-596X(03)00105-0
- Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: a depth of individual word knowledge model. *Stud. Second Lang. Acquis.* 23, 41–69. doi: 10.1017/S0272263101001024
- Yarkoni, T., Balota, D., and Yap, M. (2008). Moving beyond Coltheart’s N: a new measure of orthographic similarity. *Psychon. Bull. Rev.* 15, 971–979. doi: 10.3758/PBR.15.5.971

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Cook, Pandža, Lancaster and Gor. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Phonetic Encoding of Coda Voicing Contrast under Different Focus Conditions in L1 vs. L2 English

Jiyoun Choi^{1,2}, Sahayng Kim³ and Taehong Cho^{2*}

¹ ARC Centre of Excellence for the Dynamics of Language, Western Sydney University, Sydney, NSW, Australia, ² Hanyang Phonetics and Psycholinguistics Lab, Department of English Language and Literature, Hanyang University, Seoul, South Korea, ³ Hongik University, Seoul, South Korea

OPEN ACCESS

Edited by:

Miquel Simonet,
University of Arizona, USA

Reviewed by:

Hanyong Park,
University of Wisconsin-Milwaukee,
USA

Jeffrey Jackson Holliday,
Korea University, South Korea

Allard Jongman,
University of Kansas, USA

*Correspondence:

Taehong Cho
tcho@hanyang.ac.kr

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 29 January 2016

Accepted: 13 April 2016

Published: 13 May 2016

Citation:

Choi J, Kim S and Cho T (2016)
Phonetic Encoding of Coda Voicing
Contrast under Different Focus
Conditions in L1 vs. L2 English.
Front. Psychol. 7:624.
doi: 10.3389/fpsyg.2016.00624

This study investigated how coda voicing contrast in English would be phonetically encoded in the temporal vs. spectral dimension of the preceding vowel (in vowel duration vs. F1/F2) by Korean L2 speakers of English, and how their L2 phonetic encoding pattern would be compared to that of native English speakers. Crucially, these questions were explored by taking into account the phonetics-prosody interface, testing effects of prominence by comparing target segments in three focus conditions (phonological focus, lexical focus, and no focus). Results showed that Korean speakers utilized the temporal dimension (vowel duration) to encode coda voicing contrast, but failed to use the spectral dimension (F1/F2), reflecting their native language experience—i.e., with a more sparsely populated vowel space in Korean, they are less sensitive to small changes in the spectral dimension, and hence fine-grained spectral cues in English are not readily accessible. Results also showed that along the temporal dimension, both the L1 and L2 speakers hyperarticulated coda voicing contrast under prominence (when phonologically or lexically focused), but hypoarticulated it in the non-prominent condition. This indicates that low-level phonetic realization and high-order information structure interact in a communicatively efficient way, regardless of the speakers' native language background. The Korean speakers, however, used the temporal phonetic space differently from the way the native speakers did, especially showing less reduction in the no focus condition. This was also attributable to their native language experience—i.e., the Korean speakers' use of temporal dimension is constrained in a way that is not detrimental to the preservation of coda voicing contrast, given that they failed to add additional cues along the spectral dimension. The results imply that the L2 phonetic system can be more fully illuminated through an investigation of the phonetics-prosody interface in connection with the L2 speakers' native language experience.

Keywords: english coda voicing, L2 speech, prominence, focus, prosodic structure, phonetics-prosody interface, Korean learners of English

INTRODUCTION

Recent years have witnessed a rapidly growing body of research on the phonetics-prosody interface which illuminates how phonetic realization of segments is fine-tuned by higher-order prosodic structure, and how the prosodic structure is in turn manifested in the fine phonetic detail (e.g., Fougheron and Keating, 1997; de Jong, 2004; Cho and McQueen, 2005; Cho et al., 2014). In a widely received view on the phonetics-prosody interface, high-level prosodic structure is assumed to modulate not only fine-grained phonetic shaping of individual segments, but also phonetic encoding of paradigmatic phonological contrast. A well-known example comes from position-related phonetic modulation whereby a same segment is produced differently as a function of prosodic position (Fougheron and Keating, 1997; Cho and Keating, 2001, 2009). Another example is phonetic modulation due to prosodic structure that involves accent (or prominence in a broader sense). When the prominence of a linguistic unit (such as a syllable or a word) is expressed by a nuclear pitch accent (an element of prosodic structure in English; see Shattuck-Hufnagel and Turk, 1996), phonetic clarity of individual segments is heightened, maximizing phonological contrast through enhancement of phonetic features involved (e.g., de Jong, 1995, 2004; Cho and McQueen, 2005; Cho et al., 2014).

Accumulated evidence on such phonetics-prosody interface has led to a common consensus among researchers that a fuller understanding of the phonetic system of a given language should be accompanied by an understanding of the detailed aspects of the phonetics-prosody interface (see Fletcher, 2010; Cho, 2011, for a review). Theories of the phonetics-prosody interface, however, have been developed primarily based on L1 speech, leaving many relevant questions unanswered in L2 speech. Numerous studies on L2 speech production have indeed vigorously informed how L2 speech production is influenced by L1 phonetic knowledge both on the segmental level (e.g., Flege, 2003; Best and Tyler, 2007) and on the suprasegmental level (e.g., Munro, 1995; Trofimovich and Baker, 2006), but our understanding of the interplay between the two levels in L2 is still at an embryonic stage (cf. Davidson, 2011). In an effort to fill the gap, the present study explores the interplay between low-level phonetic realization and high-order prosodic structure in L1 (English) vs. L2 (by Korean learners of English) by investigating a case of phonetic modulation of coda voicing contrast in English.

The coda voicing contrast in English is known to be phonetically encoded in both the temporal and the spectral dimensions of the preceding vowels as reflected in vowel duration and formants (F1, F2; e.g., Chen, 1970; Wolf, 1978; Keating, 1985; Summers, 1987; Crowther and Mann, 1992; Maddieson, 1997; de Jong, 2004; Moreton, 2004). For example, in the temporal dimension, vowel duration is longer before a voiced than before a voiceless coda; and in the spectral dimension, the vowel (especially the low vowel, /æ/ or /a/) is produced with *lower* F1 (positioning the vowel *higher* in the vowel space), and *higher* F2 (positioning the vowel *more advanced* in the vowel space) before a voiced than a voiceless coda. Given its

multi-dimension cues, the case of coda voicing in English allows us to investigate how the cues in different phonetic dimensions are used differentially as a function of L2 speakers' native language experience.

The goal of the present study is therefore to investigate how the coda voicing contrast in English is indeed manifested in the fine-grained temporal vs. spectral dimensions of the preceding vowel in L1 vs. L2, with a view to understanding the influence of L1 experience on L2 phonetics-prosody interface. For L2 speakers, we chose native Korean (NK) learners of English because Korean differs from English in crucial ways which provide a basis for testing how the non-native coda voicing contrast would be modulated by L2 speakers' native language experience. Two crucial language-specific aspects relevant for the present study are (1) Korean does not employ voicing contrast in the coda position, and (2) Korean has a much smaller vowel inventory which might reduce their sensitivity to the spectral (formant) cues (see below for further discussion). In what follows, we will develop specific research questions in connection with these language-specific characteristics of Korean along with some predictions that ensue from NK speakers' native language experience.

The first question to be considered in the present study is how NK L2 speakers of English encode coda voicing contrast in the temporal vs. the spectral dimensions as compared with how NAE speakers do. In Korean, the laryngeal contrast of stops is completely neutralized to a voiceless unreleased stop (see Cho and McQueen, 2006, for a related discussion). This means that NK L2 speakers of English do not have native language experience with coda voicing contrast in any phonetic dimension, whether spectral or temporal. Given the lack of NK speakers' experience with coda voicing contrast, one might expect that NK speakers would have difficulties in encoding the coda contrast equally in the spectral and in the temporal dimension.

Alternatively, however, NK L2 speakers of English might use the temporal vs. the spectral cues in an asymmetric way. They might rely on the temporal dimension to encode coda voicing contrast, but their use of spectral dimension could be restricted, possibly attributable to the fact that Korean employs a more sparsely populated vowel space compared to English. For example, there are only two contrastive vowels (/i/ and /ɛ/) in the front region of the vowel space in Korean (e.g., Yang, 1996) as opposed to five in English (/i, ɪ, eɪ, ɛ, æ/). With the experience of a sparsely populated vowel space in their native language, NK speakers are supposed to be less sensitive to fine-grained changes in the spectral (formant) dimension than NAE speakers are. This possibility is in fact in line with an assumption in the literature that speakers of a language with a sparsely populated vowel space has larger Difference Limens (DLs, or Just Noticeable Differences). For example, native listeners of Japanese (with a relatively smaller vowel inventory) show DLs as large as 13% of the formant frequency (Nakagawa et al., 1982), while DLs as small as 1% has been reported for native listeners of American English (with a larger vowel inventory; Kewley-Port and Watson, 1994; Kent and Read, 2002 for a related discussion; see also Iverson and Evans, 2009, for a study showing an advantage of a complex L1 vowel space in L2 learning). It is therefore plausible to assume that small changes in the spectral dimension

are not easily accessible to NK L2 speakers of English, and therefore they do not utilize the spectral cues to the coda voicing contrast, or at least not in as fine-grained a way as the native speakers do.

The cues used in the temporal dimension, on the other hand, appear to be more readily available to L2 speakers as has often been noted by previous researchers (e.g., Flege and Hillenbrand, 1986; Bohn, 1995; Escudero et al., 2009, to name a few). Bohn (1995) proposed that L2 speakers' propensity to rely more on a temporal cue than on a spectral cue (when both cues are available for an L2 contrast) is attributable to the universally driven perceptual salience of durational cues, although L2 speakers' native language experience may further modulate the L2 speakers' use of durational cues (see Broselow and Kang, 2013 for a review; Escudero et al., 2009 for related discussion). For example, Flege and Hillenbrand (1986) showed that native speakers of three languages (French, Swedish, and Finnish) all exploited vowel duration in perceiving the coda voicing contrast between /z/ and /s/ in English, but the effect was smaller for native speakers of French than those of Swedish and Finnish, showing some native language effect. Crucially, speakers of these languages all showed comparable reliance on durational cues to coda voicing, independently of the amount of their exposure to English. This is again consistent with the universally driven perceptual account—i.e., because the durational cues are perceptually salient, the durational cues can be easily exploited by L2 speakers regardless of the speakers' native language and their English proficiency. The salient nature of durational cues may also be reflected in the tendency that listeners rely on durational cues more than F0 cues at prosodic junctures in lexical segmentation when processing an unfamiliar language (e.g., Tyler and Cutler, 2009; Kim et al., 2012). Non-native speakers indeed appear to exploit a temporal cue in L2 even if the specific temporal cue is not directly used in their native language. In Arabic, for example, the stop voicing contrast is maintained in coda position, but Arabic does not systematically use the vowel duration cue to coda voicing presumably because vowel duration is preserved for maintaining phonemic length (quantity) contrast between vowels (see de Jong and Zawaydeh, 2002 for a related discussion). Nevertheless, Arabic L2 speakers of English utilized the vowel duration to encode the coda voicing contrast in English (Flege and Port, 1981).

Taken together, it is reasonable to assume that the phonetic cues are different in nature in terms of whether they are expressed in the temporal vs. the spectral dimensions, so that the former tends to be universally exploitable while the latter is more prone to be language-specifically tuned. Under this view, phonetic cues in the temporal dimension in L2 are more readily accessible to L2 speakers than are those in the spectral dimension, leading to a prediction that NK speakers will be able to encode the coda voicing better in the temporal than in the spectral dimension. Furthermore, given the perceptually driven accessibility of durational cues, one might expect that the NK speakers would show a similar phonetic encoding pattern of coda voicing along the temporal dimension, regardless of their English proficiency. The fact that the coda voicing effect on vowel

duration is a near-universal phenomenon (Chen, 1970; Keating, 1985; Maddieson, 1997; Cho, 2015, for a review) indeed appears to reinforce these predictions.

Another important question of the present study concerns how NK L2 speakers of English would express the phonetics-prosody interface as to be reflected in modulation of coda voicing contrast as a function of prominence (which stems from a prosodic structure of a given utterance—e.g., Beckman, 1996; Shattuck-Hufnagel and Turk, 1996; Fletcher, 2010; Cho, 2011). It has been well-documented in the literature that phonological distinction is enhanced or *hyperarticulated* in the prominent (accented) condition while it is attenuated or *hypoarticulated* when the segment occurs in the non-prominent (unaccented) condition (e.g., de Jong, 1995, 2004; Cho and McQueen, 2005; Cho et al., 2011, 2014). For example, de Jong (2004) showed that NAE speakers hyperarticulated (or exaggerated) the durational difference due to coda voicing in order to maximize phonological distinction of the coda voicing contrast (at least in the temporal dimension) in the prominent context, whereas the coda voicing contrast was minimized or hypoarticulated in the non-prominent context. Such an interaction between coda voicing and prominence, however, was not observed in the spectral dimension (F1 and F2), although the coda voicing contrast itself was still reflected in the spectral dimension. In the present study, we extend this study to L2 speech by investigating the extent to which NK L2 speakers of English exploit the acoustic-phonetic space for phonetic encoding of coda voicing contrast as a function of the prominence system of prosodic structure.

If NK L2 speakers of English indeed fail to use the spectral dimension to encode coda voicing contrast, and if the L2 temporal cues are readily accessible to NK speakers, NK speakers are also likely to use the temporal dimension for phonetic modulation of coda voicing contrast as a function of prominence. More crucial questions, however, are how efficiently NK speakers use the temporal dimension along a hypo- to hyper-articulation (H&H) continuum (cf. Lindblom, 1990) to express *both* the phonological voicing contrast and its interaction with the prominence system, and how much the NK speakers' way of using the temporal dimension is attributable to their native language experience. In order to address these questions in connection with communicative efficiency in L2 (to be reflected in the way that the H&H continuum is exploited by L2 speakers), we integrated the prominence factor with information structure which is often assumed to be mediated by the prominence system of prosodic structure (e.g., de Jong, 2004). That is, the prominence conditions (accented vs. unaccented) were obtained with three different focus types that were assumed to stem from information structure, so that we could examine how phonetic encoding of coda voicing could be fine-tuned as a function of information associated with different focus types. Thus, the target-bearing words were produced with one of the following focus types: (1) phonologically contrastive focus in which the coda voicing contrast (*bed* vs. *bet*) was directly emphasized; (2) lexically contrastive focus in which a target-bearing word was contrastive with a semantically related word (*bed* vs. *chair*); and (3) no focus in which the target-bearing word was in the

background with a contrastive focus being placed elsewhere in the utterance.

As for the focus effects in L1, de Jong (2004) already showed that different focus types induced different degrees in the coda voicing effect on the preceding vowel duration. The vowel length difference was found to be enhanced when the targets were focused (phonologically or lexically) compared to when the targets received no focus, and more importantly, the focus effect was found to be more robust in the phonological than in the lexical focus condition. This suggests that the phonetics-prosody interface as reflected in enhancement of coda voicing under prominence is further modulated by higher-order information structure, which may be taken to be driven by an optimization of communicative efficiency in response to information structure. That is, it appears that speakers make an articulatory effort focusing on either a particular phonological contrast or the whole lexical item to enhance the locus of information as signaled by information structure (driven by the principle of contrast maximization), while they ease articulation when the target is not the locus of information (driven by the principle of effort minimization; cf. Lindblom, 1990; Flemming, 1995). The present study builds on this assumption in L1, and further explores the extent to which such communicative efficiency may be reflected in L2 production by NK speakers. The L2 system is in fact considered to operate through the interaction between principles of contrast maximization and effort minimization (see Hawkins, 2014, for a related discussion), and one might therefore expect that NK speakers would show an interaction between coda voicing and focus in a way similar to that of NAE speakers, as far as the common goal is concerned—i.e., to achieve communicative efficiency in response to information structure. But as non-native speakers, NK speakers might not be able to show as efficient a pattern as native speakers do, not only because they have less experience with the L2 communicative system as a whole, but also because their production is likely to be affected by their native language experience. As briefly mentioned above, while native (NAE) speakers use both the temporal and the spectral dimension to encode the coda voicing contrast, NK L2 speakers of English are likely to rely exclusively on the temporal dimension to maintain the coda voicing contrast in a communicatively efficient way as regulated by information structure. If this is the case, with the lack of spectral cues to coda voicing contrast, NK speakers' use of temporal dimension would be restricted to the extent that the phonological voicing contrast in the temporal dimension is not blurred when the system prefers hypoarticulation.

Finally, the present study examines the coda voicing effect on syllable-onset Voice Onset Time (VOT). One of the traditional explanations for the coda voicing effect on the preceding vowel duration may be that the rate of (V-to-C) closure formation for the voiced stop is slower (Chen, 1970), which implies that the temporal effect is localized to a later part of the vowel which roughly corresponds to the closing gesture for the coda. Most recently, however, in an acoustic study, Pycha and Dahan (2016) showed that coda voicing influences the relative timing of the nucleus and the offglide for a diphthong /aɪ/, implying that the effect is not local but global, regulating the temporal organization of the first and the second components of the vowel. The

hypothesized global articulatory effect is further consistent with a perceptual account—i.e., an acoustically defined vowel would be lengthened, enhancing the percept of voicing for a voiced coda (see Raphael, 2005, for a review). If the vowel lengthening due to coda voicing is entirely perceptually driven, the lengthening effect does not need to be localized to a later part of the vowel. None of these explanations, however, predicts the coda voicing effect on the syllable-onset VOT, as VOT is not involved in closure formation of the coda, nor does it contribute to the voiced percept or the nucleus-offglide timing as it is by nature voiceless. From an articulatory point of view, however, the onset of the vocalic (mouth opening) gesture for the vowel (i.e., the release of closure) coincides with the onset of VOT and therefore VOT may be considered as a 'voiceless' part of the vowel (for example, in the framework of Articulatory Phonology, Browman and Goldstein, 1992). If the coda voicing effect on the vowel is localized near the coda consonant, it is expected to influence the closing gesture for the coda, but not the vowel's opening gesture that includes VOT in the vowel's temporal domain. In such a case, VOT will not vary as a function of coda voicing. Alternatively, if the coda voicing effect influences the temporal structure of the entire vowel including the vowel's opening gesture, VOT as part of the vowel is expected to be longer before a voiced coda just like the acoustic vowel duration is. In the present study, we test this possibility in both L1 and L2 speech.

MATERIALS AND METHODS

Participants and Recording

Thirty-six speakers participated in the study for monetary reward. They were 12 native speakers of American English (six females, six males, aged: 21–33, mean age = 26), 12 Korean advanced learners of English with an average TOEFL score of 110 (six females, six males, aged: 21–26, mean age = 23), and 12 Korean intermediate learners of English with an average TOEFL score of 75 (six females, six males, aged: 21–28, mean age = 24). The native speakers of English were exchange students, English teachers or visitors residing in Seoul at the time of recording. The Korean learners of English were all university students. All participants were naïve as to the purpose of the present study. The speech data were recorded in a soundproof booth at the Hanyang Phonetics and Psycholinguistics Lab, with a Tascam HP-Ps digital recorder and a SHURE KSN44 microphone at a sampling rate of 44.1 kHz.

Speech Materials and Procedure

Four minimal pairs of English CVC words differing in the voicing of coda stops were used as in (1):

- (1) (a) front mid vowel /ɛ/: *bed-bet, ped-pet*
 (b) front low vowel /æ/: *bad-bat, pad-pat*

Each of the eight target words (in the four pairs) in two different vowel contexts (/ɛ, æ/) was placed in a carrier sentence, which was an answer to a question in a mini discourse situation. The mini discourse was used to induce the desired variety of accent-placement patterns with different focus types and

TABLE 1 | Example sentences with a target *bed*.

IP-initial (IPi)	PH-FOC	A: Did you write 'BET fast again'?
		B: Not exactly. ' <u>BED</u> fast again' was what I wrote.
	LEX-FOC	A: Did you write 'CHAIR fast again'?
IP-medial (IPm)		B: Not exactly. ' <u>BED</u> fast again' was what I wrote.
	NoFOC	A: Did you write 'bed SLOWLY again'?
		B: Not exactly. 'bed FAST again' was what I wrote.
IP-medial (IPm)	PH-FOC	A: Did you write 'say BET fast again'?
		B: No, I wrote 'say <u>BED</u> fast again.'
	LEX-FOC	A: Did you write 'say CHAIR fast again'?
		B: No, I wrote 'say <u>BED</u> fast again.'
	NoFOC	A: Did you write 'say bed SLOWLY again'?
		B: No, I wrote 'say <u>bed</u> FAST again.'

The target word is underlined. Focused words are in uppercase letters.

prosodic groupings. Example sentences with a target word *bed* are given in **Table 1**.

As can be seen in the table (underlined) target words always occurred in the second sentence ('B') preceded by a prompt question ('A') which was used to induce an intended focus type for the target word. Following de Jong and Zawaydeh (2002) and de Jong (2004), the focus types were manipulated as follows (see Gussenhoven, 2007 for a comprehensive review of focus types):

- **phonological (segmental) focus** (PH-FOC): with a corrective contrastive focus on the phonological voicing of codas between the target word and the contrastive counterpart in the prompt question which form a minimal pair (e.g., *BET* vs. *BED*).
- **lexical focus** (LEX-FOC): with a corrective contrastive focus on the target word which is semantically related to the contrasting word in the question (e.g., *CHAIR* vs. *BED*).
- **no focus** (background; No FOC): with a corrective contrastive focus on a word following the target word, so that the target word became the background in the focus-background structure (e.g., *bed SLOWLY* vs. *bed FAST*).

As can be seen in **Table 1**, the position of the target word was controlled to be either in the initial or in the medial position of the Intonational Phrase (IP), because prosodic position may interact with prominence. The target word was placed either in the initial position of a sentence (e.g., *Not exactly. 'Bed fast again' was what I wrote*), which is likely to be the beginning of an IP, or in the middle of an IP (e.g., *No. I wrote 'say bed fast again'*), given the likelihood that the phrase 'say bed fast again' forms an IP.

The prompt questions were pre-recorded by a female native speaker of American English who had been trained to produce intended focus-inducing patterns. During the recording, subjects first silently read the question-and-answer sentences on a computer screen. They then heard the pre-recorded prompt question, and answered it aloud as written on the screen. The first 36 trials were practice trials, so that speakers familiarized themselves with focus types in different mini discourse situations. They were asked to speak casually at a comfortable speech rate as if they were talking to a friend. The practice session was

repeated when a speaker was not fluent enough to place focus naturally. The entire set of the sentences was repeated three times in a randomized order. Whenever a speaker misplaced focus or produced the intended IP (e.g., '*Bed fast again*') with a strong prosodic juncture inside, the speaker was asked to repeat the sentence a few more times to obtain a token with the best-matched intended focus or position. A total of 5184 tokens (36 speakers \times 8 target words \times 3 focus types \times 2 positions \times 3 repetitions) were obtained. The collected tokens were further checked by all three authors on the placement of pitch accent on the focused word, and position of the test word.¹ Thirteen tokens were further discarded, as agreed by all three authors, due to inadequate prosodic junctures around the test word or a misplacement of pitch accent.

Measurements

In order to investigate effects of focus and position on the acoustic realization of the English coda voicing contrast, four acoustic parameters were measured:

- **Duration of preceding vowel (V-duration)**: vowel duration was taken as the interval from the onset of the voicing of the preceding vowel (defined as the zero-crossing point before the first positive peak of the periodic waveform which is largely aligned with the onset of F1 seen in the spectrogram) to the vowel offset (defined as the point of abrupt discontinuity in the amplitude of the waveform which coincides with the offset of F2 seen in the spectrogram).
- **F1 and F2**: F1 and F2 of the vowels (/ε, æ/ as in *bed-bet* and *bad-bat*) were taken from the steady-state region (which was largely around the midpoint of the vowel) from the spectrographic displays.² Although much discussion on the coda voicing effect on formants has been based on the results obtained with F1, we included both F1 and F2 as they both have been found to show coda voicing effects on English monophthongs (Wolf, 1978; Summers, 1988). The values were obtained by using an LPC formant tracking function with hand corrections based on visual inspection of the spectrogram for each token.
- **Voice Onset Time (VOT)**: VOT of the voiceless (aspirated) stop /p/ in the syllable onset position (*ped-pet*, *pad-pat*) was taken from the point of the stop release to the voice onset of the vowel (as defined for the vowel duration measure).

¹The prosodic transcriptions employed here were based on the conventions of the English ToBI (tones and break indices; Beckman and Ayers, 1994, Unpublished; Beckman et al., 2005). In ToBI, a tone with "*" or a starred tone (e.g., H* or L+H*) refers to a pitch accent that falls on a lexically stressed syllable along with a higher-level (phrasal) stress. H* means that a tone rises and reaches its peak largely in the vowel without a noticeable low tone that precedes it whereas L+H* means that the starred high tone (H*) is realized primarily on a lexically stressed syllable preceded by a low tone. In the present study, the pitch accent type observed with the target words was either H* or L+H*.

²Although effects on F1 and F2 could be more robust near the coda consonant, formant values at the edge of the vowel appeared to be quite variable due to V-to-C formant transitions. Furthermore, previous researchers (e.g., Wolf, 1978; Summers, 1988) indicated that the coda voicing effect on formants was reliably observed in steady-state parts of the vowel. Our initial informal inspection of formant values with some speakers' tokens also indicated that the coda effects were robust even in the middle of the vowel. We therefore decided to include F1 and F2 measures at the midpoint (steady-state) of the vowel.

VOTs for voiced stops were not included partly because they could often be negative (voice lead) adding an additional complexity, and partly because in an ‘aspiration’ language like English the already short VOTs for voiced stops are not expected to vary much as a function of various factors (e.g., Kessinger and Blumstein, 1997; Smiljanić and Bradlow, 2008).

Statistical Analyses

In order to evaluate statistically the effects of prosodic factors and vowel context on English coda voicing as produced by different groups of speakers, a series of repeated measures Analyses of Variance (RM ANOVA) was conducted using SPSS 21 statistical package for windows on the acoustic measures mentioned above. At first, statistical analyses were performed separately for each language group (Native American English, NAE, vs. Native Korean, NK) with four within-subject factors, Voicing (voiced /d/ vs. voiceless /t/ coda), Vowel type (V-type: /e/ vs. /æ/), Focus (PH-FOC vs. LEX-FOC vs. NoFOC), and Position (IP-initial vs. IP-medial). For the NK group, there was a between-subject factor, Group (NK-advanced vs. NK-intermediate). Combined analyses were then conducted, with four within-subject factors listed above and one between-subject factor, Native Language (NAE vs. NK). When there were interactions between factors, *post hoc* pairwise comparisons were performed with Bonferroni/Dunn corrections. *p*-values less than 0.05 were considered statistically significant, and the values between 0.05 and 0.08 were treated as a trend. In the following section, we first outline the results, present the statistical results separately by each language group (NAE vs. NK) for each of the acoustic parameters, and provide combined analyses with both language groups.

RESULTS

Vowel Duration

Effects of Voicing on V-duration and its possible interactions with Focus and Vowel Type are illustrated for each speaker group in **Figure 1**. As can be visually observed in **Figure 1A**, both the NAE (native American English) and the NK (native Korean) speakers showed robust coda voicing effects on V-duration. Crucially, the effect was augmented in the focused conditions but attenuated in the unfocused conditions across the board. The figure also shows the interaction between Voicing and Focus was further conditioned by the speakers’ native language: the focus-induced augmentation of the coda voicing effect on vowel duration tended to be greater for the NAE speakers (**Figure 1A1**) than for the NK speakers (**Figures 1A2,3**), whereas the reverse was true in the unfocused (NoFOC) condition in which the lengthening effect was more extremely attenuated by the NAE speakers than by the NK speakers (both advanced and intermediate). Furthermore, it is observable from **Figure 1B** that the NAE speakers maintained a clear durational division for the intrinsic vowel height between the mid and the low vowels (/e/ vs. /æ/; **Figure 1B1**), but that the division was less clear for the NK-advanced speakers (**Figure 1B2**) and it entirely disappeared for the NK-intermediate speakers (**Figure 1B3**), while the difference

in V-duration due to coda voicing remained unchanged. These observations were statistically supported by RM ANOVAs as reported below.

Effects on V-Duration by NAE (L1 ENG)

The NAE speakers showed a main effect of Voicing on V-duration, such that the vowel was longer before a voiced than before a voiceless stop (/d/ vs. /t/; mean difference 25.8 ms, $F[1,11] = 57.4$, $p < 0.001$). The Voicing effect, however, interacted with Focus ($F[2,22] = 46.9$, $p < 0.001$). As shown in **Figure 1A1**, the Voicing by Focus interaction stemmed from a focus-sensitive Voicing effect: the Voicing effect was augmented in the focused conditions [PH-FOC, mean difference 39.9 ms, $t(11) = 59.2$, $p < 0.001$; LEX-FOC, mean difference 32.8, $t(11) = 56.9$, $p < 0.001$] while the effect was extremely attenuated in the unfocused condition [NoFOC, mean difference 4.7 ms, $t(11) = 5.9$, $p < 0.05$]. Furthermore, the interaction appeared to be in part due to, on the average, a larger Voicing effect in the phonologically focused (PH-FOC) than in the lexically focused (LEX-FOC) condition (39.9 ms vs. 32.8 ms). There was also a three-way interaction between Voicing, Focus and Vowel Type ($F[2,22] = 17.2$, $p < 0.001$), such that the focus-sensitive voicing effect on V-duration was further conditioned by Vowel Type: as can be seen in **Figure 1B1**, there was a small but significant Voicing effect on V-duration for /æ/ in the NoFOC condition [NoFOC, mean difference 8.9 ms, $t(11) = 3.3$, $p < 0.01$] but not for /e/ [NoFOC, mean difference 0.4 ms, $t(11) = 0.3$, $p = 0.79$]. Another noteworthy observation was that while the Voicing effect was robust for both vowel types (/e/ vs. /æ/), there was a significant interaction between Voicing and Vowel Type ($F[1,11] = 41.3$, $p < 0.001$). As can be inferred from **Figure 1B1**, the interaction was due to the fact that the coda voicing effect was larger for the low vowel /æ/ [mean difference 36.5 ms, $t(11) = 70.3$, $p < 0.001$] than for the mid vowel /e/ [mean difference 15.1 ms, $t(11) = 23.2$, $p < 0.01$], presumably because the intrinsically longer (low) vowel has a greater degree of freedom for temporal expansion. It is also worth mentioning that there was a four-way interaction which included the Position factor: Voicing \times Focus \times Vowel Type \times Position ($F[2,22] = 13.1$, $p < 0.01$). The four-way interaction, however, was too complicated to be fully understood, but a visual inspection indicated that one of the contributing patterns (figure not shown) to the interaction was that the Voicing effect on the duration of /æ/ in the NoFOC condition turned out to have stemmed mostly from a robust voicing effect on /æ/ in the IP-initial position [/æ/, NoFOC, IP-initial, mean difference 11.7 ms, $t(11) = 3.6$, $p < 0.01$; NoFOC, IP-medial, mean difference 6.2 ms, $t(11) = 1.8$, $p = 0.09$].

Effects on V-Duration by NK (L2 ENG)

Native Korean speakers also showed a robust main effect of Voicing on V-duration, such that it was longer before a voiced than before a voiceless stop (mean difference 23.8 ms, $F[1,22] = 53.43$, $p < 0.001$). As was the case with the NAE speakers, the Voicing effect interacted with Focus (Voicing \times Focus, $F[1,44] = 16.1$, $p < 0.001$) due to the fact that the Voicing effect was augmented in the focused conditions

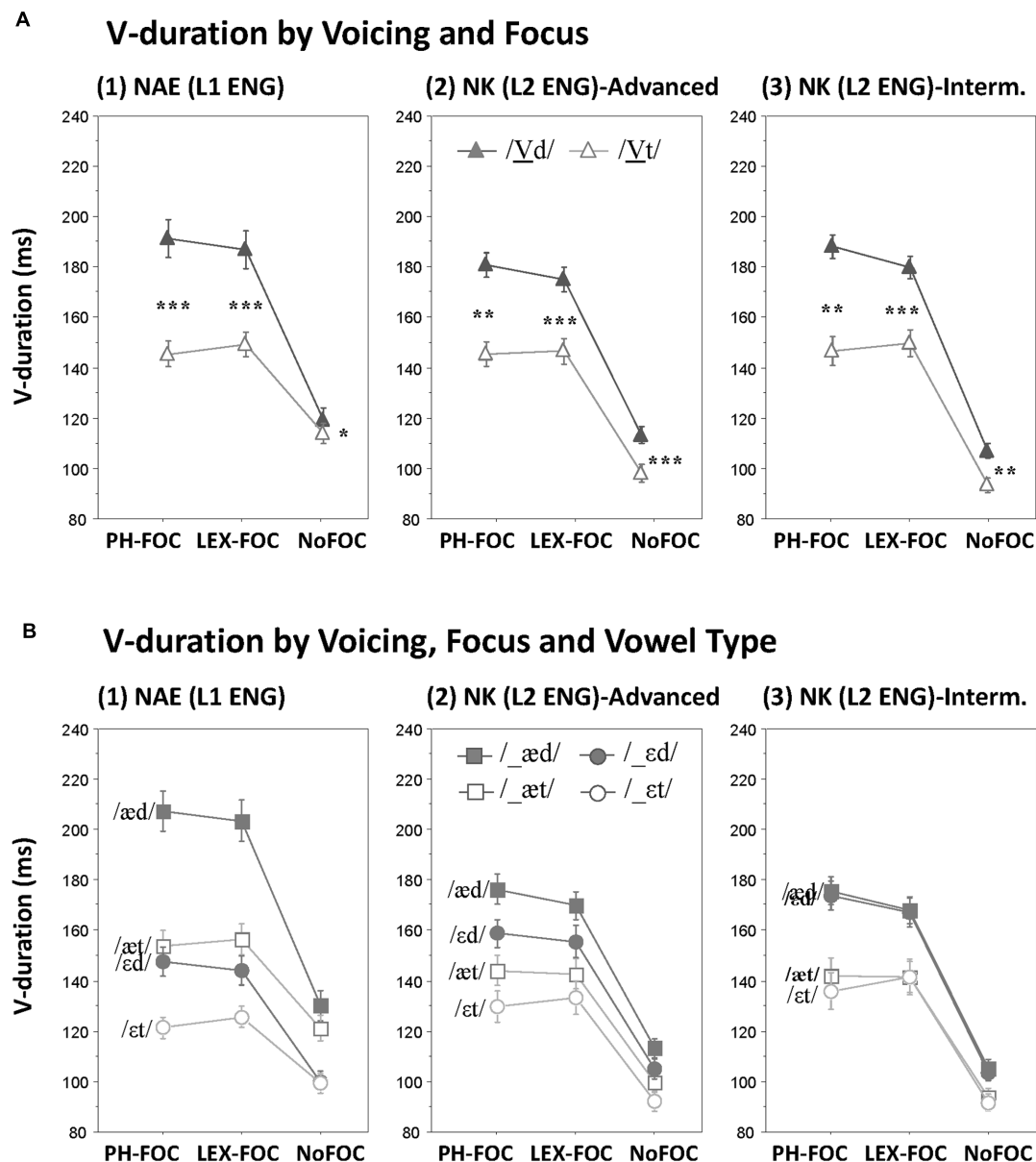


FIGURE 1 | Effects of coda voicing on vowel duration. (A) Voicing \times Focus interactions; **(B)** Voicing \times Focus \times Vowel type interactions, as produced by (1) native speakers of English, (2) Korean advanced learners of English, and (3) Korean intermediate learners of English (** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

[PH-FOC, mean difference 33.4 ms, $t(23) = 6.2$, $p < 0.001$; LEX-FOC, mean difference 25.4 ms, $t(23) = 7.3$, $p < 0.001$], but attenuated in the unfocused (NoFOC) condition [mean difference 12.5 ms, $t(23) = 7.2$, $p < 0.001$]. Furthermore, as was the case with the NAE speakers, the NK speakers also showed a similar tendency toward a larger Voicing effect in the phonologically focused (PH-FOC) than in the lexically focused (LEX-FOC) conditions (33.4 ms vs. 25.4 ms). This interaction was observed for both NK-advanced and NK-intermediate speakers as visually shown in **Figures 1A2,3** and statistically confirmed—i.e., there was no further interaction with Group (Voicing \times Focus \times Group, $F[2,44] < 1$, $p > 0.6$).

There was no other interaction effect that involved Voicing, except for a Voicing \times Position interaction ($F[1,22] = 4.92$, $p < 0.05$). Planned t -tests, however, indicated that there was no noticeable difference in the Voicing effect on V-duration as a function of Position [IP-initial, mean difference 22.4 ms, $t(23) = 7.1$, $p < 0.001$; IP-medial, mean difference 25.2 ms, $t(23) = 7.5$, $p < 0.001$]. This suggests that Position did not heavily modulate the temporal variation of the vowel due to coda voicing. It is also worth mentioning that there was a significant interaction between Vowel and Group: the NK-advanced speakers marked the intrinsic durational difference between / ε / and / æ / [mean difference 11.97, $t(11) = 7.5$, $p < 0.05$;

see **Figure 1B2**] while the NK-intermediate speakers showed a complete overlap between the two vowels [mean difference 2 ms, $t(11) = 1.74$, $p > 0.2$; see **Figure 1B3**]. Thus, although the NK-intermediate speakers failed to use the vowel duration cue for the intrinsic vowel height difference, they used the cue successfully for marking the phonological voicing contrast of the following stops even in the non-prominent (unfocused) context.

Combined Analyses on V-Duration Across NAE (L1 ENG) and NK (L2 ENG)

The above-observed patterns on the coda voicing effects on V-duration were based on RM ANOVAs, separately carried out for the NAE and NK speakers. The results of a combined analysis (a five-way ANOVA) with an additional factor Native Language (NAE vs. NK) indeed showed a significant three-way interaction: Voicing \times Focus \times Language ($F[2,68] = 4.197$, $p < 0.05$). As seen in **Figure 1A**, the augmented Voicing effect on V-duration in the focused condition was on the average *larger* for the NAE than for the NK speakers (PH-FOC, 39.9 ms vs. 33.4 ms; LEX-FOC, 32.8 ms vs. 25.4 ms), whereas the attenuated voicing effect in the unfocused condition was on the average *smaller* for the NAE than for the NK speakers (NoFOC, 4.7 ms vs. 12.5 ms). The results of the combined analysis also showed a four-way interaction: Voicing \times Focus \times Vowel Type \times Language ($F[1,68] = 9.44$, $p < 0.001$). As seen in **Figure 1B**, the NAE speakers showed a three-way interaction between Voicing, Focus and Vowel Type while the NK speakers did not.

Voice Onset Time

Effects of Voicing on VOT and its possible interactions with Focus and Vowel Type are illustrated for each speaker group in **Figure 2**. As can be visually observed in **Figure 2A**, the most striking pattern was that VOT (which may be taken as the initial component of the articulatory vocalic gesture in the temporal dimension) was indeed influenced by the voicing of the following coda, such that VOT for the voiceless stop (/p/) was on the average longer before a voiced than before a voiceless coda (/d/ vs. /t/) for both the NAE and the NK groups. Unlike the Voicing effect on V-duration, however, the results for both the NAE and the NK speakers did not show a noticeable interaction between Voicing and Focus. (Recall that the voicing effect on V-duration was augmented in the focused condition but attenuated in the unfocused condition). **Figure 2B** shows a possible difference that came from the speakers' native languages, especially in terms of whether Voicing further interacted with Focus and Vowel Type. As can be seen in **Figure 2B1**, for the mid vowel /e/ (but not for the low vowel /æ/), the NAE speakers indeed show an augmented Voicing effect on VOT in the *focused* conditions (PH-FOC and LEX-FOC) as compared with the Voicing effect in the *unfocused* (NoFOC) condition. The NK speakers, as can be seen in **Figures 2B2,3**, showed no such interaction, although the NK-advanced speakers shows some resemblance to the NAE's interaction pattern. These observations were statistically supported by RM ANOVAs as reported below.

Effects on VOT by NAE (L1 ENG)

The NAE speakers showed a significant main effect of Voicing on VOT ($F[1,11] = 9.7$, $p < 0.05$) such that VOT for the voiceless stop /p/ in the onset position was longer when the coda was voiced than when it was voiceless (as shown in **Figure 2A1**). There was, however, a significant two-way interaction between Voicing and Vowel Type ($F[1,11] = 8.3$, $p < 0.05$). Planned *t*-tests indicated that the two-way interaction stemmed from the fact that the Voicing effect was reliable only for the mid vowel /e/ [mean difference 8.7 ms, $t(11) = 4.2$, $p < 0.01$], but not for the low vowel /æ/ [mean difference 1.9 ms, $t(11) = 0.9$, $p = 0.37$], as can be seen in **Figure 2B1**. There was also a three-way interaction between Voicing, Focus and Vowel Type ($F[2,22] = 3.8$, $p < 0.05$) which was due to the fact that the Voicing effect on VOT before the mid vowel /e/ was larger in the focused conditions than in the unfocused (NoFOC) conditions [PH-FOC, mean difference 10.4 ms, $t(11) = 3.4$, $p < 0.01$; LEX-FOC, mean difference 11.2, $t(11) = 3.1$, $p < 0.05$; NoFOC, mean difference 4.4 ms, $t(11) = 2.6$, $p < 0.05$]. There was no other significant interactions that involved the Voicing factor.

Effects on VOT by NK (L2 ENG)

Like the NAE speakers, the NK speakers showed a significant main effect of Voicing on VOT ($F[1,22] = 20.64$, $p < 0.001$), such that VOT for the voiceless stop in the onset position was longer when the coda was voiced than when it was voiceless (**Figures 2A2,3**). Unlike the case with the NAE speakers, the Voicing effect did not interact with Vowel Type and Focus: there was no Voicing by Vowel Type interaction ($F[2,44] < 1$, $p > 0.3$), nor was there a three-way interaction between Voicing, Focus and Vowel Type ($F[2,44] < 1$, $p > 0.5$), as can be inferred from **Figures 2B2,3**. There was no further interaction with Group, indicating that speakers of both the NK-advanced and the NK-intermediate groups did not modulate the Voicing effect as a function of Focus and Vowel Type. There was no other significant interaction that involved the Voicing factor.

Combined Analyses on VOT Across NAE (L1 ENG) and NK (L2 ENG)

A combined analysis with Native Language as an additional factor returned a significant main effect of Voicing ($F[1,34] = 28.26$, $p < 0.001$) with no interaction between Voicing and Language ($F[1,64] < 1$, $p > 0.7$). This confirmed the robust coda voicing effect on the onset VOT across speakers of both native and non-native groups (NAE and NK). The combined analysis also showed a significant three-way interaction: Voicing \times Vowel Type \times Language ($F[1,34] = 11.7$, $p < 0.005$), reflecting the fact that the NAE speakers showed an interaction between Voicing and Vowel (i.e., the voicing effect was significant only for the mid vowel /e/), while the NK showed the effect for both /e/ and /æ/. However, there was no four-way interaction of Voicing \times Focus \times Vowel Type \times Language ($F[2,68] = 1.83$, $p > 0.1$), despite the fact that there was a significant three-way interaction of Voicing \times Focus \times Vowel Type for the NAE speakers, but not for the NK speakers. Thus, the results of the combined analysis indicated that the differential Voicing effects on VOT as a function of speakers' native language was most

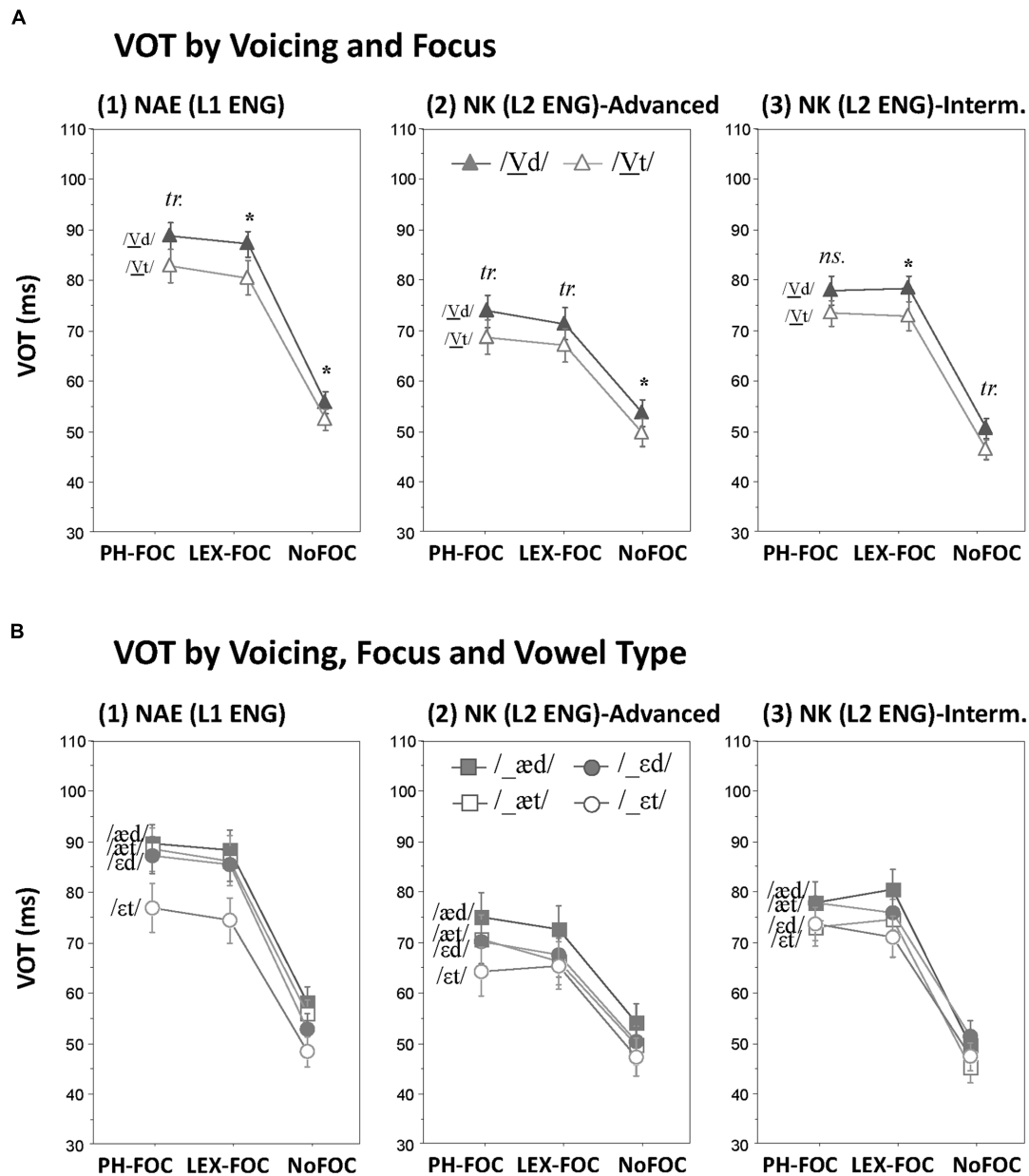


FIGURE 2 | Effects of coda voicing on VOT. (A) Voicing × Focus interactions; (B) Voicing × Focus × Vowel type interactions, as produced by (1) native speakers of English, (2) Korean advanced learners of English, and (3) Korean intermediate learners of English ($t_r = p < 0.08$, $*p < 0.05$).

reliably evident in the presence or absence of a Voicing × Vowel interaction for NAE vs. NK.

F1

Effects of Voicing on F1 and its possible interactions with Focus and Vowel Type are illustrated for each speaker group in Figure 3. As can be visually observed in Figure 3B1, the NAE speakers employed the F1 cue not only for marking the phonemic contrast between the mid vowel and the low vowel (/ɛ/-/æ/), but also for marking the voicing contrast of the following codas with F1 being lower before a voiced than before a voiceless coda, thus

positioning the vowels higher in the vowel space. On the other hand, the NK speakers (Figures 3A2,3) did not use the F1 cue at all for marking the coda voicing contrast, although the NK-advanced speakers did use the F1 cue for making a distinction between the mid and the low vowels. These observations were statistically supported by RM ANOVAs as reported below.

Effects on F1 by NAE (L1 ENG)

There was a main effect of Voicing on F1 of the preceding vowel ($F[1,11] = 70.9$, $p < 0.001$), such that F1 was lower (thus positioning the vowel higher in the vowel space) before

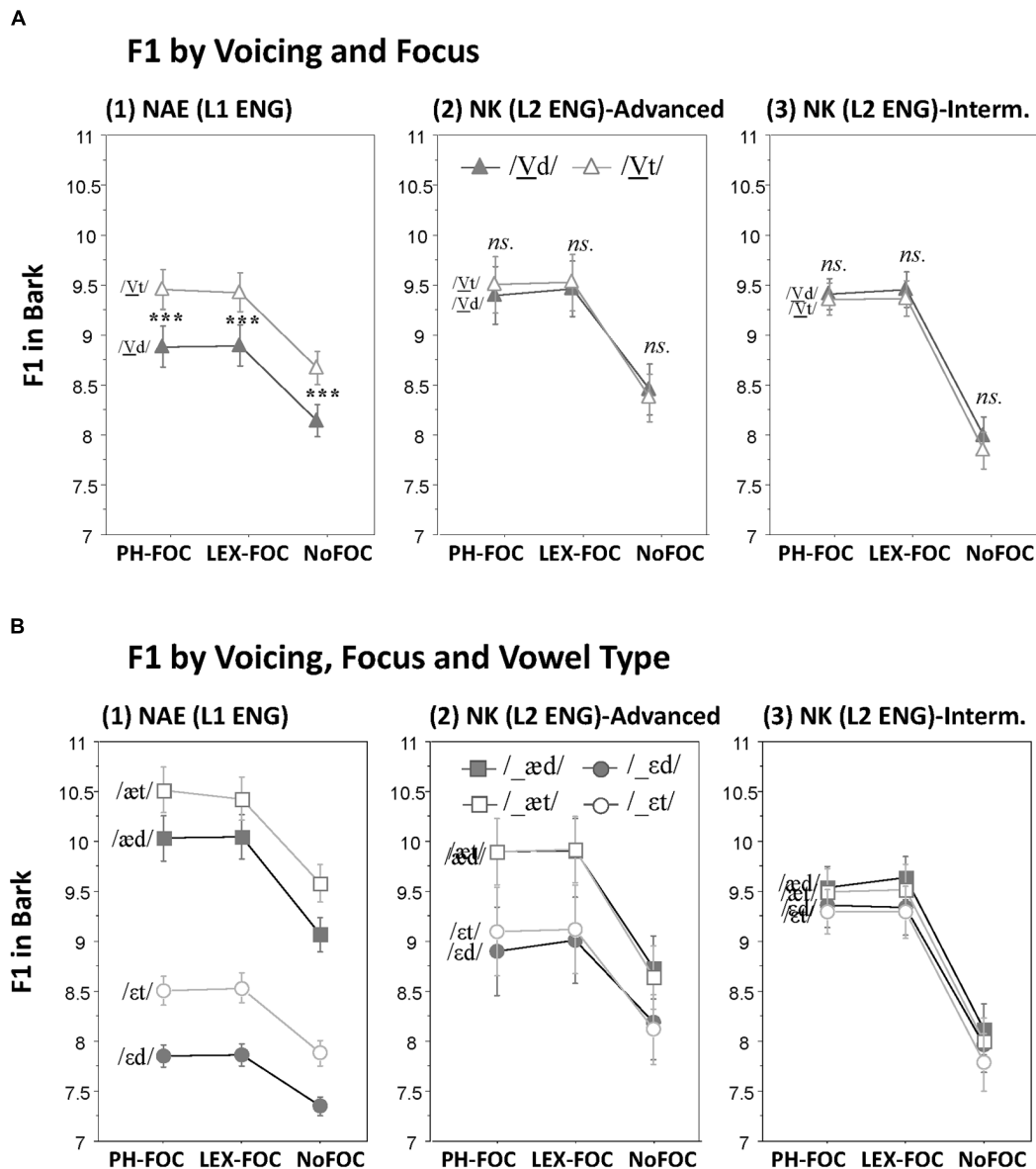


FIGURE 3 | Effects of coda voicing on F1. (A) Voicing \times Focus interactions; **(B)** Voicing \times Focus \times Vowel type interactions, as produced by (1) native speakers of English, (2) Korean advanced learners of English, and (3) Korean intermediate learners of English (** $p < 0.001$).

a voiced than before a voiceless coda (/d/ vs. /t/; see the general pattern in **Figure 3A1** for NAE). Unlike V-duration, F1 showed no Voicing by Focus interaction ($F[2,22] < 1$, $p > 0.4$), suggesting that the Voicing effect (lower F1 before a voiced coda) remained unchanged across different focus types (as can be seen in **Figure 3A1**). RM ANOVAs also returned a significant three-way interaction between Voicing, Focus and Vowel Type ($F[2,22] = 3.6$, $p < 0.05$), but planned t -tests indicated that the Voicing effect on F1 remained significant in each focus condition for each vowel type (all at $p < 0.001$) showing the same direction. As can be visually inferred from **Figure 3B1**, the three-way interaction effect appeared to have stemmed from the fact that the F1 difference due to coda voicing was on the average larger

in the phonologically focused (PH-FOC) than in the lexically focused (LEX-FOC) for /æ/ (mean difference 0.49 vs. 0.38 Bark, respectively), but not for /ɛ/ (mean difference 0.66 vs. 0.67 Bark, respectively). (Compare the F1 difference in the PH-FOC vs. LEX-FOC conditions for the /ɛd/-/ɛt/ pair vs. the /æd/-/æt/pair in **Figure 3B1**).

Effects on F1 by NK (L2 ENG)

For NK speakers, there was no main effect of Voicing on F1 ($F[1,22] < 1$, $p > 0.1$) nor was there any interaction between factors that involved Voicing. In particular, the fact that the Voicing factor did not interact with Group (NK-advanced vs. NK-intermediate; $F[1,22] = 1.87$, $p > 0.1$) indicates that neither

group of NK speakers employed the F1 cue for marking the coda voicing contrast. This null voicing effect on F1 is illustrated in **Figures 3A2,3** for each group. Furthermore, it is interesting to note that the NK-advanced speakers made a clear phonemic distinction between / ϵ / and / \ae / (**Figure 3B2**), although the NK-intermediate speakers did not (**Figure 3B3**). But they both failed to use the F1 cue for the coda voicing contrast. In other words, the NK-advanced speakers did use the F1 cue for the phonemic vowel contrast, but not for the voicing contrast of the following codas. A visual inspection of the results, however, suggested that the NK-advanced speakers may possibly employ the F1 cue for the voicing contrast at least in one particular condition—i.e., in the PH-FOC condition for the / ϵ d/-/ ϵ t/ pair, as can be seen in **Figure 3B2**. Planned *t*-tests indeed showed that there was a small but significant voicing effect only in this particular condition ($F[1,11] = 5.0, p < 0.05$).

Combined Analyses on F1 Across NAE (L1 ENG) and NK (L2 ENG)

As reported above, the results of RM ANOVAs run separately for each native language group (NAE and NK) showed clearly that the spectral F1 cue for the coda voicing contrast was employed by the NAE but not by the NK speakers. A five-way RM ANOVA with Native Language as an additional factor returned a significant Voicing and Language interaction on F1 ($F[1,34] = 51.4, p < 0.001$) confirming the speakers' differential use of the F1 cue as a function of their native language.

F2

Effects of Voicing on F2 and its possible interactions with Focus and Vowel Type are illustrated for each speaker group in **Figure 4**. As can be visually observed in the figure (and as was the case with F1), the NAE speakers employed the F2 cue for marking both the phonemic contrast (/ ϵ /-/ \ae /; **Figure 4B1**) and the voicing contrast of the following codas with F2 being higher before a voiced coda (**Figure 4A1**), thus positioning the vowels more advanced in the vowel space before a voiced than before a voiceless coda. On the other hand, the NK speakers did not use the F2 cue at all for marking the coda voicing contrast (**Figures 4A2,3**), although the NK-advanced speakers did use the F2 cue for making a distinction between the mid and the low vowels (/ ϵ / vs. / \ae /; **Figure 4B2**). These observations were statistically supported by RM ANOVAs as reported below.

Effects on Voicing by NAE (L1 ENG)

The NAE speakers showed a main effect of Voicing on F2, such that F2 was higher before a voiced coda /d/ than before a voiceless coda /t/ ($F[1,11] = 40.4, p < 0.001$), which positioned the vowel before a voiced coda more advanced in the vowel space (**Figure 4A1**). As was the case with F1, F2 also showed a vowel-independent voicing effect. That is, there was no interaction between Voicing and Vowel Type ($F[1,11] = 2.55, p > 0.1$), indicating that the Voicing effect on F2 (higher F2 for a voiced coda) was applicable to both the mid vowel / ϵ / and the low vowel / \ae / as shown in **Figure 4B1**. Just like on F1, the Voicing effect on F2 did not interact with Focus ($F[2,22] < 1, p > 0.9$), nor was there a three-way interaction between Voicing, Focus

and Vowel Type ($F[2,22] < 1, p > 0.8$), indicating no further modulation of the Voicing effect as a function of Focus and Vowel Type. Voicing did not interact with Position, either ($F[1,11] < 1, p > 0.9$), showing a position-independent voicing effect (figure not shown). There was no other interaction effect that involved Voicing.

Effects on Voicing by NK (L2 ENG)

Unlike the NAE speakers, the NK speakers did not generally employ the F2 cue in marking the coda voicing contrast: there was no main effect of Voicing on F2 ($F[1,22] < 1, p > 0.9$). There was no significant interaction between Voicing and Group ($F[1,22] < 1, p > 0.3$), either, indicating that both the NK-advanced and NK-intermediate speakers failed to use the F2 cue. There was, however, a three-way interaction: Voicing \times Focus \times Group ($F[2,44] = 3.27, p < 0.05$). The interaction was due to the fact that while the NK-advanced speakers showed no Voicing effect on F2 in each focus condition (**Figure 4A2**), the NK-intermediate speakers showed a significant Voicing effect in the LEX-FOC condition (**Figure 4A3**). But as can be seen in **Figure 4A3**, the Voicing effect in the LEX-FOC condition was the opposite of what was found for the NAE speakers. It is also interesting to note that just as they used F1, the NK-advanced speakers also used F2 to make a phonemic distinction between / ϵ / and / \ae / ($F[1,11] = 6.0, p < 0.05$; **Figure 4B2**), resembling the NAE's / ϵ /-/ \ae / distinction, whereas the NK-intermediate speakers did not ($F[1,11] = 1.04, p > 0.3$). The NK-advanced speakers, however, failed to utilize the F2 cue for the coda voicing contrast. There was no other interaction that involved the Voicing factor.

Combined Analyses on F2 Across NAE (L1 ENG) and NK (L2 ENG)

As reported above, the results of RM ANOVAs run separately for each native language group (NAE and NK) showed clearly that the spectral F2 cue for the coda voicing contrast was employed by the NAE but not by the NK speakers. A five-way RM ANOVA with Native Language as an additional factor returned a significant Voicing and Language interaction on F2 ($F[1,34] = 25.46, p < 0.001$) confirming the speakers' differential use of the F2 cue as a function of their native language.

DISCUSSION

In the present study we investigated how coda voicing contrast in English would be manifested in the acoustic-phonetic detail of the preceding vowel in both the temporal and the spectral dimensions. Crucially, we compared speech productions in L1 (by 12 native speakers of American English, NAE) and L2 (by 24 non-native Korean learners of English, NK) with a view to understanding the phonetics-prosody interface in L1 and L2. To this end, we tested effects of prominence that stemmed from prosodic structure closely related to information structure as reflected in different focus types: phonological focus (PH-FOC), lexical focus (LEX-FOC), and no focus (NoFOC). We also controlled for the prosodic position factor, so that the test

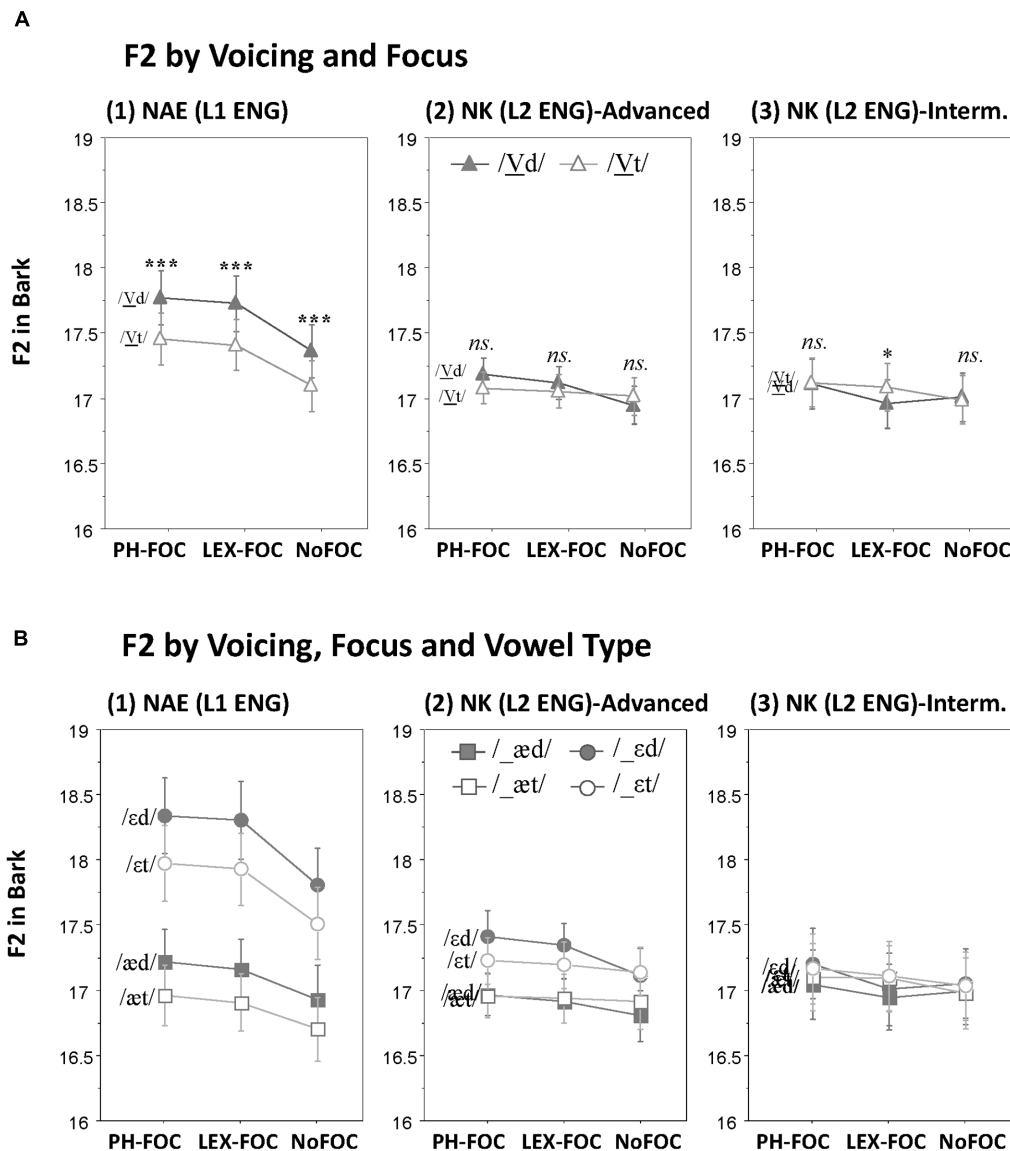


FIGURE 4 | Effects of coda voicing on F2. (A) Voicing \times Focus interactions; **(B)** Voicing \times Focus \times Vowel type interactions, as produced by (1) native speakers of English, (2) Korean advanced learners of English, and (3) Korean intermediate learners of English (** $p < 0.001$, * $p < 0.05$).

words in different focus conditions occurred both in the phrase-initial and the phrase-medial positions (i.e., the IP-initial vs. the IP-medial position). In what follows, we recapitulate several important findings that have emerged from the results along with some discussion on implications for phonetic encoding of phonological contrast and its interaction with higher order linguistic structure in L2 speech.

Differential Use of Phonetic Dimensions in L1 vs. L2

One of the basic findings of the present study is that both the native (NAE) and the non-native (NK) speakers showed robust coda voicing effects on the temporal realization of the preceding

vowel. The vowel duration was systematically longer before a voiced than before a voiceless coda stop in the production of both the NAE and the NK speakers. The effect was independent of the prosodic position (IP-initial vs. IP-medial) in which the target bearing word occurred. (See the next section for discussion on an interaction of voicing and focus.) The coda voicing effect in the temporal dimension was further evident in the syllable-onset VOT. Both the native (NAE) and the non-native (NK) speakers showed a significant main effect of coda voicing on the syllable-onset VOT which was longer before a voiced than before a voiceless coda. Interestingly, however, the NAE speakers showed an interaction effect on VOT between Voicing and Vowel: the NAE speakers showed the voicing effect on VOT for the mid vowel pair (*ped-pet*), but not for the low vowel pair

(*pad-pat*), whereas the non-native (NK) speakers showed no such interaction. We do not have any principled explanation to offer for why there is such an asymmetric coda voicing effect on VOT in the native (NAE) speakers' production, but one cannot entirely rule out the possibility that the asymmetry has stemmed from the lexical differences (e.g., word frequency) between the two pairs. On the other hand, the fact that the non-native (NK) speakers showed a consistent coda voicing effect on VOT regardless of the word pair may then be interpretable as stemming from the possibility that non-native speakers are less sensitive to the lexical differences in speech production. While these possibilities need further corroborations, what appears to be clear is that the coda voicing effect on the syllable-onset VOT is less robust than that on the vowel next to the coda at least for the NAE speakers, possibly reflecting the proximity effect—i.e., VOT is not adjacent to the source of coda voicing.

From an articulatory gestural point of view, as discussed in the introduction, VOT may be taken to be part of the vowel, given that the onset of the vocalic opening gesture for the vowel coincides with the onset of VOT. The lengthened VOT before a voiced coda therefore suggests that coda voicing affects the entire temporal structure of the vowel (cf. Pycha and Dahan, 2016), rather than being localized to a later part of the vowel (cf. Chen, 1970). From an acoustic point of view, on the other hand, VOT is considered as part of the syllable-onset voiceless stop, so that the effect on VOT defined as such further implies that coda voicing may modify the temporal structure of the entire syllable even beyond the preceding vowel. This rather long distant effect is in line with the case for the syllable-onset /l/ whose phonetic realization was found to be modulated by the voicing of the syllable-coda (e.g., Nguyen and Hawkins, 2004). Importantly, although the non-native (NK) speakers have no experience with such a phonological coda voicing contrast in their native (Korea) language, they appear to modulate the temporal structure of the entire vowel (or possibly the entire syllable) in a comparable way as the native speakers do.

Unlike the coda voicing effects in the temporal dimension, however, the way that coda voicing contrast was manifested in the spectral dimension was clearly bifurcated between L1 (NAE) and L2 (NK) speaker groups. The native (NAE) speakers showed robust effects of coda voicing on both F1 and F2 for the monophthong vowels /ε, æ/ largely in line with the previous studies (Wolf, 1978; Summers, 1987; Crowther and Mann, 1992). Both the mid and the low vowels /ε, æ/ were produced with *lower* F1 and *higher* F2 before a voiced than a voiceless stop (thus positioning the vowel *higher* (lower F1) and *more advanced* (higher F2) before a voiced stop in the acoustic vowel space). It is also worth pointing out that previous studies observed lower F1 before a voiced coda only for low vowels (/æ/ or /ɑ/) with an interpretation that the coda voicing effect was due to 'hyperarticulation' of the vowel before a *voiceless* coda (as reflected in *higher* F1 before a voiceless coda and *lower* F1 before a voiced one), possibly enhancing the [+low] feature for the low vowel (e.g., see Thomas, 2000; Moreton, 2004 for a related discussion). The present study demonstrated that the same holds for the non-low (mid) vowel /ε/, indicating that

the assumed hyperarticulation does not necessarily enhance the vowel's distinctive feature—i.e., the increase in F1 before a voiceless stop is taken to enhance the [+low] for the low vowel /æ/, but not for the mid vowel /ε/.

Most crucially, however, unlike the NAE speakers, the non-native (NK) speakers did not show any evidence of their use of spectral cues to the coda voicing contrast. A question that arises here is then why there is discrepancy in the way that the non-native (NK) speakers employ the temporal dimension vs. the spectral dimension for encoding coda voicing contrast in L2 English—i.e., they successfully encode coda voicing contrast in the temporal dimension, but fail to do so in the spectral dimension. The asymmetric use of the temporal vs. the spectral dimension by the non-native (NK) speakers may be accounted for by different natures of the phonetic cues in the temporal vs. the spectral dimensions. On the one hand, as Bohn (1995) noted, cues in the temporal dimension may be taken to be perceptually more salient. The temporal dimension in fact is exploited to express a wide range of linguistic contrast (whether syntagmatic or paradigmatic) across languages (see Cho, 2015 for a review), presumably because of its universally driven perceptual salience. This view is consistent with previous observations: speakers rely more on temporal cues in processing an unfamiliar language (e.g., Tyler and Cutler, 2009; Kim et al., 2012); and infants are indeed sensitive to prosodic variation of speech input (including variation along the temporal dimension) even at an embryonic stage of L1 acquisition, and exploit prosodic cues in lexical segmentations (see Johnson, 2016 for a review). Furthermore, the fact that the coda voicing effect on the preceding vowel duration is a near-universal tendency (e.g., Chen, 1970; Keating, 1985; Maddieson, 1997) implies that the temporal cue for the coda voicing in L2 is likely to be unmarked and hence easily accessible to non-native speakers. Thus, the universally applicable use of temporal dimension appears to make it easier for the non-native (NK) speakers to encode the coda voicing contrast along the temporal dimension.

The failure of using the spectral dimension, on the other hand, appears to have stemmed from the speakers' native language experience. Specifically, this possibility is in line with the view that speakers of a language with a sparsely populated vowel space has larger Difference Limens (DLs, or Just Noticeable Differences), thus being less sensitive to a small change in formant frequencies than speakers of a language with a densely populated vowel space (see Kent and Read, 2002 for a related discussion). The observed null effect of coda voicing on F1 and F2 for the non-native (NK) speakers can therefore be interpreted as having stemmed from the NK speakers' native language experience whose smaller vowel inventory induces perceptual insensitivity to formant frequencies. On a related point, it is also worth pointing out that the NK-advanced learners of English (but not the NK-intermediate speakers) indeed used the spectral cues (F1, F2) to make a categorical phonemic distinction between the mid and the low vowels /ε, æ/, though the phonetic distance between the two vowels was not as large as that produced by the native (NAE) speakers. But even the NK-advanced speakers failed to use the spectral cues in a finer-grained way

for marking coda voicing contrast. This is again in line with the prediction regarding differential perceptual sensitivities as a function of the size of the vowel inventory of the speakers' native language.

These possibilities, taken together, suggest that the difference in how NK speakers use temporal and spectral dimensions stems from the fact that, in this case, one of the cues is universally driven and the other is L1-specific. It is therefore plausible that phonetic encoding of phonological contrast in L2 is constrained by an intricate relationship between the universal applicability of a phonetic cue for a given contrast and the non-native speakers' language experience.³

The Phonetics-Prosody Interface with Reference to Information Structure in L1 vs. L2

Another important finding of the present study was that both the NAE and the NK speakers showed a significant Voicing \times Focus interaction in the temporal dimension. The coda voicing contrast was temporally enhanced under prominence, such that the vowel lengthening effect due to coda voicing was augmented in the focused conditions (both phonologically focused and lexically focused) whereas the effect was extremely attenuated in the unfocused condition. In other words, insofar as the temporal dimension was concerned, both the native (NAE) and the non-native (NK) speakers showed a comparable phonetics-prosody interface as reflected in the interplay between the phonetic realization of the coda voicing contrast and the prosodic prominence factor. The way that coda voicing interacted with focus may be interpreted as being driven by an interaction of two important principles of the linguistic communicative system: *contrast maximization* and *effort minimization* (e.g., Lindblom, 1990; de Jong, 1995; Flemming, 1995). In the focused condition—i.e., when signaled by the prominence (accentuation) factor of the prosodic structure in connection with information structure, both the NAE and the non-native (NK) speakers *hyperarticulate* by making effort to maximize the distinctiveness of coda voicing contrast. In the unfocused condition—i.e., when the prosodic structure signals that voicing contrast is no longer the locus of information, they ease articulatory effort or *hypoarticulate*. Furthermore, the fact that both the NK-advanced and the NK-intermediate speakers showed a similar interaction pattern as the native (NAE) speakers did suggests that the interplay between phonetics and prosody in L2 speech operates in a communicatively optimized way, regardless of the non-native speakers' English proficiency, by making reference to higher-order information structure.

Another noteworthy finding was that both the NAE and the non-native (NK) speakers showed a trend toward a greater enhancement of coda voicing contrast in the phonologically

focused (PH-FOC) than in the lexically focused (LEX-FOC) condition consistent with findings of previous studies (e.g., de Jong and Zawaydeh, 2002; de Jong, 2004). This result has some implications for the interaction between information structure and prosodic structure. Even if the focus realization from information structure was mediated by a nuclear pitch accent as part of the prominence system in the prosodic structure, the same nuclear pitch accent induced a finer-grained phonetic effect as a function of focus type (see Mücke and Grice, 2014, for a related discussion). This suggests that the prosodic structure effect is fine-tuned by making reference to information structure. Furthermore, the fact that the non-native (NK) speakers show a similar pattern indicates that such a fine-tuning according to information structure is characteristic of a human linguistic system, and thus is readily reflected in L2 speech.

The interaction of coda voicing and focus on vowel duration, however, was further modulated by the speakers' native language. The non-native (NK) speakers enhanced the coda voicing contrast in the focused conditions but not as much as the native (NAE) speakers did, and they reduced the coda voicing contrast in the unfocused condition but not as extremely as the native (NAE) speakers did. Recall that the NAE speakers, when in the unfocused condition, did not even show any vowel lengthening effect due to coda voicing for the mid vowel / ϵ /, while the non-native (NK) speakers consistently maintained the voicing contrast for the vowel in the unfocused condition. These results therefore suggest that the native (NAE) speakers use the acoustic temporal space along a hypo- to hyper-articulation continuum in a polarized way for optimization of communication efficacy, while the non-native (NK) speakers do not seem to utilize the space as efficiently as the native speakers do. In other words, although the non-native (NK) speakers do encode coda voicing contrast by making reference to information structure mediated by the phonetics-prosody interface, it appears that the native-like encoding of coda voicing requires a further phonetic fine-tuning of vowel duration in response to communicative functional load that stems from information structure.

The difference in the voicing by focus interaction between the native (NAE) and the non-native (NK) speakers, however, does not seem to be entirely attributable to the non-native speakers' less efficient way of utilizing the phonetic space, but it may also be at least in part due to the constraint from the L2 system in which the way that the non-native (NK) speakers maintain the phonological voicing contrast is different from that of the native (NAE) speakers. In the present study, the native (NAE) speakers did not show an interaction between coda voicing and focus in the spectral dimension (F1 and F2), but they used the F1 and F2 spectral cues consistently, which helps preserving coda voicing contrast even in the unfocused condition. Thus, even an extreme reduction of the voicing effect in the temporal dimension in the L1 system (as was the case for / ϵ / in the unfocused condition) is not detrimental to the maintenance of the phonological voicing contrast as the difference due to coda voicing is invariantly present in the spectral dimension of the speech signal. On the other hand, the non-native (NK) speakers did not employ the spectral cue to the

³One might wonder whether it is the temporal difference due to coda voicing that is taught explicitly in EFL class in Korea, which, if so, would have influenced the NK-speakers' performance. To our best knowledge, however, the vowel length difference due to coda voicing is never taught explicitly in both the primary and the secondary school.

coda voicing in their L2 system. With the lack of the spectral cue, too extreme a reduction of the voicing effect in the temporal dimension would undermine the phonological coda voicing contrast. In other words, an optimization of temporal realization of coda voicing in response to information structure appears to be constrained by the way that coda voicing is phonetically encoded in the L2 phonetic system, such that the phonological contrast is invariantly maintained. The phonetic optimization of the phonetics-prosody interface in the L2 phonetic system can therefore be taken to be modulated by the non-native speakers' native language experience.

Implications for Phonological Abstraction and Lexical Representation in L2 System

The fact that the NK speakers' sensitivity to the spectral vs. the temporal dimension in L2 is modulated by L1 (Korean) sound system indicates that phonetic encoding of coda voicing contrast is internalized in their L2 system in an L1-specific way, so that NK speakers' phonetic manifestations of phonological abstraction deviate from those of the native (NAE) speakers. Such an L1-specific abstraction appears to be further supported by our anecdotal observation that while the native (NAE) speakers showed an asymmetric coda voicing effect on VOT for the *pad-pat* vs. the *ped-pet* pair (presumably in part due to the lexical differences), non-native (NK) speakers showed a consistent effect on VOT regardless of lexical pair. Non-native (NK) speakers' impoverished lexical knowledge therefore appears to increase the role of phonological abstraction in phonetic encoding, hence the across-the-board voicing effect, although this possibility is subject to corroboration by further studies.

These observations have some implications for the nature of lexicon in L2. Recent years have witnessed a constructive debate in the literature on the nature of lexical representations, especially regarding how much phonetic detail is stored in the lexicon. One of recent approaches to this question is an exemplar-based approach (e.g., Goldinger, 1996, 1997; Pierrehumbert, 2001, 2003). It generally assumes a phonetically rich lexicon which stores phonetically detailed exemplars of specific speech events (also known as 'episodes') in a multidimensional phonetic space. A phonological contrast then emerges as phonetic categories are formed as a result of generalizations over a frequency weighted distribution of exemplars. Such a model is especially useful in accounting for effects of lexical frequency and individual differences which are prevalent in both speech production and perception. The perception-based exemplars are used in speech production, so that a phonological contrast is phonetically encoded based on random sampling from the frequency weighted distribution of exemplars associated with different (contrastive) phonetic categories. Although the theory has been developed primarily based on L1 speech, phonetic categories in L2, in principle, should be formed in a similar way. To the extent that the theory holds, however, the results of the present study indicate that phonetic detail of exemplars stored in the lexicon in L2 (developed by NK speakers) is different from that in L1 lexicon. In other words, the phonetic dimensions along

which the perceived exemplars form a category appear to be constrained by the L2 speakers' native language experience—i.e., L2 speakers' perceptual bias due to their L1 experience constrains the distribution of exemplars in a multidimensional phonetic space. Furthermore, the fact that NK speakers showed no clear English proficiency effect (and no lexical item effect on VOT) implies some degree of phonological abstraction, leading to a question as to the extent to which L2 speakers' phonetic encoding is indeed based on random sampling from a frequency weighted distribution of exemplars from the phonetically rich lexicon.

CONCLUSION

In the present study, we have demonstrated that the low-level phonetic encoding of phonological coda voicing contrast in L1 vs. L2 English (by Korean learners of English) is modulated by the prominence factor of prosodic structure in connection with information structure. Specifically, the results suggest that the non-native (NK) speakers' phonetic encoding of coda voicing contrast is modulated by an intricate interaction between the universal-applicability of phonetic cues used for the contrast along the temporal dimension and the non-native speakers' native language experience which constrains a finer-grained use of the spectral cue (presumably due to its scarcely populated vowel inventory). Furthermore, just like the native (NAE) speakers, the non-native (NK) speakers showed that their phonetic encoding of coda voicing was modulated by information structure mediated by the phonetics-prosody interface. Regardless of the non-native speakers' English proficiency, the L2 use of the acoustic phonetic space was polarized in a communicatively efficient way, in response to functional loads dictated by information structure. This suggests that once the relative acoustic phonetic cue is learned, the phonetics-prosody interface by making reference to higher order information structure appears to follow relatively easily, presumably because such an interaction is characteristic of the human linguistic communicative system, not specific to an individual language. However, the communicative efficacy of using the temporal dimension in L2 by the non-native (NK) speakers appeared to be less optimal compared to that in L1 speech. We proposed that such difference is also attributable to the non-native (NK) speakers' native language experience. Given that the spectral dimension is not used by the non-native (NK) speakers for marking coda voicing contrast, the communicative efficacy along the temporal dimension is achieved in a way that is not detrimental to the maintenance of the phonological contrast of the coda voicing. The non-native use of the temporal dimension therefore appears to be 'optimized' for a particular L2 communicative system by the NK speakers.

All in all, the present study has built on a gradually growing body of L2 phonetic literature with respect to the phonetics-prosody interface in L2. There is no doubt that speech production of L1 and L2 alike is modulated by a human communicative system in such a way that the information that comes from higher-order linguistic or information structure is encoded in

speech signal, and it is eventually available to the listener. Nevertheless, our understanding of L2 speech has been fairly limited, especially with respect to how low-level phonetic realization is systematically modulated by higher-order linguistic structure. The results of the present study therefore have further implications for theories of L2 speech (e.g., Flege, 2003; Best and Tyler, 2007; Davidson, 2011), for which there appears to be much room for further development regarding how the low-level phonetic implementation interacts with prosodic structure, how higher level linguistic information is further mediated by the phonetics-prosody interface, and how such interactions are constrained by the L2 speakers' native language experience.

AUTHOR CONTRIBUTIONS

JC: the first author, designed the study from the beginning, carried out the experiment, analyzed the data, and wrote up an earlier version of this manuscript; SK: the second author,

participated in the study at every stage, designing the study, analyzing the data, interpreting the results, discussing the results, and editing earlier versions of the manuscript; TC: the corresponding author, supervised the entire project at every stage, and edited the entire manuscript with elaborations on introduction, research questions, predictions, results section, interpretations of the results, and discussion with implications.

FUNDING

This work was supported by the research fund of Hanyang University (HY-2016) to the corresponding author (TC).

ACKNOWLEDGMENT

We would like to thank our RAs, Yuna Baek and Jiyoung Jang for their assistance in data acquisition and acoustic measurements.

REFERENCES

- Beckman, M. E. (1996). The parsing of prosody. *Lang. Cogn. Process.* 11, 17–67. doi: 10.1080/016909696387213
- Beckman, M. E., Hirschberg, J., and Shattuck-Hufnagel, S. (2005). "The original ToBI system and the evolution of the ToBI framework," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed. S.-A. Jun (Oxford: Oxford University Press), 9–54.
- Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: commonalities and complementarities," in *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production*, eds M. J. Munro and O. S. Bohn (Amsterdam: John Benjamins), 13–34.
- Bohn O. S. (1995). "Cross-language speech perception in adults: first language transfer doesn't tell it all," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. W. Strange (Timonium, MD: York Press), 379–410.
- Broselow, E., and Kang, Y. (2013). "Second language phonology and speech," in *The Cambridge Handbook of Second Language Acquisition*, eds J. Herschensohn and M. Young-Scholten (Cambridge, MA: Cambridge University Press), 529–554.
- Browman, C. P., and Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica* 49, 155–180. doi: 10.1159/000261913
- Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22, 129–159. doi: 10.1159/000259312
- Cho, T. (2011). "Laboratory phonology," in *The Continuum Companion to Phonology*, eds N. C. Kula, B. Botma, and K. Nasukawa (London: Continuum), 343–368.
- Cho, T. (2015). "Language effects on timing at the segmental and suprasegmental levels," in *The Handbook of Speech Production*, ed. M. A. Redford (Hoboken, NJ: Wiley-Blackwell), 505–529.
- Cho, T., and Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *J. Phon.* 29, 155–190. doi: 10.1006/jpho.2001.0131
- Cho, T., and Keating, P. (2009). Effects of initial position versus prominence in English. *J. Phon.* 37, 466–485. doi: 10.1016/j.wocn.2009.08.001
- Cho, T., Lee, Y., and Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *J. Phon.* 39, 344–361. doi: 10.1016/j.wocn.2011.02.005
- Cho, T., Lee, Y., and Kim, S. (2014). Prosodic strengthening on the /s/-stop cluster and the phonetic implementation of an allophonic rule in English. *J. Phon.* 46, 128–146. doi: 10.1016/j.wocn.2014.06.003
- Cho, T., and McQueen, J. (2005). Prosodic influences on consonant production in Dutch: effects of prosodic boundaries, phrasal accent and lexical stress. *J. Phon.* 33, 121–157. doi: 10.1016/j.wocn.2005.01.001
- Cho, T., and McQueen, J. (2006). Phonological versus phonetic cues in native and nonnative listening: Korean and Dutch listeners' perception of Dutch and English consonants. *J. Acoust. Soc. Am.* 119, 3085–3096. doi: 10.1121/1.2188917
- Crowther, C. S., and Mann, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *J. Acoust. Soc. Am.* 92, 711–722. doi: 10.1121/1.403996
- Davidson, L. (2011). Phonetic and phonological factors in the second language production of phonemes and phonotactics. *Lang. Linguist. Compass* 5, 126–139. doi: 10.1111/j.1749-818X.2010.00266.x
- de Jong, K. (1995). The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *J. Acoust. Soc. Am.* 97, 491–504. doi: 10.1121/1.412275
- de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *J. Phon.* 32, 493–516. doi: 10.1016/j.wocn.2004.05.002
- de Jong, K., and Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration. *J. Phon.* 30, 53–75. doi: 10.1006/jpho.2001.0151
- Escudero, P., Benders, T., and Lipski, S. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: the case of Dutch, German, and Spanish listeners. *J. Phon.* 37, 452–466. doi: 10.1016/j.wocn.2009.07.006
- Flege, J. (2003). "Assessing constraints on second-language segmental production and perception," in *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, eds A. Meyer and N. Schiller (Berlin: Mouton de Gruyter), 319–355.
- Flege, J. E., and Hillenbrand, J. (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *J. Acoust. Soc. Am.* 79, 508–517. doi: 10.1121/1.393538
- Flege, J., and Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Lang. Speech* 24, 125–146.
- Flemming, E. (1995). *Auditory Representation in Phonology*. Doctoral dissertation, University of California, Los Angeles, CA.
- Fletcher, J. (2010). "The prosody of speech: timing and rhythm," in *The Handbook of Phonetic Sciences*, 2nd Edn, eds W. J. Hardcastle, J. Laver, and F. E. Gibbon (Oxford: Blackwell), 523–602.
- Fougeron, C., and Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *J. Acoust. Soc. Am.* 101, 3728–3740. doi: 10.1121/1.418332
- Goldinger, S. D. (1996). Words and Voices: episodic traces in spoken word identification and recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1166–1183.
- Goldinger, S. D. (1997). "Words and voices: perception and production in an episodic lexicon," in *Talker Variability in Speech Processing*, eds K. Johnson and J. W. Mullennix (New York, NY: Academic Press), 33–56.

- Gussenhoven, C. (2007). "Intonation," in *The Cambridge Handbook of Phonology*, ed. P. de Lacy (Cambridge: Cambridge University Press), 253–280.
- Hawkins, J. A. (2014). "Patterns in competing motivations and the interaction of principles," in *Competing Motivations in Grammar and Usage*, eds B. MacWhinney, A. Malchukov, and E. Moravcsik (Oxford: Oxford University Press), 54–69.
- Iverson, P., and Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: auditory training for native Spanish and German speakers. *J. Acoust. Soc. Am.* 126, 866–877. doi: 10.1121/1.3148196
- Johnson, E. K. (2016). Constructing a proto-lexicon: an integrative view of infant language development. *Annu. Rev. Linguist.* 2, 391–412. doi: 10.1146/annurev-linguistics-011415-040616
- Keating, P. (1985). "Universal phonetics and the organization of grammars," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, ed. V. A. Fromkin (Orlando, FL: Academic Press), 115–132.
- Kent, R., and Read, C. (2002). *The Acoustic Analysis of Speech*. San Diego, CA: Singular Press.
- Kessinger, R. H., and Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *J. Phon.* 25, 143–168. doi: 10.1006/jpho.1996.0039
- Kewley-Port, D., and Watson, C. S. (1994). Formant-frequency discrimination for isolated English vowels. *J. Acoust. Soc. Am.* 95, 485–496. doi: 10.1121/1.408942
- Kim, S., Broersma, M., and Cho, T. (2012). The use of prosodic cues in learning new words in an unfamiliar language. *Stud. Second Lang. Acquisit.* 34, 415–444. doi: 10.1017/S0272263112000137
- Lindblom, B. (1990). "Explaining phonetic variation: a sketch of the H and H theory," in *Speech Production and Speech Modeling*, eds W. J. Hardcastle and A. Marchal (Dordrecht: Kluwer Academic), 403–440.
- Maddieson, I. (1997). "Phonetic universals," in *The Handbook of Phonetic Sciences*, eds W. J. Hardcastle and J. Laver (Oxford: Blackwell), 619–639.
- Moreton, E. (2004). Realization of the English postvocalic [voice] contrast in F1 and F2. *J. Phon.* 32, 1–33. doi: 10.1016/S0095-4470(03)00004-4
- Mücke, D., and Grice, M. (2014). The effect of focus marking on supralaryngeal articulation: is it mediated by accentuation? *J. Phon.* 44, 47–61. doi: 10.1016/j.wocn.2014.02.003
- Munro, M. J. (1995). Nonsegmental factors in foreign accent: ratings of filtered speech. *Stud. Second Lang. Acquisit.* 17, 17–34. doi: 10.1017/S0272263100013735
- Nakagawa, T., Saito, S., and Yoshino, T. (1982). Tonal difference limens for second formant frequencies of synthesized Japanese vowels. *Annu. Bull. R. Inst. Logoped. Phon.* 16, 81–88.
- Nguyen, N., and Hawkins, S. (2004). Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *J. Phon.* 32, 199–231. doi: 10.1016/S0095-4470(03)00031-7
- Pierrehumbert, J. B. (2001). "Exemplar dynamics: word frequency, lenition, and contrast," in *Frequency Effect and the Emergence of Lexical Structure*, eds J. Bybee and P. Hopper (Amsterdam: John Benjamins), 137–157.
- Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Lang. Speech* 46, 115–154. doi: 10.1177/00238309030460020501
- Pycha, A., and Dahan, D. (2016). Differences in coda voicing trigger changes in gestural timing: a test case from the American English diphthong /aɪ/. *J. Phon.* 56, 15–37. doi: 10.1016/j.wocn.2016.01.002
- Raphael, L. J. (2005). "Acoustic cues to the perception of segmental phonemes," in *Handbook of Speech Perception*, eds D. B. Pisoni and R. Remez (Oxford: Blackwell), 182–206.
- Shattuck-Hufnagel, S., and Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *J. Psycholinguist. Res.* 25, 193–247. doi: 10.1007/BF01708572
- Smiljanić, R., and Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *J. Phon.* 36, 91–113. doi: 10.1016/j.wocn.2007.02.002
- Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: articulatory and acoustic analyses. *J. Acoust. Soc. Am.* 82, 847–863. doi: 10.1121/1.395284
- Summers, W. V. (1988). F1 structure provides information for final-consonant voicing. *J. Acoust. Soc. Am.* 84, 485–492. doi: 10.1121/1.396826
- Thomas, E. R. (2000). Spectral differences in /aɪ/ offsets conditioned by voicing of the following consonant. *J. Phon.* 28, 1–25. doi: 10.1006/jpho.2000.0103
- Trofimovich, P., and Baker, W. (2006). Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *Stud. Second Lang. Acquisition* 28, 1–30. doi: 10.1017/S0272263106060013
- Tyler, M. D., and Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *J. Acoust. Soc. Am.* 126, 367–376. doi: 10.1121/1.3129127
- Wolf, C. G. (1978). Voicing cues in English final stops. *J. Phon.* 6, 299–309.
- Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *J. Phon.* 24, 245–261. doi: 10.1006/jpho.1996.0013

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Choi, Kim and Cho. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Influence of the *Pinyin* and *Zhuyin* Writing Systems on the Acquisition of Mandarin Word Forms by Native English Speakers

Rachel Hayes-Harb * and Hui-Wen Cheng

Department of Linguistics, University of Utah, Salt Lake City, UT, USA

OPEN ACCESS

Edited by:

Annie Tremblay,
University of Kansas, USA

Reviewed by:

Karen Elizabeth Mulak,
University of Western Sydney,
Australia
Peggy Mok,
The Chinese University of Hong Kong,
China

Min Wang,
University of Maryland, USA

*Correspondence:

Rachel Hayes-Harb
r.hayes-harb@utah.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 31 December 2015

Accepted: 10 May 2016

Published: 03 June 2016

Citation:

Hayes-Harb R and Cheng H-W (2016)
The Influence of the *Pinyin* and *Zhuyin*
Writing Systems on the Acquisition of
Mandarin Word Forms by Native
English Speakers.
Front. Psychol. 7:785.
doi: 10.3389/fpsyg.2016.00785

The role of written input in second language (L2) phonological and lexical acquisition has received increased attention in recent years. Here we investigated the influence of two factors that may moderate the influence of orthography on L2 word form learning: (i) whether the writing system is shared by the native language and the L2, and (ii) if the writing system is shared, whether the relevant grapheme-phoneme correspondences are also shared. The acquisition of Mandarin via the *Pinyin* and *Zhuyin* writing systems provides an ecologically valid opportunity to explore these factors. We first asked whether there is a difference in native English speakers' ability to learn *Pinyin* and *Zhuyin* grapheme-phoneme correspondences. In Experiment 1, native English speakers assigned to either *Pinyin* or *Zhuyin* groups were exposed to Mandarin words belonging to one of two conditions: in the "congruent" condition, the *Pinyin* forms are possible English spellings for the auditory words (e.g., <nai> for [nai]); in the "incongruent" condition, the *Pinyin* forms involve a familiar grapheme representing a novel phoneme (e.g., <xio> for [ciou]). At test, participants were asked to indicate whether auditory and written forms matched; in the crucial trials, the written forms from training (e.g., <xio>) were paired with possible English pronunciations of the *Pinyin* written forms (e.g., [ziou]). Experiment 2 was identical to Experiment 1 except that participants additionally saw pictures depicting word meanings during the exposure phase, and at test were asked to match auditory forms with the pictures. In both experiments the *Zhuyin* group outperformed the *Pinyin* group due to the *Pinyin* group's difficulty with "incongruent" items. A third experiment confirmed that the groups did not differ in their ability to perceptually distinguish the relevant Mandarin consonants (e.g., [c] from the foils (e.g., [z]), suggesting that the findings of Experiments 1 and 2 can be attributed to the effects of orthographic input. We thus conclude that despite the familiarity of *Pinyin* graphemes to native English speakers, the need to suppress native language grapheme-phoneme correspondences in favor of new ones can lead to less target-like knowledge of newly learned words' forms than does learning *Zhuyin*'s entirely novel graphemes.

Keywords: second language acquisition (SLA), mandarin, *Pinyin*, *Zhuyin*, orthographic input, second language phonology, second language word learning

INTRODUCTION

Adult second language (L2) learners can exploit the availability of orthographic input in learning the phonological forms of L2 words (e.g., Escudero et al., 2008). However, we have also seen that there are limits to the utility of orthographic input in supporting learners' target-like acquisition of words' forms—the literature provides cases where written input either had no beneficial effect (Simon et al., 2010; Hayes-Harb and Hacking, 2015; Showalter and Hayes-Harb, 2015) or in fact interfered with the target-like acquisition of L2 word forms (e.g., Hayes-Harb et al., 2010; Young-Scholten and Langer, 2015). Two factors that have emerged as possibly associated with whether or not orthographic input supports or interferes with word form learning are (i) whether the writing system is shared by the native language and the L2, and (ii) if the writing system is shared, whether the relevant grapheme-phoneme correspondences are also shared. The case of native English speakers learning Mandarin via the *Zhuyin* and *Pinyin* writing systems provides an ecologically valid opportunity to explore the relative impact of these two factors on L2 word form learning. *Pinyin* uses the Roman alphabet, shared with English, while *Zhuyin* uses an entirely different set of graphemes. Each writing system poses its own set of challenges to native English learners: *Zhuyin* requires learners to acquire an entirely novel grapheme set; *Pinyin*, on the other hand, involves only familiar graphemes, but learners must suppress a number of English grapheme-phoneme correspondences in favor of new ones (e.g., in Mandarin, *Pinyin* <x>¹ maps to /ɕ/). In the present study we explored the consequences of these characteristics of *Pinyin* and *Zhuyin* for native English speakers' ability to learn the phonological forms of a set of Mandarin words, with the goal of elucidating the relative difficulty associated with each writing system.

ORTHOGRAPHIC INPUT AND L2 PHONOLOGICAL ACQUISITION

The role of orthographic input in L2 phonological and lexical acquisition has received increased attention in recent years (Bassetti, 2008; Bassetti et al., 2015). While a number of studies have demonstrated a facilitative effect of orthographic input for L2 learners (e.g., Escudero et al., 2008; Showalter and Hayes-Harb, 2013), others have found limited or no effect of orthographic input (e.g., Simon et al., 2010; Pytlyk, 2011; Escudero, 2015; Hayes-Harb and Hacking, 2015; Showalter and Hayes-Harb, 2015). Indeed, there are also circumstances where orthographic input can interfere with L2 phonological and lexical acquisition (Bassetti, 2006; Escudero and Wanrooij, 2010; Hayes-Harb et al., 2010; Mathieu, 2016). We begin by reviewing studies on the influence of orthographic input in L2 word form learning, followed by a discussion of the small number of studies that have considered the influence of *Zhuyin* and *Pinyin* input on L2 Mandarin phonological and lexical acquisition.

¹“<x>” denotes a written form.

Orthographic Input and L2 Word Form Learning

In cases where orthographic input facilitates L2 word form learning, learners may benefit from familiarity of the graphemes in addition to familiarity of the grapheme-phoneme correspondences. For example, native Dutch speakers who saw written forms during an English word learning task (e.g., <tandek> and <tenzer>) were more likely to have established lexical representations that distinguish between English /æ/ and /e/ (corresponding to the letters <a> and <e>) than those who did not have access to written forms (Escudero et al., 2008). In this case, the L2 English graphemes were familiar to the native Dutch learners, and additionally, while the particular grapheme-vowel mappings differ between Dutch and English, the graphemes <a> and <e> capture a phonological contrast in both languages, presumably allowing participants to infer the English phonological contrast from the differential spellings.

More recent studies, however, have provided evidence of the limitations of written input in facilitating second language word learning. For example, a number of studies have found no effect of orthographic input in some cases where the graphemes and/or grapheme-phoneme correspondences are unfamiliar (e.g., Simon et al., 2010; Showalter and Hayes-Harb, 2013). Others have even found detrimental effects when the grapheme-phoneme correspondences of the L1 and L2 differ (Young-Scholten, 2002; Hayes-Harb et al., 2010; Hayes-Harb et al., submitted), or when the orthography is entirely unfamiliar (e.g., Mathieu, 2016). For example, Hayes-Harb et al. (2010) demonstrated that using a familiar orthography with unfamiliar grapheme-phoneme correspondences can lead learners to misremember the phonological forms of newly learned words. In this study, native English speakers were taught a set of auditory English non-words along with pictured meanings, and were later tested on their ability to match auditory forms to the pictures. In the “congruent” condition, participants always saw written forms (presented immediately below the picture) that were spelled according to English grapheme-phoneme correspondences (e.g., the auditory form [faza] was accompanied by the written form <faza>). In the “incongruent” condition, participants saw some written forms did not conform to English grapheme-phoneme correspondences (e.g., the auditory form [faza] was accompanied by <fasha>). In the control condition, participants saw <xxx> instead of written forms. At test, participants in the incongruent condition were more likely than participants in the other two conditions to misremember the phonological forms of the words in ways that reflected the (incongruent) spellings (e.g., accept [faʃa] as a possible pronunciation of the word [faza]). In this way, the incongruent spellings of the newly learned words appear to have interfered with participants' ability to correctly remember the words' phonological forms at test. Hayes-Harb et al. (submitted), following up on earlier studies such as those of Young-Scholten (2002) and Young-Scholten and Langer (2015), demonstrated that access to spelled forms in the L2 input can interfere with native English speakers' acquisition of German final obstruent devoicing. Hayes-Harb et al. (submitted) taught native English speakers German nonwords in two conditions:

in one condition, participants saw spelled forms (e.g., hear [krat]; see <krad>); in the other condition, participants did not see spelled forms. At test, participants who had seen <krad> during the word learning phase were more likely than those in the no spelled forms condition to pronounce it as [krad]. They conclude that in cases where auditory forms and written forms conflict, inferences about the pronunciation of words from written input may override the auditory input. Escudero et al. (2014) provide additional evidence that “congruency” between the grapheme-phoneme correspondences of the L1 and L2 influence the effect of written input on L2 word form learning. They taught native Spanish speakers auditory Dutch nonwords and pictured meanings in two conditions (one with and one without written forms), and later tested them on their ability to distinguish between minimal pairs of test words. In this study, “congruency” was defined somewhat differently than in other studies mentioned here in that it related to whether or not a graphemic contrast signals a phonemic contrast in both the L1 and L2, not to the grapheme-phoneme correspondences themselves. Some pairs of test words were “congruent” in the sense that the corresponding orthographic forms signal a phonemic contrast in both Spanish and English (e.g., Dutch <i> – <uu> = /i/ – /y/ and Spanish <i> – <u> = /i/ – /u/), while others were “incongruent” in that the orthographic forms signal a phonemic contrast in Dutch but not in Spanish (e.g., Dutch: <u> – <uu> = /Y/ – /y/; Spanish: <u> = /u/). The native Spanish participants performed more accurately at test on congruent than incongruent items. Escudero et al. (2014) thus found further evidence that L2 learners experience a benefit associated with congruency between the L1 and L2 writing systems when learning new words.

Why should auditory and orthographic input interact in these ways in second language word learning? The influence of orthography on spoken word recognition is well documented. For example, Ziegler and Ferrand (1998) demonstrated that native French speakers respond faster in an *auditory* lexical decision task to words whose rimes have a single possible spelling (e.g., <age> for for the rime /aʒ/) than to words whose rimes can be spelled variously (e.g., <omb> or <om> for the rime /om/). In addition, the effect of orthography on phonological processing begins in childhood along with early literacy. For example, Racine et al. (2014) found that native French readers (9–10 years old) show evidence for the influence of words’ spelled forms on their auditory processing of French production variants resulting from schwa deletion, while native French pre-readers (5–6 year olds) do not.

As noted by Veivo and Jarvikivi (2013), a “consequence of many L2 learners being literate is that the teaching and the learning of L2 are often based on written language to a significant degree” (p. 866). Thus L2 learners’ (alphabetic) literacy, presumably including their knowledge of specific grapheme-phoneme correspondences and/or the expectation that written input will provide phonologically relevant information about the forms of L2 words, may exert an influence beginning with their earliest exposure to L2 words. In light of the vast literature documenting learners’ propensity for transferring aspects of their L1 into L2 acquisition (see, e.g., Eckman and Iverson, 2013), it

may be unsurprising that learners appear to transfer their L1 grapheme-phoneme correspondences to L2 learning.

Relative to the number of studies that have considered the impact of orthographic congruency on L2 word form learning, very few have investigated the effect of unfamiliar orthographies. Hayes-Harb and Hacking (2015) investigated the influence of diacritic stress marks on Russian written words on native English speakers’ ability to learn Russian lexical stress. They secondarily asked whether the effect of stress marks differed depending on whether the words were written in Cyrillic or Roman letters. They found no beneficial effect of the diacritic stress marks, and no difference in performance associated with the Cyrillic vs. Roman letter condition, suggesting at a minimum that the familiarity of the graphemes did not influence word form learning. Showalter and Hayes-Harb (2015) similarly did not find a difference in word learning performance between groups of naïve native English speakers exposed to Arabic vs. Roman written forms when learning Arabic words minimally distinguished by the difficult velar /k/—uvular /q/ contrast. While these two studies do not indicate a word learning disadvantage associated with novel orthographies vs. familiar ones, it is worth noting that the measure of learning in both of these studies involved perceptually discriminating difficult novel phonological contrasts. In the present study, we focus not on the role of orthographic input in learners’ ability to differentiate words containing difficult novel contrasts, but rather on the issues of orthographic congruency and familiarity and their effects on L2 word form learning.

In summary, the growing literature on the influence of orthographic input in L2 word form learning has highlighted two factors that may be associated with whether written input supports or interferes with word form learning: (1) whether the writing system is shared by the native language and the L2, and (2) if the writing system is shared, whether the relevant grapheme-phoneme correspondences are shared by the two languages. The acquisition of L2 Mandarin provides an opportunity to explore these factors, given that the *Pinyin* writing system involves familiar graphemes with a number of novel grapheme-phoneme correspondences, and *Zhuyin* involves an entirely new set of graphemes. The following section reviews the small number of studies that have considered the influence of these two writing systems on L2 Mandarin acquisition.

Orthographic Input and the Acquisition of L2 Mandarin

Chinese characters are known for their opacity in terms of grapheme-phoneme correspondences. Indeed, the phonetic component of a Chinese character provides reliable cues to the pronunciation of the character <30% of the time (Cheng, 2012). To facilitate the learning of Chinese characters, a phonetic script that transparently presents the phonological forms of Chinese words is usually introduced to beginning learners (including both L1 and L2 learners). *Pinyin* and *Zhuyin* are the scripts that are most commonly used for this purpose. *Pinyin* (formally known as *Hanyu Pinyin*) is a Romanization system used in China and Singapore, and has been adopted by the International Organization for Standardization for the

Romanization of Chinese ISO (2015). *Zhuyin* (also called *Zhuyin fuhao* or *Bopomofo*) consists of components of ancient Chinese characters, and is widely used in Taiwan. Crucially for the present purposes, while both are transparent phonographic writing systems, *Pinyin* and *Zhuyin* differ from one another in the graphemes they employ. There are also organizational differences between *Pinyin*, which is an alphabet, and *Zhuyin*, which is a semi-syllabary, or a combination of an alphabet and a syllabary; (Taylor and Taylor, 2014). Here we explore the differential effects of the two writing systems on the acquisition of Mandarin word forms by native English speakers. In particular, we ask whether the orthographic differences between *Pinyin* and *Zhuyin* influence Mandarin word learning. This question is particularly intriguing in the context of adult L2 Mandarin acquisition, because these learners are equipped with the knowledge of their L1 writing system, which may interact with the characteristics of *Pinyin* and *Zhuyin*. For example, native English-speaking learners, whose L1 employs the Roman alphabet, may find *Pinyin* less difficult than *Zhuyin* initially given the familiarity of the *Pinyin* symbols, which form a subset of the English alphabet. However, for a subset of *Pinyin* graphemes, the grapheme-phoneme correspondences differ from those of English. For example, in Chinese the *Pinyin* grapheme <x> maps to the voiceless alveopalatal fricative /ç/, a phoneme that does not exist in English; the same grapheme maps to /ks/ as in “tax” or /z/ as in “xylophone” in English. Thus native English speakers learning Mandarin who are exposed to *Pinyin* may benefit from the familiarity of the graphemes but experience difficulty learning novel grapheme-phoneme correspondences. In other words, native English speakers may show evidence of the negative transfer of English grapheme-phoneme correspondences when learning Mandarin with *Pinyin*.

On the other hand, no such opportunity for negative transfer is associated with *Zhuyin*, whose graphemes do not overlap with English graphemes. For instance, the voiceless alveolar affricate /ts/, which is written <z> in *Pinyin*, is written <ㄗ> in *Zhuyin*. **Table 1** provides example *Zhuyin* and *Pinyin* graphemes, along with their corresponding phonemes. *Zhuyin*, however, presents its own challenge for native English speakers—that of learning a new set of graphemes. At present we are interested in the relative difficulty associated with learning new graphemes vs. learning new grapheme-phoneme correspondences on native English speakers’ ability to learn the phonological forms of new Mandarin words.

A small number of studies have specifically investigated the influence of orthographic input on the acquisition of Mandarin by native English speakers (Bassetti, 2006; Pytlyk, 2011; Showalter and Hayes-Harb, 2013). Bassetti (2006) and Pytlyk (2011) specifically explored the acquisition of Mandarin by native English speakers via the written medium of *Pinyin*, focusing on the potential for interference due to the negative transfer of native English grapheme-phoneme correspondences. Bassetti (2006) investigated whether *Pinyin* spelling conventions for rimes influences native English speakers’ Mandarin phonological representations, focusing on the confusion they may cause for native English speakers with respect to the number of segments contained in the rimes. In *Pinyin*, rimes may be spelled differently

TABLE 1 | Example *Pinyin* and *Zhuyin* graphemes and their corresponding Mandarin phonemes.

<i>Pinyin</i>	<i>Zhuyin</i>	Corresponding mandarin phoneme
n	ㄋ	Alveolar nasal /n/
s	ㄙ	Voiceless alveolar fricative /s/
l	ㄌ	Alveolar lateral /l/
m	ㄇ	Bilabial nasal /m/
z	ㄗ	Voiceless dental affricate /ts/
c	ㄘ	Voiceless aspirated dental affricate /ts ^h /
q	ㄑ	Voiceless aspirated alveopalatal affricate /tç ^h /
x	ㄒ	Voiceless alveopalatal fricative /ç/

depending on their context, in particular with respect to the inclusion of a letter representing what is called the “main vowel.” For example, following a consonantal onset, the rime /uei/ is spelled <ui> (without a letter corresponding to the main vowel /e/), as in <kui>. The same rime is spelled <wei> (with the letter <e> representing the main vowel) in onsetless syllables. Bassetti (2006) asked native English speakers who were beginning learners of Mandarin to perform two phonological tasks. In the phoneme counting task, participants were asked to read (logographic) Chinese characters and to count the number of “sounds” in each. In the phoneme segmentation task, participants were asked to pronounce the characters’ sounds one-by-one. Bassetti found that for syllables where the *Pinyin* spellings do not represent the main vowel, participants counted one fewer vowel in the rime than when the *Pinyin* spellings represent the main vowel. The segmentation task confirmed that the vowel omitted by learners was indeed the main vowel, or the one that is not represented in *Pinyin* spellings. Bassetti concluded that the native English speakers’ phonological representations for Chinese syllables was affected by the *Pinyin* spelling conventions with respect to main vowels.

Pytlyk (2011) investigated whether exposure to *Pinyin*, in particular in cases where English and Mandarin have different grapheme-phoneme correspondences, negatively influences native English speakers’ ability to perceive Mandarin consonants. Pytlyk predicted that while native English speakers may benefit from the positive transfer of knowledge of the Roman alphabet in learning Mandarin via *Pinyin* (a “shared” orthography), they may experience difficulty where the grapheme-phoneme correspondences of *Pinyin* and the English alphabet differ. Specifically, the prediction was that “learners who learn Mandarin via *Pinyin*...will tend to equate a similar Mandarin (L2) phoneme with its English counterpart because the shared orthographic symbols would make perceiving the differences between the similar sounds even more difficult” (p. 545). In contrast, it was predicted that learner groups who were exposed to *Zhuyin* or to no written forms at all would outperform the *Pinyin* learners in Mandarin consonant perception because neither of these groups would experience the orthographic interference associated with *Pinyin*. Native English speakers with no previous Chinese language experience participated in a language training phase followed by a perception test. During

the language training phase, they were taught the Mandarin phoneme inventory via *Pinyin*, *Zhuyin*, or no written input. At test, participants performed an oddity discrimination task, in which they heard three stimuli and were asked to determine which one differed from the other two. There were no significant differences in test performance among the participants trained via *Pinyin*, *Zhuyin*, or no written input. While Pytlyk (2011) did not find the predicted differences in perception performance, this study nonetheless highlights the utility of Mandarin and its *Pinyin* and *Zhuyin* writing systems for addressing questions concerning the role of orthographic transfer in second language phonological learning.

Research Questions

The Bassetti (2006) and Pytlyk (2011) studies investigated the influence of orthographic input on phonological representations of Mandarin syllables and on the ability of learners to perceive Mandarin phonological contrasts, respectively. In focus in the present work is the influence of orthographic input in early lexical-phonological development—specifically, the influence of *Pinyin* and *Zhuyin* on native English speakers' ability to accurately remember the phonological forms of newly learned Mandarin words. The broadest question guiding our research is thus: Is there a difference in the difficulty associated with learning the grapheme-phoneme correspondences for novel graphemes (as in *Zhuyin*) and learning new grapheme-phoneme correspondences for familiar graphemes (as in *Pinyin*)? The first research question that this study is designed to answer is whether there is a difference in native English speakers' ability to learn *Pinyin* vs. *Zhuyin* grapheme-phoneme correspondences, specifically whether native English speakers exposed to *Pinyin* experience particular difficulty with “incongruent” grapheme-phoneme correspondences. This question is addressed in Experiment 1. Our second research question is whether there is a difference in native English speakers' ability to learn the phonological forms of new words when exposed to *Pinyin* vs. *Zhuyin*, specifically whether native English speakers exposed to *Pinyin* experience particular difficulty learning the phonological forms of words with “incongruent” spellings (Experiment 2). Our final research question, addressed in Experiment 3, is whether participants exposed to *Pinyin* vs. *Zhuyin* differ in their ability to perceive Mandarin consonant contrasts.

EXPERIMENTS

This study was carried out with approval from the University of Utah Institutional Review Board and with written informed consent from all participants.

Participants

Thirty monolingual native English speakers were recruited from the University of Utah community and received course credit for participating in the study. All participated in all three experiments in the same order. A background questionnaire confirmed that none of the participants had previously studied Chinese, and none reported speech, language, hearing, or neurological disorders. The participants were randomly assigned

to the *Pinyin* group or the *Zhuyin* group ($n = 15$ each). Each group consisted of 5 males and 10 females. The mean age in the *Pinyin* group was 23.7 years old ($SD = 4.7$), and the mean age in the *Zhuyin* group was 25.7 years old ($SD = 8.7$). Participants assigned to the *Pinyin* group reported experience with Spanish (8 participants), Japanese (2), French (2), and one each with Arabic, Latin, Korean, German, Modern Greek, Samoan, Turkish, and Swahili; two participants reported no second language experience. Participants in the *Zhuyin* group reported experience with Spanish (12), French (3), and one each had experience with Russian, Armenian, ASL, German, or Italian; two reported no second language experience.

Materials

For the purposes of the study, we developed a set of 16 Mandarin syllables (“words”), along with their written forms in *Pinyin* and *Zhuyin* and randomly-assigned line-drawing visual referents (i.e., the words’ “meanings”). The words belonged to two conditions: congruent and incongruent. In the congruent condition, the *Pinyin* forms are possible English spellings for the auditory words (e.g., <nai> for [nai]); in the incongruent condition, the *Pinyin* forms involve a familiar (English) grapheme representing a novel (Mandarin) consonant (e.g., the <x> in <xiu> for [ciou]). It is important to note that words are categorized as congruent and incongruent on the basis of their *Pinyin* spellings only—the novel *Zhuyin* graphemes are neither congruent nor incongruent from the point of view of participants. To determine the use of *Pinyin* graphemes in the congruent vs. incongruent word conditions, we first conducted a norming study. In this study, 10 native English speakers (who did not participate in the three experiments) were asked to use English graphemes to transcribe the initial consonants in 105 aurally-presented Mandarin CV syllables. The syllables were produced by a male Mandarin-English bilingual speaker reading from *Pinyin* transcriptions. Following a brief practice session using English nonwords to familiarize them with the task, the native English speakers were asked to respond to the entire block of 105 syllables, presented twice and in a different random order each time.

We calculated the percentage of participants' English letter responses that matched the *Pinyin* letters used to transcribe the initial consonants in Mandarin. For example, the auditory syllable /lin/, which is spelled with an initial <l> in *Pinyin*, was always transcribed by the native English participants with an initial <l>, and thus received a “match” score of 100%. On the other hand, the initial consonant in [tɕ^{hi}ie], transcribed as <q> in *Pinyin*, was transcribed by the native English participants as <ch, C, sh, t, ts>, but never as <q>, and thus received a match score of 0%. The four graphemes that received the highest match scores were selected for use in the congruent condition: <l> (100%), <m> (100%), <s> (98%), and <n> (96%). The four receiving the lowest match scores were selected for use in the incongruent condition: <c> (0%), <q> (0%), <x> (0%), and <z> (13%). [Note: Although *Pinyin* <zh> also had a low match score (5%), its corresponding Mandarin consonant phoneme had a similar response profile to <q>, indicating that the Mandarin phonemes represented by <zh> and <q> are

potentially confusable by native English speakers. For this reason, we excluded <zh> from the study materials].

We next created 16 Mandarin syllables using the Mandarin phonemes represented by the congruent and incongruent graphemes that were selected via the norming study. To control for lexical tone, all word stimuli were produced in Tone 4 (high-falling); in this tone, some of the words were actual words in Mandarin and others were nonwords; all are referred to here as “words” since our participants were unfamiliar with Mandarin. Due to restrictions on vowel distributions in Mandarin, words with initial <z, c, n, s> (/ts, ts^h, n, s/, respectively) contained the rimes <ai> or <ao> (/ai/ or /au/), and those with initial <q, x, l, m> (/tɕ^h, ɕ, l, m/) contained the rimes <ie> or <iu> (/iɛ/ or /iou/). Each of the eight initial consonants was combined with its two corresponding rimes to create the 16 Mandarin words; a full list of the words is provided in **Table A1** in Appendix. These words served as the stimuli presented in the exposure, criterion, and test phases of the three experiments described below.

In addition, we created a set of 16 foil words for use in the test phases. For the incongruent condition, we chose the phonemes that the incongruent *Pinyin* graphemes usually represent in English to serve as foils. The foil phoneme for <z> and <x> is thus /z/, and the foil phoneme for <c> and <q> is thus /k/ (Note: as there is no /z/ phoneme in Mandarin, some of the words used in the study are in fact impossible in Mandarin; as a whole, the stimulus set is thus quasi-Mandarin). The foil phonemes for the congruent graphemes were selected randomly: /d/ for <n> and <l>, and /ə/ for <s> and <m>. **Table 2** summarizes the construction of the Mandarin words and their foils. Words are categorized as congruent and incongruent on the basis of their *Pinyin* spellings only, given that the native English speakers who participated in the present experiments do not have existing grapheme-phoneme correspondences for the (unfamiliar) *Zhuyin* graphemes.

Each of the 16 words was randomly assigned a “meaning” from among a set of nonobject line drawings; the word-meaning pairings were the same for all participants. The words were produced by a male Mandarin-English bilingual speaker reading from *Pinyin* transcriptions.

Experiment 1 (Grapheme-Phoneme Correspondence Learning) Procedures

In Experiment 1, we exposed participants to the set of auditory Mandarin words and their written forms, and later tested them on their ability to accurately determine whether the auditory and written forms were correctly matched. The experiment involved three phases: exposure, criterion, and test. All three experiments were conducted in a sound-attenuated booth; the entire session lasted ~ 1 h, with brief participant-controlled breaks between experiments.

Exposure Phase

Participants were asked to learn the 16 words. In each exposure trial, a written form was presented on the computer screen while the auditory word was played over headphones at a comfortable listening level. The written form remained on the screen for 2 s, followed by the next trial. The 16 words constituted one block,

and there were four blocks in the exposure phase. see **Table 3** for example exposure phase trials.

Criterion Phase

The criterion phase consisted of 16 matched and 16 mismatched trials. Participants were asked to indicate whether a written word matched the auditory word by pressing “yes” or “no” buttons on the keyboard. **Table 3** illustrates example criterion phase trials. Congruent-matched and congruent-mismatched trials were expected to be easy for participants (e.g., see pinyin/zhuyin written form for [nai] and hear [nai] (matched) or [ts^hai] (mismatched)). In incongruent-matched trials, participants saw a written form and heard its corresponding auditory word (e.g., see the pinyin/zhuyin written form for the word [ciou] and hear [ciou]). In the incongruent-mismatched trials, as in the congruent-mismatched trials, participants saw a written form and heard a word beginning with an entirely different consonant that was also not the foil (e.g., see the pinyin/zhuyin written form for the word [ciou] but hear [miou]). In this way, no criterion phase trials were designed to be difficult for participants in either exposure condition; rather, the criterion phase was used to ensure that all participants achieved a similar level of ability to distinguish learned forms from quite different foils before continuing to the test phase. Participants repeated the exposure and criterion phases until they reached 90% accuracy on the criterion test.

Test Phase

The test phase was identical to the criterion phase except that the test phase was designed to determine whether participants experienced confusion due to differences between Pinyin and English grapheme-phoneme correspondences. Congruent-matched and congruent-mismatched trials were again expected to be easy for participants (e.g., see Pinyin/Zhuyin written form for [nai] and hear [nai] (matched) or [dai] (mismatched)), as were the incongruent-matched trials (e.g., see the Pinyin/Zhuyin written form for the word [ciou] and hear [ciou]). However, the incongruent-mismatched trials were designed to be difficult for participants in the Pinyin condition if they experienced interference from English grapheme-phoneme correspondences. In these trials, participants saw a written form and heard a word beginning with a consonant reflecting English grapheme-phoneme correspondences (e.g., see the Pinyin/Zhuyin written form for the word [ciou], which is spelled <xu> in Pinyin, but hear [ziou], a possible English pronunciation of the Pinyin written form). See **Table 3** for an illustration of test phase trials.

Results

The first analysis concerns the number of exposure-criterion phase cycles that participants required to reach the criterion necessary to continue to the final test. Participants in the *Pinyin* group (mean = 1.6 cycles; SD = 0.632) required significantly fewer cycles than did participants in the *Zhuyin* group [mean = 3.47; SD = 1.807; $F_{(1, 28)} = 14.255$, $p = 0.001$, *partial* $\eta^2 = 0.337$].

We converted the final test phase accuracy data (see **Table 4**) to d-primes using Signal Detection Theory (see **Figure 1**; for

TABLE 2 | The Pinyin and Zhuyin graphemes used in the study, along with the foil phonemes assigned to each grapheme and the vowels added to create the Mandarin word stimuli.

	Grapheme		Corresponding phoneme(s)		Foil phoneme	Vowels added to create word stimuli		
	Pinyin	Zhuyin	Mandarin	English		Vowel	Pinyin	Zhuyin
Congruent items	n	ㄋ	/n/	/n/	/d/	/aɪ/ /aʊ/	ai ao	ㄞ ㄠ
	s	ㄙ	/s/	/s/	/θ/			
	l	ㄌ	/l/	/l/	/d/	/iɛ/ /iou/	ie iu	ㄟ ㄢ
	m	ㄇ	/m/	/m/	/θ/			
Incongruent items	z	ㄗ	/ts/	/z/	/z/	/aɪ/ /aʊ/	ai ao	ㄞ ㄠ
	c	ㄘ	/ts ^h /	/k/, /s/	/k/			
	q	ㄑ	/tɕ ^h /	/k/	/k/	/iɛ/ /iou/	ie iu	ㄟ ㄢ
	x	ㄒ	/ɕ/	/z/, /ks/	/z/			

TABLE 3 | Experiment 1 example stimuli, by phase.

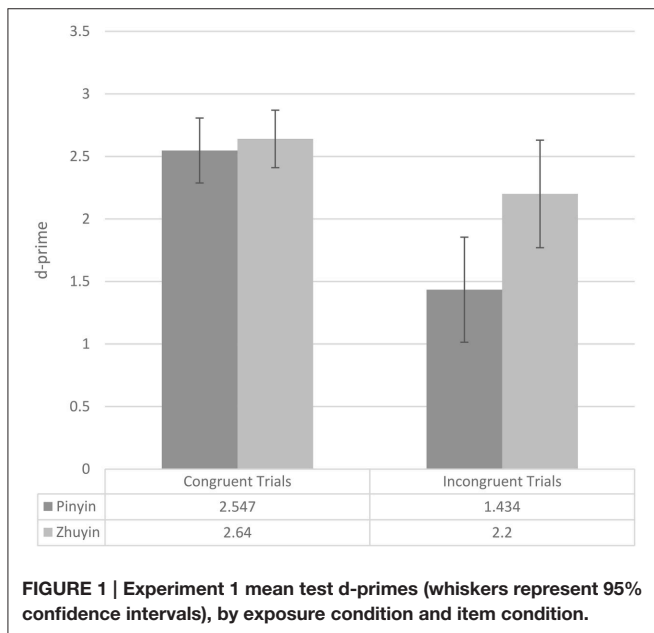
EXPOSURE PHASE						
Exposure condition	Example congruent trials			Example incongruent trials		
	See	Hear		See	Hear	
Pinyin	nai			xiu		
Zhuyin	ㄋ ㄞ	[nai]		ㄒ ㄢ	[ciou]	
CRITERION TEST PHASE						
Exposure condition	Example congruent trials			Example incongruent trials		
	See	Matched hear	Mismatched hear	See	Matched hear	Mismatched hear
Pinyin	nai			xiu		
Zhuyin	ㄋ ㄞ	[nai]	[ts ^h ai]	ㄒ ㄢ	[ciou]	[miou]
FINAL TEST PHASE						
Exposure condition	Example congruent trials			Example incongruent trials		
	See	Matched hear	Mismatched hear	See	Matched hear	Mismatched hear
Pinyin	nai			xiu		
Zhuyin	ㄋ ㄞ	[nai]	[dai]	ㄒ ㄢ	[ciou]	[ziou]

more information about d-prime, please see MacMillan and Creelman, 2004). The d-primes were submitted to ANOVA with exposure condition (two levels: *Pinyin*, *Zhuyin*) as a between-subjects variable and item condition (congruent, incongruent) as a within-subjects variable. There was a main effect of exposure group, with participants in the *Zhuyin* group performing more accurately than participants in the *Pinyin* group overall [$F_{(1, 28)} = 4.275$, $p = 0.048$, $partial \eta^2 = 0.132$], a main effect of item

type, with higher d-primes on congruent than incongruent items [$F_{(1, 28)} = 32.027$, $p < 0.0005$, $partial \eta^2 = 0.534$], and an interaction of the two [$F_{(1, 28)} = 5.991$, $p = 0.021$, $partial \eta^2 = 0.176$]. Following up on the interaction, we looked at the effect of exposure condition in the two item conditions separately. On congruent items, there was no effect of exposure condition [$F_{(1, 28)} = 0.284$, $p = 0.598$, $partial \eta^2 = 0.010$]. However, on incongruent items, the effect of exposure condition

TABLE 4 | Experiment 1 test accuracy (proportion correct responses; 95% confidence intervals in parentheses), by exposure condition and item condition.

Exposure condition	Congruent trials		Incongruent trials	
	Matched	Mismatched	Matched	Mismatched
Pinyin	0.975 (0.03)	0.867 (0.07)	0.933 (0.06)	0.533 (0.12)
Zhuyin	0.942 (0.04)	0.925 (0.05)	0.858 (0.05)	0.867 (0.10)



was significant [$F_{(1, 28)} = 6.277, p = 0.018, \text{partial } \eta^2 = 0.183$], with participants in the *Zhuyin* condition outperforming those in the *Pinyin* condition².

Experiment 2 (Word Learning) Procedures

Our second research question concerned whether there is a difference in native English speakers' ability to learn the phonological forms of new words when exposed to *Pinyin* vs. *Zhuyin*. Experiment 2 was identical to Experiment 1 except that participants additionally saw line drawings depicting word meanings during the exposure phase, and at test were asked to match auditory forms with the line drawings.

Exposure Phase

Participants were asked to learn the 16 auditory words and their pictured meanings. For each word, a written word form, a picture representing the word meaning, and an auditory word were presented simultaneously and stayed on the screen for 4 s,

²At the suggestion of an anonymous reviewer, we explored whether the English word status of the letter sequence <lie> (relative to the English nonword status of all other words' spelled forms) may have impacted performance on items involving <lie>. Analysis of proportion correct scores among Pinyin group participants in Experiment 1 indicates that <lie> items elicited accuracy within the range of that of the other items.

followed by the next trial. The 16 words constituted one block, and there were four blocks in the exposure phase.

Criterion Phase

The criterion phase trials were identical to those in Experiment 1 except that instead of matching auditory words to written forms, participants were asked to determine the accuracy of the match between auditory words and pictures. Again, congruent-matched and congruent-mismatched trials were expected to be easy for participants (e.g., see the picture associated with the auditory word [nai] and hear [nai] (matched) or [ts^hai] (mismatched)). In incongruent-matched trials, e.g., participants saw the picture associated with the auditory word [ciou] and heard [ciou]. In the incongruent-mismatched trials, e.g., participants saw the picture associated with the auditory word [ciou] but heard [miou]. Participants repeated the exposure and criterion phases until they reached 90% accuracy on the criterion test.

Test Phase

The test phase trials were identical to those in Experiment 1, again with the exception that participants' task was to determine the accuracy of the match between auditory words and pictures. Congruent-matched and congruent-mismatched trials involved, e.g., seeing the picture associated with [nai] and hearing [nai] (matched) or [dai] (mismatched). Incongruent-matched trials involved, e.g., seeing the picture associated with [ciou] and hearing [ciou]. In incongruent-mismatched trials, participants saw a picture associated with, e.g., [ciou], but heard [ziou], a possible English pronunciation of the pinyin written form <xiu>. Table 5 illustrates the stimuli encountered during the exposure, criterion, and final test phases in Experiment 2.

Results

Again, we first consider the number of exposure-criterion phase cycles participants in the two exposure conditions required. In this experiment, participants in the *Pinyin* group (mean = 2.53; SD = 0.834) required on average more cycles than did participants in the *Zhuyin* group (mean = 2.00; SD = 0.655); however, this difference was only marginally significant [$F_{(1, 28)} = 3.797, p = 0.061, \text{partial } \eta^2 = 0.119$]. Table 6 presents the final test accuracy data and Figure 2 the d-primes. The d-primes were submitted to ANOVA with exposure condition (two levels: *Pinyin*, *Zhuyin*) as a between-subjects variable and item condition (congruent, incongruent) as a within-subjects variable. There was a main effect of exposure group, with participants in the *Zhuyin* group performing more accurately than participants in the *Pinyin* group overall [$F_{(1, 28)} = 14.410, p = 0.001, \text{partial } \eta^2 = 0.340$], a main effect of item type, with higher d-primes on congruent than incongruent items [$F_{(1, 28)} = 56.571, p < 0.0005, \text{partial } \eta^2 = 0.669$], and an interaction of the two [$F_{(1, 28)} = 2.362, p = 0.001, \text{partial } \eta^2 = 0.318$]. Following up on the interaction, we looked at the effect of exposure condition in the two item conditions separately. On congruent items, there was no effect of exposure condition [$F_{(1, 28)} = 1.688, p = 0.204, \text{partial } \eta^2 = 0.056$]. However, on incongruent items, the effect of exposure condition was significant [$F_{(1, 28)} = 32.027, p < 0.0005$,

TABLE 5 | Experiment 2 example stimuli, by phase.













EXPOSURE PHASE						
Exposure condition	Example congruent trials		Example incongruent trials			
	See	Hear	See	Hear		
Pinyin						
	nai	[nai]	xiu	[ciou]		
Zhuyin	 ㄋㄞ		 ㄒㄧㄡ			
CRITERION TEST PHASE						
Exposure condition	Example congruent trials			Example incongruent trials		
	See	Matched hear	Mismatched hear	See	Matched hear	Mismatched hear
Pinyin		[nai]	[ts ^h ai]		[ciou]	[miou]
Zhuyin						
FINAL TEST PHASE						
Exposure condition	Example congruent trials			Example incongruent trials		
	See	Matched hear	Mismatched hear	See	Matched hear	Mismatched hear
Pinyin		[nai]	[dai]		[ciou]	[ziou]
Zhuyin						

TABLE 6 | Experiment 2 test accuracy (proportion correct responses; 95% confidence intervals in parentheses), by exposure condition and item condition.

Exposure condition	Congruent trials		Incongruent trials	
	Matched	Mismatched	Matched	Mismatched
Pinyin	0.975 (0.03)	0.942 (0.06)	0.850 (0.06)	0.683 (0.12)
Zhuyin	0.967 (0.03)	0.992 (0.02)	0.900 (0.05)	0.925 (0.07)

partial $\eta^2 = 0.534$], with participants in the *Zhuyin* condition outperforming those in the *Pinyin* condition.

Experiment 3 (Consonant Discrimination) Procedures

The purpose of Experiment 3 was to determine whether the participants in the *Pinyin* group and those in the *Zhuyin* group differed in their ability to perceptually distinguish the consonants contained in the newly learned words from the foil consonants contained in the incongruent-mismatched trials. Because the foil

consonants (e.g., [z]) were sometimes phonetically similar to the relevant Mandarin consonants (e.g., [ʑ]), performance in the test phase may have been confounded by perceptual confusability, which would undermine our ability to attribute Experiments 1 and 2 performance to the influence of the written input. Experiment 3 involved 16 matched and 16 mismatched trials. In each trial, two auditory words were presented, and participants were asked to decide whether the two words that they heard were the same. In the matched trials, each of the 16 words was presented twice. In the mismatched trials, each of the 16 words was presented with its foil (from Experiments 1 and 2; see Table A1 in Appendix for each word's foil).

Results

In this final experiment, participants were tested on their ability to discriminate the Mandarin consonant contrasts. As seen in Table 7 and Figure 3, participants in both groups were near ceiling in their discrimination ability. The d-primes were submitted to ANOVA with exposure condition (two levels: *Pinyin*, *Zhuyin*) as a between-subjects variable and item type (congruent, incongruent) as a within-subjects variable. There was no significant main effect of either exposure condition [$F_{(1, 28)} =$

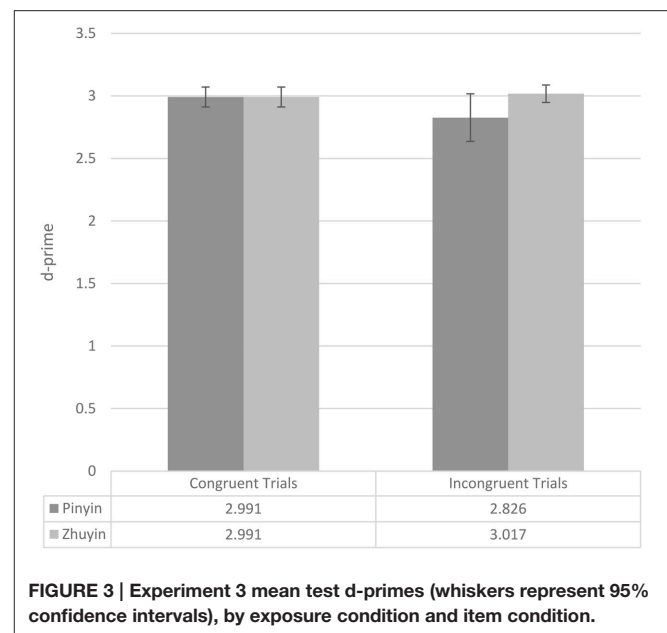
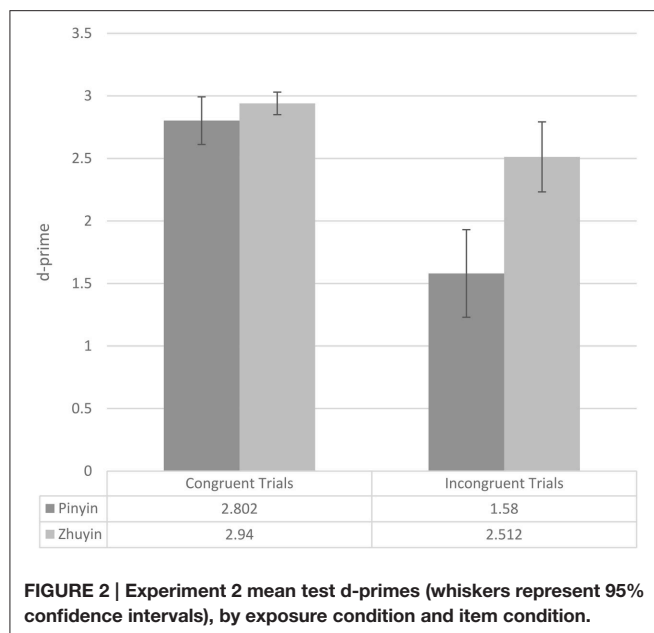


TABLE 7 | Experiment 3 accuracy (proportion correct responses; 95% confidence intervals in parentheses), by exposure condition and item condition.

Exposure condition	Congruent trials		Incongruent trials	
	Matched	Mismatched	Matched	Mismatched
Pinyin	0.975 (0.03)	1.000 (0.00)	0.992 (0.02)	0.933 (0.05)
Zhuyin	0.975 (0.03)	1.000 (0.00)	0.992 (0.02)	0.975 (0.03)

2.683, $p = 0.113$, $\text{partial } \eta^2 = 0.087$] or item condition [$F_{(1, 28)} = 1.357$, $p = 0.254$, $\text{partial } \eta^2 = 0.046$], and the interaction was also nonsignificant [$F_{(1, 28)} = 2.529$, $p = 0.123$, $\text{partial } \eta^2 = 0.083$]. Thus any differences in performance between the two groups on Experiments 1 and 2 is not attributable to differences in the two groups' perceptual sensitivities to the Mandarin consonant contrasts.

DISCUSSION

Recall that we first asked whether there is a difference in native English speakers' ability to learn *Pinyin* and *Zhuyin* grapheme-phoneme correspondences, specifically whether native English speakers exposed to *Pinyin* experience particular difficulty with "incongruent" grapheme-phoneme correspondences. Experiment 1 was designed to address this question. Analysis of the number of exposure-criterion phase cycles required to reach the 90% accuracy criterion indicates that participants exposed to *Zhuyin* required more than twice as many cycles as did those exposed to *Pinyin*. However, on the final test, those exposed to *Zhuyin* did not experience interference from English grapheme-phoneme correspondences on the "incongruent" items, as did

participants exposed to *Pinyin*. Thus while participants initially required more exposure to *Zhuyin* than to *Pinyin* to learn the grapheme-phoneme correspondences, they ultimately were able to avoid difficulty associated with the negative transfer of native language grapheme-phoneme correspondences.

We next asked whether there is a difference in native English speakers' ability to learn the phonological forms of new words when exposed to *Pinyin* vs. *Zhuyin*, specifically whether native English speakers exposed to *Pinyin* experience particular difficulty learning the phonological forms of words with "incongruent" spellings. In Experiment 2, which immediately followed Experiment 1, we examined the word learning ability of participants exposed to *Zhuyin* vs. *Pinyin* written forms. In the exposure phase of this experiment, participants heard auditory forms and saw pictures indicating the words' meanings. The pictures were accompanied by either the *Zhuyin* written form or the *Pinyin* written form. As in Experiment 1, we were interested in whether those in the *Pinyin* group would experience interference from English grapheme-phoneme correspondences on words in the incongruent condition. Indeed, at test, participants in the *Pinyin* group incorrectly accepted auditory forms reflecting English grapheme-phoneme correspondences (the foils) as the labels for newly learned words (e.g., they indicated that [ziou] was a correct pronunciation for a picture they had learned was pronounced [ɕiou], presumably due to its *Pinyin* spelling <xio>) significantly more often than did those in the *Zhuyin* group, while there was no difference in performance between groups on words in the congruent condition.

It is interesting to note, however, that in Experiment 2, the pattern with respect to the number of exposure-criterion phase cycles required by the two groups was opposite that observed in Experiment 1. In Experiment 2, the *Pinyin* group in fact required *more* exposure-criterion phase cycles than did the *Zhuyin* group, though this difference only approached significance at $p = 0.061$.

Thus the learning speed disadvantage experienced by *Zhuyin* participants in Experiment 1 (when learning grapheme-phoneme correspondences and not word meanings) did not persevere into the word learning experiment. One might intuitively anticipate initial difficulty associated with exposure to unfamiliar graphemes—indeed, in a similarly-structured study of native English speakers learning of Arabic words, Showalter and Hayes-Harb (2015) hypothesized that the unfamiliarity of the Arabic script and its conventions may have been responsible for low test accuracy levels. However, in a follow-up experiment, when the Arabic letters were replaced with Roman transliteration, they saw no increase in word learning accuracy, indicating that difficulty associated with the novel symbols was not fully responsible for the observed test difficulty. In another similarly-structured study, Hayes-Harb and Hacking (2015) did not find substantial differences in either number of exposure-criterion phase cycles or in final test accuracy between native English speakers exposed to Russian words spelled in Cyrillic vs. Roman letters. Together, the present findings, in addition to those of these Arabic and Russian studies, do not provide evidence of an initial learning detriment associated with the presence of novel graphemes in the visual input during word learning. We in fact see evidence to the contrary in the present study: relative to (familiar but incongruent) *Pinyin*, exposure to (unfamiliar) *Zhuyin* ultimately afforded a word form learning advantage. We have thus provided additional evidence for the detrimental effects of orthographic incongruency between the L1 and L2, consistent with the findings of a number of earlier studies (e.g., Hayes-Harb et al., 2010; Escudero et al., 2014).

Our final research question was whether there is a difference in native English speakers' ability to perceive Mandarin consonant contrasts when exposed to *Pinyin* vs. *Zhuyin*. Experiment 3 was designed to determine whether any differential perceptual sensitivity to the Mandarin consonant contrast existed between the two groups of participants that might undermine the interpretation of the results of Experiments 1 and 2. However, there was no effect of exposure group on perceptual sensitivity to the distinction between the consonants contained in the newly learned words and their foils, confirming that the differential performance by the two exposure groups in Experiments 1 and 2 are attributable to *Pinyin* vs. *Zhuyin* exposure rather than to differences between the groups of participants in auditory discrimination ability. It is worth noting that our finding that differences in orthographic experience of participants in the two groups did not lead to a differential ability to perceive the consonant contrasts is consistent with the findings reported by Pytlyk (2011). We thus provide evidence that incongruencies between the L1 and L2 grapheme-phoneme correspondences can impact participants' memory for words' phonological forms in the absence of impacting their perceptual sensitivity to the relevant novel phonological contrasts. This suggests that, at least under the circumstances of the present study, the difficulty associated with suppressing native language grapheme-phoneme correspondences in favor of new ones played out at the level of the lexicon, with conflicts between orthographic and phonological information often resolved in favor of orthography, which was, crucially, interpreted via

grapheme-phoneme correspondence rules transferred from the native language.

CONCLUSION

The study of orthographic input in L2 phonological and word form acquisition has emerged only recently, and the present study represents an additional step in the direction of understanding the specific circumstances under which L2 learners' lexical development is helped or hindered by written input. Our aim was to investigate the influence of two factors that may moderate the influence of written input on L2 word form learning: (i) whether the writing system is shared by the native language and the L2, and (ii) if the writing system is shared, whether the relevant grapheme-phoneme correspondences are also shared. We did so via a series of experiments in which native English speakers were exposed to Mandarin words via auditory and visual (picture, written) input. Native speakers of English who had access to *Pinyin* (familiar writing system, some unfamiliar grapheme-phoneme correspondences) experienced difficulty learning the words' phonological forms due to interference from English grapheme-phoneme correspondences. Those who had access to *Zhuyin* (unfamiliar writing system) experienced no such interference, though they did initially take somewhat longer to learn the words' written forms.

In light of the fact that both *Pinyin* and *Zhuyin* are used in pedagogical settings to support Mandarin language acquisition, our findings can contribute to an understanding of the costs and benefits of each for this purpose. In particular, given that literate L2 learners are likely, especially in instructed settings, to be exposed to new words' phonological forms and their written forms more or less simultaneously, it is crucial that we understand the ways in which these two types of input impact the establishment and subsequent use of L2 lexical representations. Short laboratory-based studies like the one presented here differ importantly from real-world language acquisition; however, they do permit us to isolate and examine the factors that may contribute to L2 learning success or difficulty. One might next ask whether the patterns identified in the present study with respect to *Pinyin*'s and *Zhuyin*'s influence on L2 word form learning play out in actual native English-speaking learners of Mandarin, and whether Mandarin language experience (see Veivo and Jarvikivi, 2013) or other factors can moderate the effects of orthographic input.

AUTHOR CONTRIBUTIONS

RH and HC collaborated on this project while HC was a post-doctoral researcher under the supervision of RH. RH and HC were involved at all stages of the project, from its conception through design, data collection, and analysis. RH was responsible for preparing the manuscript for publication, in consultation with HC.

ACKNOWLEDGMENTS

We gratefully acknowledge the contributions of Richard Chi, Aaron Kaplan, Jeffrey Green, and Jing Zhao to the development of the study stimuli. We are also grateful to members of the

Speech Acquisition Lab at the University of Utah for their help with data collection, the Department of Linguistics and the Second Language Teaching and Research Center at the University of Utah for postdoctoral funding for HC, and to the two anonymous reviewers and editor for helpful suggestions.

REFERENCES

















- Basetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Writ. Lang. Lit.* 9, 95–114. doi: 10.1075/wll.9.1.07bas
- Basetti, B. (2008). "Orthographic input and second language phonology," in *Input Matters in SLA*, eds T. Piske and M. Young-Scholten (Clevedon: Multilingual Matters), 191–206.
- Basetti, B., Escudero, P., and Hayes-Harb, R. (2015). Second language phonology at the interface between acoustic and orthographic input. *Appl. Psycholinguist.* 36, 1–7. doi: 10.1017/S0142716414000393
- Cheng, H-W. (2012). *Semantic and Phonological Activation in First and Second Language Reading*. Unpublished dissertation. Boston University, Boston.
- Eckman, F., and Iverson, G. K. (2013). The role of native language phonology in the production of L2 contrasts. *Stud. Second Lang. Acquis.* 35, 67–92. doi: 10.1017/S027226311200068X
- Escudero, P. (2015). Orthography plays a limited role when learning the phonological forms of new words: the case of Spanish and English learners of novel Dutch words. *Appl. Psycholinguist.* 36, 7–22. doi: 10.1017/S014271641400040X
- Escudero, P., Hayes-Harb, R., and Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *J. Phon.* 36, 345–360. doi: 10.1016/j.wocn.2007.11.002
- Escudero, P., Simon, E., and Mulak, K. (2014). Learning words in a new language: orthography doesn't always help. *Bilingualism* 17, 384–395. doi: 10.1017/S1366728913000436
- Escudero, P., and Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Lang. Speech* 53, 343–365. doi: 10.1177/0023830910371447
- Hayes-Harb, R., and Hacking, J. F. (2015). The influence of written stress marks on native English speakers' acquisition of Russian lexical stress contrasts. *Slavic East Eur. J.* 59, 91–109.
- Hayes-Harb, R., Nicol, J., and Barker, J. (2010). Learning the phonological forms of new words: effects of orthographic and auditory input. *Lang. Speech* 53, 367–381. doi: 10.1177/0023830910371460
- ISO (2015). 7098: *Information and documentation – Romanization of Chinese*. Retrieved from: <https://www.iso.org/obp/ui/#iso:std:iso:7098:ed-3:v1:en> (Accessed December 31, 2015).
- MacMillan, N. A., and Creelman, C. D. (2004). *Detection Theory: A User's Guide*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mathieu, L. (2016). The influence of foreign scripts on the acquisition of a second language phonological contrast. *Second Lang. Res.* 32, 145–170. doi: 10.1177/0267658315601882
- Pytyk, C. (2011). Shared orthography: do shared written symbols influence the perception of L2 sounds? *Mod. Lang. J.* 95, 541–557. doi: 10.1111/j.1540-4781.2011.01244.x
- Racine, I., Bürki, A., and Spinelli, E. (2014). The implication of spelling and frequency in the recognition of phonological variants: evidence from pre-readers and readers. *Lang. Cogn. Neurosci.* 29, 893–898. doi: 10.1080/01690965.2013.832784
- Showalter, C. E., and Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: a novel lexicon study. *Second Lang. Res.* 29, 185–200. doi: 10.1177/0267658313480154
- Showalter, C. E., and Hayes-Harb, R. (2015). Native English speakers learning Arabic: the influence of novel orthographic information on second language phonological acquisition. *Appl. Psycholinguist.* 36, 23–42. doi: 10.1017/S0142716414000411
- Simon, E., Chambless, D., and Alves, U. K. (2010). Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Lang. Sci.* 32, 380–394. doi: 10.1016/j.langsci.2009.07.001
- Taylor, I., and Taylor, M. M. (2014). *Writing and Literacy in Chinese, Korean and Japanese: Revised Edition*. Amsterdam: John Benjamins Publishing.
- Veivo, O., and Jarviki, J. (2013). Proficiency modulates early orthographic and phonological processing in L2 spoken word recognition. *Bilingualism* 16, 864–883. doi: 10.1017/S1366728912000600
- Young-Scholten, M. (2002). "Orthographic input in L2 phonological development," in *An Integrated View of Language Development: Papers in Honour of Henning Wode*, eds P. Burmeister, T. Piske, and A. Rohde (Trier: Wissenschaftlicher Verlag Trier), 263–279.
- Young-Scholten, M., and Langer, M. (2015). The role of orthographic input in second language German: evidence from naturalistic adult learners' production. *Appl. Psycholinguist.* 36, 93–114. doi: 10.1017/S0142716414000447
- Ziegler, J. C., and Ferrand, L. (1998). Orthography shapes the perception of speech: the consistency effect in auditory word recognition. *Psychon. Bull. Rev.* 5, 683–689. doi: 10.3758/BF03208845

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Hayes-Harb and Cheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

TABLE A1 | Complete list of Mandarin words' auditory forms, their written forms in *Pinyin* and *Zhuyin*, their auditory foils, and their pictured meanings.

	<i>Pinyin</i>	<i>Zhuyin</i>	Auditory form	Auditory foil	Pictured meaning
Congruent items	nai	ㄋㄞˊ	/naɪ/	/daɪ/	
	nao	ㄋㄠˊ	/nau/	/dau/	
	sai	ㄙㄞˊ	/saɪ/	/θaɪ/	
	sao	ㄙㄠˊ	/sau/	/θau/	
	lie	ㄌㄧㄝˊ	/liɛ/	/diɛ/	
	liu	ㄌㄧㄡˊ	/liou/	/diou/	
	mie	ㄇㄧㄝˊ	/mie/	/θiɛ/	
	miu	ㄇㄧㄡˊ	/miou/	/θiou/	
Incongruent items	zai	ㄗㄞˊ	/tsaɪ/	/zaɪ/	
	zao	ㄗㄠˊ	/tsau/	/zau/	
	cai	ㄘㄞˊ	/ts ^h aɪ/	/kaɪ/	
	cao	ㄘㄠˊ	/ts ^h au/	/kau/	
	qie	ㄑㄧㄝˊ	/tɕ ^h iɛ/	/kiɛ/	
	qiu	ㄑㄧㄡˊ	/tɕ ^h iou/	/kiou/	
	xie	ㄒㄧㄝˊ	/ɕiɛ/	/ziɛ/	
	xiu	ㄒㄧㄡˊ	/ɕiou/	/ziou/	



Phonotactic Constraints Are Activated across Languages in Bilinguals

Max R. Freeman^{1*}, Henrike K. Blumenfeld² and Viorica Marian¹

¹ Bilingualism and Psycholinguistics Research Group, Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA, ² Bilingualism and Cognition Laboratory, School of Speech, Language, and Hearing Sciences, San Diego State University, San Diego, CA, USA

OPEN ACCESS

Edited by:

Isabelle Darcy,
Indiana University, USA

Reviewed by:

Barbara C. Malt,
Lehigh University, USA
Gorka Elordieta,
University of the Basque Country
UPV/EHU, Spain

*Correspondence:

Max R. Freeman
freeman@u.northwestern.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 10 December 2015

Accepted: 26 April 2016

Published: 18 May 2016

Citation:

Freeman MR, Blumenfeld HK
and Marian V (2016) Phonotactic
Constraints Are Activated across
Languages in Bilinguals.
Front. Psychol. 7:702.
doi: 10.3389/fpsyg.2016.00702

During spoken language comprehension, auditory input activates a bilingual's two languages in parallel based on phonological representations that are shared across languages. However, it is unclear whether bilinguals access phonotactic constraints from the non-target language during target language processing. For example, in Spanish, words with s+ consonant onsets cannot exist, and phonotactic constraints call for epenthesis (addition of a vowel, e.g., *stable/estable*). Native Spanish speakers may produce English words such as *estudy* ("study") with epenthesis, suggesting that these bilinguals apply Spanish phonotactic constraints when speaking English. The present study is the first to examine whether bilinguals access Spanish phonotactic constraints during English *comprehension*. In an English cross-modal priming lexical decision task, Spanish–English bilinguals and English monolinguals heard English cognate and non-cognate primes containing s+ consonant onsets or controls without s+ onsets, followed by a lexical decision on visual targets with the /e/ phonotactic constraint or controls without /e/. Results revealed that bilinguals were faster to respond to /es/ non-word targets preceded by s+ cognate primes and /es/ and /e/ non-word targets preceded by s+ non-cognate primes, confirming that English primes containing s+ onsets activated Spanish phonotactic constraints. These findings are discussed within current accounts of parallel activation of two languages during bilingual spoken language comprehension, which may be expanded to include activation of phonotactic constraints from the irrelevant language.

Keywords: bilingualism, phonology, epenthesis, parallel language activation, comprehension

INTRODUCTION

Across many contexts and discourse situations, bilinguals activate both languages simultaneously, even when only one language is used overtly, a phenomenon known as parallel activation (e.g., Green, 1998; Dijkstra and van Heuven, 2002; Blumenfeld and Marian, 2007; Kroll et al., 2008; Shook and Marian, 2013). Bilinguals have previously demonstrated parallel activation of phonological (Marian and Spivey, 2003; Blumenfeld and Marian, 2007, 2013; Darcy et al., 2015), lexical (Linck et al., 2008; Bartolotti and Marian, 2012), semantic (Martín et al., 2010), and syntactic (Linck et al., 2008; Kootstra et al., 2012) information across their two languages. In the current study, we explore whether cross-linguistic activation of phonological structures

extends to phonotactic constraints (i.e., legal ways for combining speech sounds) of the non-target language during spoken word comprehension in bilinguals. Specifically, we address the question: *Do Spanish–English bilinguals access Spanish phonotactic constraints during English comprehension?*

Phonotactic constraints can differ across languages, which may become a stumbling block for second language (L2) speakers during initial stages of L2 acquisition and use (e.g., Flege and Davidian, 1984). Specifically, language production studies suggest that when the phonology of the L2 does not align with or is not present in the native language (L1), L2 learners and bilinguals may experience interference from the non-target language (e.g., Yavas and Someillan, 2005). For example, while word-initial s+ consonant clusters are legal in English, a phonotactic constraint for Spanish is that s+ consonant clusters cannot exist at word onsets and an epenthetic /e/ (i.e., the addition of a vowel) must be added to render the word acceptable in Spanish. This incongruence between phonotactic constraints in the L1 and L2 might result in Spanish-like pronunciations and perceptions of English words during spoken word production and comprehension (e.g., *stable*, Spanish: *estable*).

Comprehension

During receptive language processing, Spanish-only speakers have been shown to activate the epenthetic /e/ when viewing real Spanish words, even when the /e/ is removed from the word onset. Spanish speakers who performed a visual lexical decision task on words containing as+ and es+ consonant onsets showed facilitation of the epenthetic /e/ when primed with a Spanish word that had the /e/ onset removed (e.g., Spanish *stable*/"*estable*"; Hallé et al., 2008). The Spanish monolinguals in Hallé et al.'s (2008) study likely activated the epenthesis onset because Spanish was *overtly* presented and participants were judging lexicality in Spanish (not English).

Within-language activation of phonotactic constraints has been observed with monolinguals in other consonant–vowel contexts. For example, Japanese monolinguals applied an epenthesis constraint by adding a vowel (e.g., /u/) to an illegal consonant cluster in the coda of syllables when hearing Japanese-like non-words (e.g., they heard 'mikdo,' but perceived it as 'mikudo'; Dupoux et al., 2001; Carlson et al., 2015). Hallé et al. (2008) discuss the process of epenthesis within consonant clusters as phonological repair (i.e., modifying auditory input that is phonologically illegal to conform to native language rules). Moreover, Parlato-Oliveira et al. (2010) examined how bilingual experience influenced the way the epenthesis constraint was repaired. Native Japanese-speaking adults who had been exposed to Portuguese (L2) when entering school demonstrated similar epenthesis patterns as native Portuguese listeners when processing illegal consonant clusters. Moreover, simultaneous Japanese-Brazilian bilinguals who were exposed to both languages from birth also demonstrated epenthesis repair similar to that observed in native Brazilian speakers (adding an /i/). Thus, previous results suggest that monolinguals and bilinguals potentially access and repair auditory input to align with their native or more proficient language.

While cross-linguistic activation of phonotactic constraints has yet to be established in comprehension, parallel language activation has been identified in other areas of phonology. Studies suggest that non-native listeners may rely on phonological categories from the non-target L1 during L2 auditory comprehension. For example, the two distinct vowels /ɛ/ and /æ/ are contrastive phonemes in English, but are non-contrastive allophones in Dutch. Consequently, Dutch learners of English, but not English monolinguals, erroneously activated 'deaf' when primed with 'daf' (Broersma and Cutler, 2011). If the highly proficient Dutch-English bilinguals tested in this study had mastered the /ɛ/ and /æ/ phonological category distinction of their L2 (English), then the findings would suggest access of L1 phonological categories during L2 processing. Alternatively, it is possible that even proficient L2 learners routinely rely on L1 categories during phonological processing in L2. Thus, previous research indicates that individuals are attuned to the phonotactic constraints of their L1 during native-language listening tasks (Hallé et al., 2008), and that bilinguals may potentially activate L1 phonological categories during L2 comprehension (e.g., Broersma and Cutler, 2011; Darcy et al., 2015). In the current study, we ask if bilinguals are also attuned to the phonotactic epenthesis constraint of the L1 (Spanish) during L2 (English) comprehension.

Production

Evidence from word production also suggests that bilinguals are susceptible to cross-linguistic activation of phonological structures. Fabra and Romero (2012) found that L1 Catalan speakers of English produced English words with vowels (/i/, /ɛ/, /a/, /ʌ/) that were less peripheral (i.e., sounded more like Catalan vowel phonology), than native English monolinguals. The less peripheral vowel effect disappeared as proficiency in English increased. Notably, all of the vowels except /ʌ/ are shared across English and Catalan, thus the results suggest access of L1 phonological categories. As in comprehension (Broersma and Cutler, 2011), spillover effects of L1 phonological categories into L2 productions have been identified; but would there also be a similar effect with bilinguals accessing *phonotactic constraints* from the non-target language? Native Spanish speakers speaking English may at times produce words such as *estric* in English ("strict"), adding an additional /e/ to the onset of words (Yavas and Someillan, 2005; see Roelofs and Verhoef, 2006, for review of bilingual cross-linguistic phonological access during production). While we have seen evidence for irrelevant-language phonological category and phonotactic constraint access during *production*, it is not clear whether bilinguals also access cross-linguistic phonotactic constraints during comprehension.

Previous investigations have explored the contexts in which cross-linguistic phonological activation could be facilitated. For example, cognates, which are words that overlap in form and meaning across languages (e.g., English: *stable*/Spanish: *estable*), have been used to test phonological co-activation during production (e.g., Amengual, 2012) and comprehension (e.g., Blumenfeld and Marian, 2007). It has been hypothesized that joint activation of similar-sounding translation equivalents enhances activation of phonological representations across

languages. Amengual (2012) examined voice onset times (VOTs) of cognates and non-cognates produced by Spanish-English bilinguals. The results suggest that bilinguals produced longer (more English-like) VOTs on Spanish voiceless stops when producing cognates (e.g., English/Spanish *tumor*). In the presence of cognates, bilinguals may thus be more likely to experience activation of the non-target language. In an eye tracking study, English-German bilinguals' looks to pictures representing cognate targets and cross-linguistic competitors suggested that cognates increased phonological co-activation of a less proficient non-target L2 during auditory word comprehension (Blumenfeld and Marian, 2007). It is possible that activation of cross-linguistic phonotactic constraints may become enhanced when phonological representations of the other language are co-activated. Including cognates in the current study provides a condition in which phonological co-activation of languages is most likely to occur.

The large body of research on parallel language activation in bilinguals, including phonological co-activation, has been captured by current models of bilingual language comprehension and production (e.g., Dijkstra and van Heuven, 2002; Shook and Marian, 2013). While current models of bilingual language *comprehension* do not specifically account for phonotactic constraints, one model of bilingual language *production*, the WEAVER++ model, does indeed propose that bilinguals access non-target language phonology (Roelofs and Verhoeft, 2006). During bilingual production, activation of non-target language phonotactic constraints is thought to occur between encoding of the phonological word form for production and its phonetic realization. WEAVER++ posits that non-target language phonological representations and/or phonotactic constraints may intrude during encoding of words for production, and may combine with the phonological representations or phonotactic constraints of the target language to affect phonetic realization (e.g., applying the Spanish epenthetic /e/ to an English s+ consonant cluster, *estudy*).

In summary, while current experimental and theoretical work on bilingual language comprehension suggests that bilinguals co-activate phonological representations of the non-target language, it remains unclear whether they access cross-linguistic phonotactic constraints during language comprehension. The current study has the potential to expand upon the existing knowledge base for the types of cross-linguistic phonological interactions that occur during bilingual language comprehension.

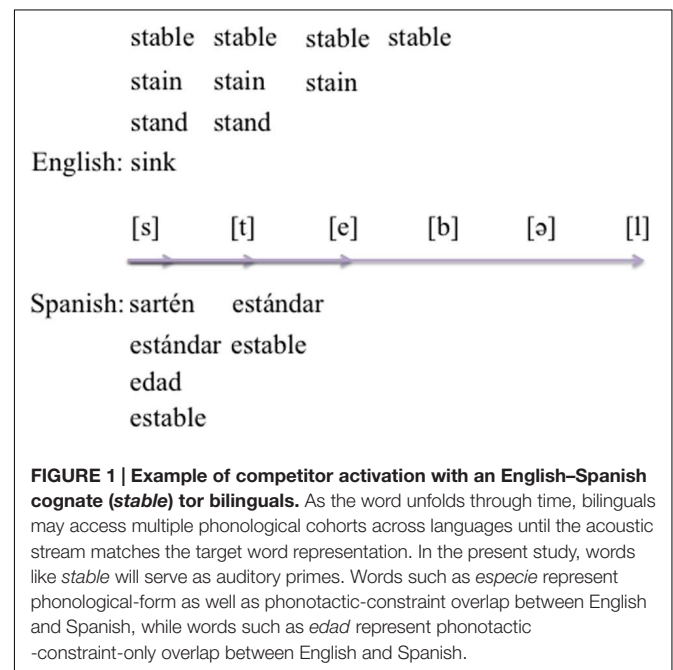
Current Study

In the current study, we explore for the first time whether bilinguals co-activate phonotactic constraints from the non-target language during comprehension. Furthermore, while phonotactic constraint activation has been observed empirically during production (e.g., Yavas and Someillan, 2005), we test whether bilinguals also access phonotactic constraints during comprehension. Thus, the current study attempts to provide evidence for the extent to which cross-linguistic structures are accessed during language comprehension in bilinguals.

In order to measure if bilinguals activated phonotactic constraints in the non-target language (Spanish), we employed a

cross-modal phonological priming lexical decision (PPLD) task. We used cognates and non-cognates to index availability of phonotactic constraints in different contexts of cross-linguistic phonological activation (e.g., Van Hell and Dijkstra, 2002; Blumenfeld and Marian, 2007). For example, when Spanish-English bilinguals hear the cognate *stable* unfold through the acoustic stream, they may initially activate phonological cohorts from both languages (e.g., *stand*, *stain*, *sink/sárten*, e.g., Blumenfeld and Marian, 2007, 2013) and the Spanish translation equivalent (i.e., *estable*; e.g., Linck et al., 2009). Critically, when hearing 'stable,' they may also activate phonological cohorts that overlap with Spanish through phonotactic constraints and phonological form (e.g., *estándar*/standard) and potentially even cohorts that overlap with Spanish through phonotactic constraints only (e.g., *edad*/age). As an alternative to activation of phonological and phonotactic cohorts upon hearing 'stable' in English, native Spanish speakers may perceptually repair 'stable' to "e-stable," (/esteibəl/) and therefore may not hear 'stable' (Hallé et al., 2008). Whether bilinguals access neighbors containing phonotactic constraints through spreading activation and mediated priming (English 'stable' activates Spanish /e/ onset words) or repair the auditory input to have the epenthesis onset, both scenarios suggest that bilinguals may access the phonotactic constraint of /e/ onset from their L1 and apply it during L2 processing.

Here, we examine both phonotactic-constraint-*and*-form access as well as phonotactic-constraint-*only* access across English and Spanish in order to dissociate constraint from form overlap (e.g., *edad* and *estándar*, respectively, see Figure 1). We will henceforth refer to the phonotactic-constraint-*and*-form manipulation as the PCF condition, and to the phonotactic-constraint-*only* manipulation as the PC condition. We focused on the Spanish epenthesis constraint (/e/ onset, e.g., English



'*estudy*') because it is a commonly observed phenomenon that occurs in production with native Spanish speakers speaking English, and thus presents a good starting point in exploring a phonotactic constraint during comprehension. The Spanish epenthesis constraint is particularly suitable to the current experimental manipulation because of its potential to be primed with English words that violate the Spanish phonotactic constraint.

We hypothesized that Spanish–English bilinguals would access Spanish (L1) phonotactic constraints during English (L2) comprehension. The goal was to examine the presence or absence of non-target language phonotactic constraint activation when phonological and lexico-semantic (cognate) or no (non-cognate) overlap was present between auditory primes and their translation equivalents. Moreover, we predicted that when bilinguals were primed with an /st/ or /sp/ word, they would access shared phonological (e.g., 'strong'/stand/estándar), lexical (e.g., 'strong'/fuerte), and potentially phonotactic constraint (e.g., 'strong'/edad) neighbors across languages. Presentation of visual /est/, /esp/, or /e/ non-word targets (e.g., *esteriors*) would then limit cross-linguistic activation to strictly phonological forms (/es/ onset) and/or phonotactic constraints (/e/ onset) that had been previously activated by the prime (e.g., Dijkstra and van Heuven, 2002; Shook and Marian, 2013). Restricted activation of phonological representations (/e/ and /es/ onsets) across primes and targets would in turn facilitate lexical selection, and thus yield faster reaction times when making a lexical decision. Given that the phonology of critical targets (e.g., *esteriors*) was expected to activate partial phonological form and phonotactic constraints of Spanish, but no specific Spanish lexical items, we predicted that there would be no lexical interference from Spanish. These predictions are supported by previous research using a lexical decision task and manipulating the amount of word-initial phoneme overlap across languages (e.g., no-overlap, 1-phoneme overlap, 2 phoneme-overlap, and 3-phoneme overlap). When Russian–English bilinguals processed words in the non-native language (English), cross-linguistic phonological overlap of word onsets was associated with faster reaction times as compared to no phonological overlap (Marian et al., 2008). In the current study, we expected that s+ consonant priming would restrict activation to words with /e/ and /es/ onsets. Therefore, Spanish–English bilinguals would be able to quickly search through a constrained space within the lexicon of s+ consonant, es+ consonant, and e+ consonant onset words to make a lexical decision. In contrast, for control non-words that did not conform to the epenthesis constraint, phonological representations would need to be activated for the first time, delaying the subsequent lexical search, and resulting in slower reaction times.

Including the cognate and non-cognate priming conditions, as well as the target conditions with PCF and PC overlap, ensured that bilingual participants would experience local (i.e., intermittent) co-activation of Spanish throughout the task. We predicted that cognates (e.g., *stable* /steɪbəl/ *estable* /estaβle/) would facilitate activation of Spanish translation equivalents more strongly than non-cognates (e.g., *strong*/fuerte) based on phonological form overlap (e.g., *stable* /steɪbəl/ *estable* /estaβle/).

Following the /sp/ and /st/ primes, PCF non-word targets that overlapped with Spanish /esp/ or /est/ onsets would potentially activate Spanish phonological form in addition to the constraint. The PC targets shared just the Spanish /e/ onset (epenthesis constraint), therefore activating Spanish to a lesser degree. We expected that, if bilingual participants would locally co-activate Spanish, effects on /e/ and /es/ non-word targets would be present only when directly preceded by /sp/ or /st/ primes, but not when preceded by control primes (e.g., workers).

We specifically predicted, across conditions on the PPLD task, that if cognate auditory primes (e.g., *stable*) facilitated non-target language phonotactic constraint and phonological form access, then bilinguals would demonstrate faster reaction times to visual letter strings that contained the previously-activated phonological cohorts (e.g., PCF non-words: *esteriors*), as compared to conditions in which less or no phonological overlap was present (e.g., controls: *stable/hereander* or *workers/hainsail*). In addition, we expected that if the non-cognate auditory primes (e.g., *strong*) facilitated phonotactic constraint access, then the bilingual group would demonstrate faster reaction times to non-word targets with PCF overlap (e.g., *estimagle*), relative to control trials (e.g., *strong/atongside*). However, this facilitation effect was predicted to be less strong than the cognate prime/PCF trials because of the absence of overlap between translation equivalents in the non-cognate prime. If bilinguals routinely activated phonotactic constraints across their two languages, then we would also expect to see similar reaction time facilitation effects for non-word targets that overlapped only with the phonotactic constraint when paired with cognate and non-cognate primes (e.g., /e/-only onset: *stable/elopevent* and *strong/encimpass*, respectively). We expected that this facilitation effect would be less robust in comparison to the PCF overlap condition, since phonological form overlap was not present. We included a control-prime condition, which was not expected to activate Spanish due to either phonotactic constraint or lexico-semantic overlap, as no overt overlap between English and Spanish was present in the control stimuli.

MATERIALS AND METHODS

Participants

Participants included 22 Spanish–English bilinguals and 23 English monolinguals, ages 18–33. Monolinguals and bilinguals were recruited via word-of-mouth, e-mails to local student and community organizations, flyers posted around campus and the community, as well as through existing participant databases. This study was carried out in accordance with the recommendations of Northwestern University's Institutional Review Board with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. Any of the monolingual English participants who had a self-reported Spanish speaking proficiency of greater than 3 (0–10 scale) on the language experience and proficiency questionnaire (LEAP-Q; Marian et al., 2007) did not participate in the experiment. Bilinguals were

TABLE 1 | Linguistic and cognitive background of Spanish–English bilingual ($n = 22$) and English monolingual ($n = 23$) participants.

	Bilinguals mean (SE)	Monolinguals mean (SE)
Age	24.09 (0.84)	22.95 (0.74)
Age of Spanish acquisition	0.45 (0.12)	–
Age of English acquisition**	6.05 (0.49)	0.18 (0.08)
Current exposure to Spanish	36.77% (6.40)	–
Current exposure to English**	62.50% (6.80)	98.65% (0.69)
Foreign accent in Spanish (0–10 scale)	2.10 (0.44)	–
Foreign accent in English (0–10 scale)*	2.82 (0.56)	0.73 (0.56)
Spanish receptive vocabulary (NIH Toolbox)	116.77 (2.84)	–
English receptive vocabulary (NIH Toolbox)	110.14 (3.55)	118.86 (3.39)
Self-reported Spanish proficiency (0–10 scale)	9.03 (0.14)	–
Self-reported English proficiency (0–10 scale)	8.95 (1.10)	9.83 (0.05)
WASI, matrix reasoning	29.27 (0.53)	28.78 (0.61)
Backward digit span	7.33 (1.20)	10.14 (1.10)

** $p < 0.001$; * $p < 0.01$.

native Spanish speakers, were exposed to Spanish at least 30% of the time daily, and acquired English at age 5 or later. See **Table 1** for additional participant information. Monolinguals and bilinguals differed on English age of acquisition ($p < 0.001$), current exposure to English ($p < 0.001$), and foreign accent in English ($p < 0.01$). Participants were matched on age, non-verbal cognitive reasoning (WASI; PsychCorp, 1999), and working memory (backward digit span; Woodcock et al., 2001/2007).

Materials

The English cross-modal PPLD task was designed to measure cross-linguistic activation of the Spanish phonotactic constraint (the epenthetic /e/) in the presence of phonological and lexico-semantic overlap between languages (cognate auditory primes) or in the absence of phonological overlap between languages (non-cognate auditory primes) through accuracy and reaction time to target identification. The within-subjects independent variables included prime type (cognate, non-cognate, control) and target type (PCF overlap non-word, PC non-word, non-word control, word control). The /st/ and /sp/ consonant clusters were chosen because they are illegal consonant clusters in Spanish without the obligatory epenthetic /e/ at the word onset. In addition, the two

consonant clusters are present in a sufficient number of English cognates and non-cognates to generate stimuli for the current study.

The cross-modal PPLD task was programmed in MatLab (Psychtoolbox add-on; Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). The auditory primes were recorded in a soundproof room (44,100 Hz, 16 bits) by a native female speaker of English. The audio recording was split into individual audio files and all files were normalized (via audio compression) in Praat (Boersma and Weenink, 2013) and exported into MatLab (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). Each prime type was paired with each target type (3x4), resulting in 12 different pairing combinations and the repetition of each prime four times and each target three times throughout the duration of the experiment. **Table 2** depicts examples of stimulus pairings for each prime and target type.

A total of 360 critical trial pairs were created, comprised of cognate primes (30 items), non-cognate primes (30 items), and control primes (30 items). Each of the auditory primes was paired with a visual target that included non-words that overlapped with Spanish via phonotactic constraint and phonological form (/es/ onset, 30 items), via phonotactic constraint only (/e/ onset, 30 items), non-words that did not overlap with Spanish via phonotactic constraint or form (non-word control, 30 items), or a real word in English that did not overlap with Spanish (word control, 30 items). The PCF (/es/ onset) non-word targets were controlled in such a way that they overlapped cross-linguistically with only the first three letters of the Spanish translation of the cognate prime [e.g., cognate prime *stable* (*estable*) was paired with /es/ non-word target (*esteriors*)]. Controlling the targets in this manner would avoid any priming effects due to additional phonological and orthographic overlap. The PC non-word targets overlapped with the cognate prime's translation equivalent only at the /e/ onset [e.g., cognate prime *stable* (*estable*) was paired with /e/ non-word target *elopevent*]]. To a) balance the proportion of word (50%) versus non-word (50%) trials, and b) prevent the participants from noticing any patterns concerning the critical stimulus pairs, 45 auditory prime fillers and 45 visual target fillers (180 total trial pairs) were also generated. Twelve additional pairs were created as practice trials. The experiment was divided into four blocks and the items were pseudo-randomized such that no two consecutive trials contained cognate primes. Consistent with cross-linguistic priming studies employing lexical decision tasks, cognate and non-cognate trials were presented in an intermixed order (Duyck et al., 2007; Siyambalapitiya et al., 2009; Davis et al., 2010; Dijkstra et al., 2010). Finally, trial order was counterbalanced (reversed) across participants.

TABLE 2 | Example stimulus pairings and total number of each item type.

Auditory prime	Phonotactic constraint + Phonological form target	Phonotactic constraint only target	Non-word control target	Word control target
30 Cognates (<i>stable</i>)	30 (<i>esteriors</i>)	30 (<i>elopevent</i>)	30 (<i>hereander</i>)	30 (<i>flattened</i>)
30 Non-cognates (<i>strong</i>)	30 (<i>estimagle</i>)	30 (<i>encimpass</i>)	30 (<i>atongside</i>)	30 (<i>daughters</i>)
30 Controls (<i>workers</i>)	30 (<i>esported</i>)	30 (<i>ebvision</i>)	30 (<i>hainsail</i>)	30 (<i>kneeling</i>)

All stimuli were controlled for various lexical characteristics. The three types of auditory primes did not differ on any of the lexical characteristics listed in Appendix A (all $ps > 0.05$).

The four types of lexical decision targets also did not differ on any of the lexical characteristics ($ps > 0.05$), with the exception of lexical decision reaction time (LDT RT) and lexical decision z-score (LDT Zscore) in which non-words had slower lexical decision response times than words in the normed sample, $ps < 0.05$ (Balota et al., 2007). See Appendix B for means and standard deviations. Similar to previous studies (e.g., Martín et al., 2010), we did not control for part of speech (auditory primes) due to the number of lexical characteristics on which the stimuli needed to be matched.

Procedure

Tasks were administered in the following order:

- (1) the LEAP-Q (Marian et al., 2007) to obtain linguistic background information and current language exposure, and to ensure that each participant met the criteria for the study;
- (2) the cross-modal PPLD task (auditory prime, visual target) to examine cross-linguistic phonotactic constraint access;
- (3) a non-linguistic Stroop task to index competition resolution abilities (adapted from Blumenfeld and Marian, 2014);
- (4) a backward digit span task (numbers reversed, Woodcock et al., 2001/2007) to index working memory;
- (5) the Wechsler Abbreviated Scale of Intelligence (WASI; PsychCorp, 1999) to index non-verbal cognitive reasoning; and
- (6) the NIH Cognition Toolbox Battery (National Institutes of Health Toolbox Cognition Battery (NIH Toolbox CB), 2013) picture vocabulary test, as a measure of English (bilinguals and monolinguals) and Spanish (bilinguals only) proficiency.

Participants were seated in a quiet room with a single iMac computer and were asked to pay attention to the word they heard and then respond by indicating whether what they saw on the screen was a word or non-word in English as quickly and as accurately as possible. After the instructions and 12 practice trials, participants performed the experimental task in which they first heard an auditory prime (cognate, non-cognate, control, filler) and then saw a visual written target (PCF overlap non-word, PC non-word, non-word control, word control, filler) on the screen after a 350 ms inter-stimulus interval (ISI). During presentation of the auditory prime through the 350 ms ISI, participants viewed a central fixation crosshair on the computer screen. Previous studies using similar priming techniques have shown effects of parallel activation 350–500 ms post-stimulus onset (e.g., Martín et al., 2010; Blumenfeld and Marian, 2013). The visual targets were presented in the center of a white screen in black, size 16 font, Courier, and the left/right shift keys represent yes/no responses. Presentation of written words lasted until the participant made a response or for 3,000 ms after the onset of the display (see Figure 2).

Participants were given three short, but untimed, breaks in between each of the four blocks. The total time to complete this task was approximately 30 min. Participants performed the remaining tasks, then were debriefed about the study and compensated. The total study duration was approximately 2 h.

Coding and Analysis

For the PPLD task, reaction times and accuracy rates were analyzed. Reaction times were measured from the onset of the visual lexical decision target (PPLD task). Filler trials were not analyzed, as they only served to balance the word/non-word ratio. Incorrect trials and trials 2.5 standard deviations above and below the mean reaction time were disregarded for both tasks. Means and standard deviations for each condition (12 critical conditions) were then calculated.

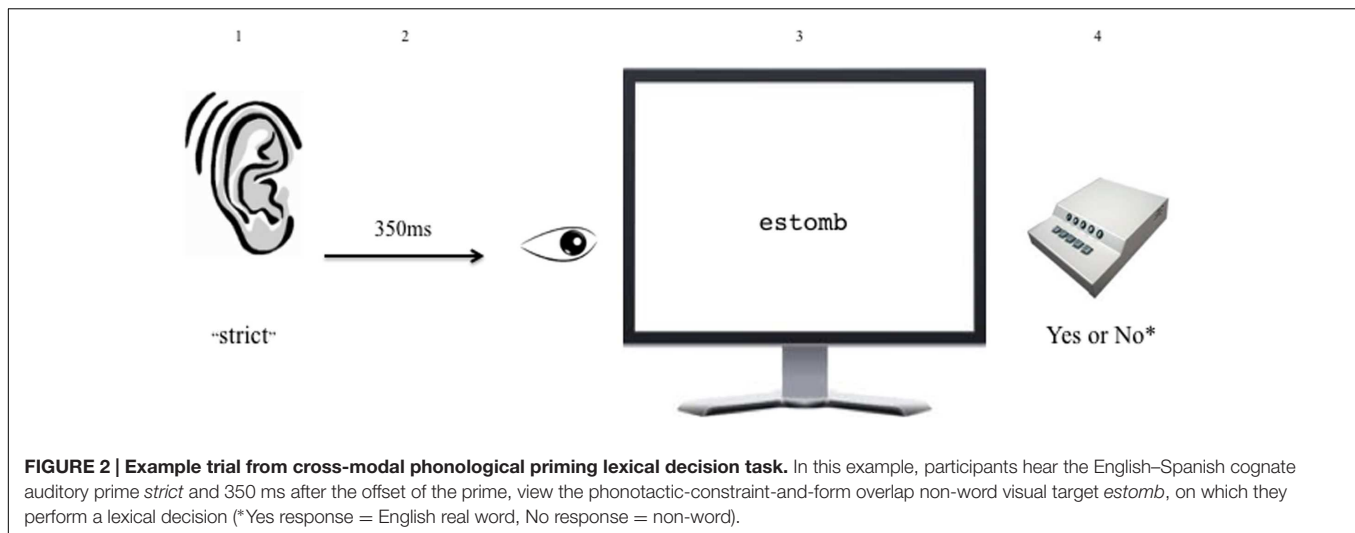
RESULTS

Overall Accuracy Effects on the PPLD Task

We examined lexical decision accuracy, expecting that decisions on non-words would be less accurate than on real words based on previous research (de Groot et al., 2000). A 3 (auditory prime: cognate, non-cognate, control) \times 4 (visual target: PCF overlap non-word, PC non-word, non-word control, word control) \times 2 (language group: monolingual, bilingual) repeated measures ANOVA was conducted on the lexical decision targets. There was a main effect of target, $F(3,129) = 4.26$, $p < 0.01$, $\eta_p^2 = 0.09$, with Bonferroni-corrected pairwise *post hoc* comparisons revealing that participants were more accurate on PCF overlap non-word trials (e.g., *esteriors*; $M = 96.89\%$, $SE = 0.47$) than on word control trials (e.g., *flattened*; $M = 94.87\%$, $SE = 0.90$), $p = 0.045$. While we did not anticipate higher accuracy for non-words, we reason that this accuracy effect may have been due to participants using more time to make a decision on non-words than on words, as evidenced by increased reaction times for non-words (see below).

Overall Reaction Time Effects on the PPLD Task

We next examined whether monolinguals would be faster overall in their lexical decision response rates than bilinguals, as bilinguals were performing a lexical decision in their L2 (Dijkstra et al., 1999). Further, we tested whether participants were slower to respond to non-words than words, a pattern demonstrated in previous research (Dijkstra et al., 1999). There was a main effect of language group, $F(1,43) = 11.70$, $p < 0.01$, $\eta_p^2 = 0.21$, indicating that monolinguals ($M = 655.96$ ms, $SE = 46.10$) indeed responded to targets more quickly than bilinguals ($M = 881.31$ ms, $SE = 47.10$), $p < 0.01$. A main effect of visual target condition was also identified, $F(3,129) = 16.02$, $p < 0.001$, $\eta_p^2 = 0.27$, with Bonferroni-corrected pairwise *post hoc* comparisons indicating the following



patterns: participants were faster to respond to PCF overlap non-word trials (e.g., *esteriors*; $M = 758.99$ ms, $SE = 31.49$) than to non-word controls (e.g., *hereander*; $M = 800.52$ ms, $SE = 37.11$), $p < 0.001$; faster to respond to PC non-word trials (e.g., *elopevent*; $M = 770.47$ ms, $SE = 32.54$) than to non-word controls (e.g., *hereander*; $M = 800.52$ ms, $SE = 37.11$), $p < 0.01$; faster to respond to word-control trials (e.g., *flattened*; $M = 744.50$ ms, $SE = 31.90$) than to PC non-word trials (e.g., *elopevent*; $M = 770.47$ ms, $SE = 32.54$), $p < 0.05$; and faster to respond to word-control trials (e.g., *flattened*; $M = 744.50$ ms, $SE = 31.90$) than to non-word-control trials with other word onsets (e.g., *hereander*; $M = 800.52$ ms, $SE = 37.11$), $p < 0.001$. Thus, reaction time differences across target conditions confirmed faster overall responses in monolinguals than bilinguals and faster responses to words over non-words. Effects of target condition warranted further follow-up analyses across monolinguals and bilinguals.

Monolingual versus Bilingual Reaction Time Performance

Next, related to our prediction of greater cross-linguistic activation effects in bilinguals than monolinguals, we examined whether differences in performance across target conditions would be greater for bilinguals than monolinguals. Indeed, an interaction emerged for reaction times between target type and language group, $F(3,129) = 4.18$, $p < 0.01$, $\eta_p^2 = 0.09$. Bonferroni-corrected pairwise comparisons revealed that, relative to monolinguals, bilinguals showed additional reaction time effects across target conditions, with faster reaction times to PCF overlap non-words ($M = 866.49$ ms, $SE = 59.16$) than to non-word control trials ($M = 928.81$ ms, $SE = 68.28$), $p < 0.01$, and a marginal effect of faster reaction times to PC non-word trials ($M = 885.55$ ms, $SE = 60.30$) than to non-word control trials ($M = 928.81$ ms, $SE = 68.28$), $p = 0.058$. Monolinguals did not demonstrate such effects, $p > 0.05$.

Phonotactic Constraint Activation between Cognate and Non-cognate Primes and Target Conditions

Finally, we tested our key prediction following the hypothesis of bilinguals' activation of irrelevant-language phonotactic constraints during comprehension. We conducted planned follow-up t -test comparisons within monolingual and bilingual groups to probe for reaction time effects across prime and target conditions of interest. It was expected that some priming effects would occur for monolinguals, as there was /st/ or /sp/ overlap between the prime and target. Indeed, a significant difference was observed for monolinguals, with faster reaction times to PCF overlap targets (e.g., *estimagle*) preceded by non-cognate primes (e.g., *strong*; $M = 662.62$ ms, $SE = 27.10$) than to non-word controls (e.g., *atongside*) preceded by non-cognate primes ($M = 677.61$ ms, $SE = 32.10$), $t(22) = -2.51$, $p = 0.02$. However, bilinguals demonstrated several significant reaction time differences across prime and target conditions in line with non-target language phonotactic constraint activation. Bilinguals were faster to respond to PCF overlap non-word trials (e.g., *esteriors*) preceded by cognate primes (e.g., *stable*; $M = 848.45$ ms, $SE = 57.70$) than to non-word controls (e.g., *hereander*) preceded by cognate primes ($M = 922.29$ ms, $SE = 66.42$), $t(21) = -3.94$, $p = 0.001$. Bilinguals were also marginally faster to respond to PC non-word trials (e.g., *elopevent*) preceded by cognate primes ($M = 883.83$ ms, $SE = 56.68$) than to non-word control trials preceded by cognate primes ($M = 922.29$ ms, $SE = 66.42$), $t(21) = -1.83$, $p = 0.082$. Finally, bilinguals were faster to respond to PCF overlap non-word targets (e.g., *estimagle*) and PC non-word targets (e.g., *encimpass*) preceded by non-cognate primes (e.g., *strong*; $M = 876.33$ ms, $SE = 61.85$; $M = 881.14$ ms, $SE = 62.67$, respectively) than to non-word controls preceded by non-cognate primes ($M = 944.39$ ms, $SE = 72.11$), $t(21) = -4.63$, $p < 0.001$; $t(21) = -3.56$, $p < 0.01$, respectively. (See **Figures 3A,B** for the bilingual versus monolingual reaction time by condition comparison.)

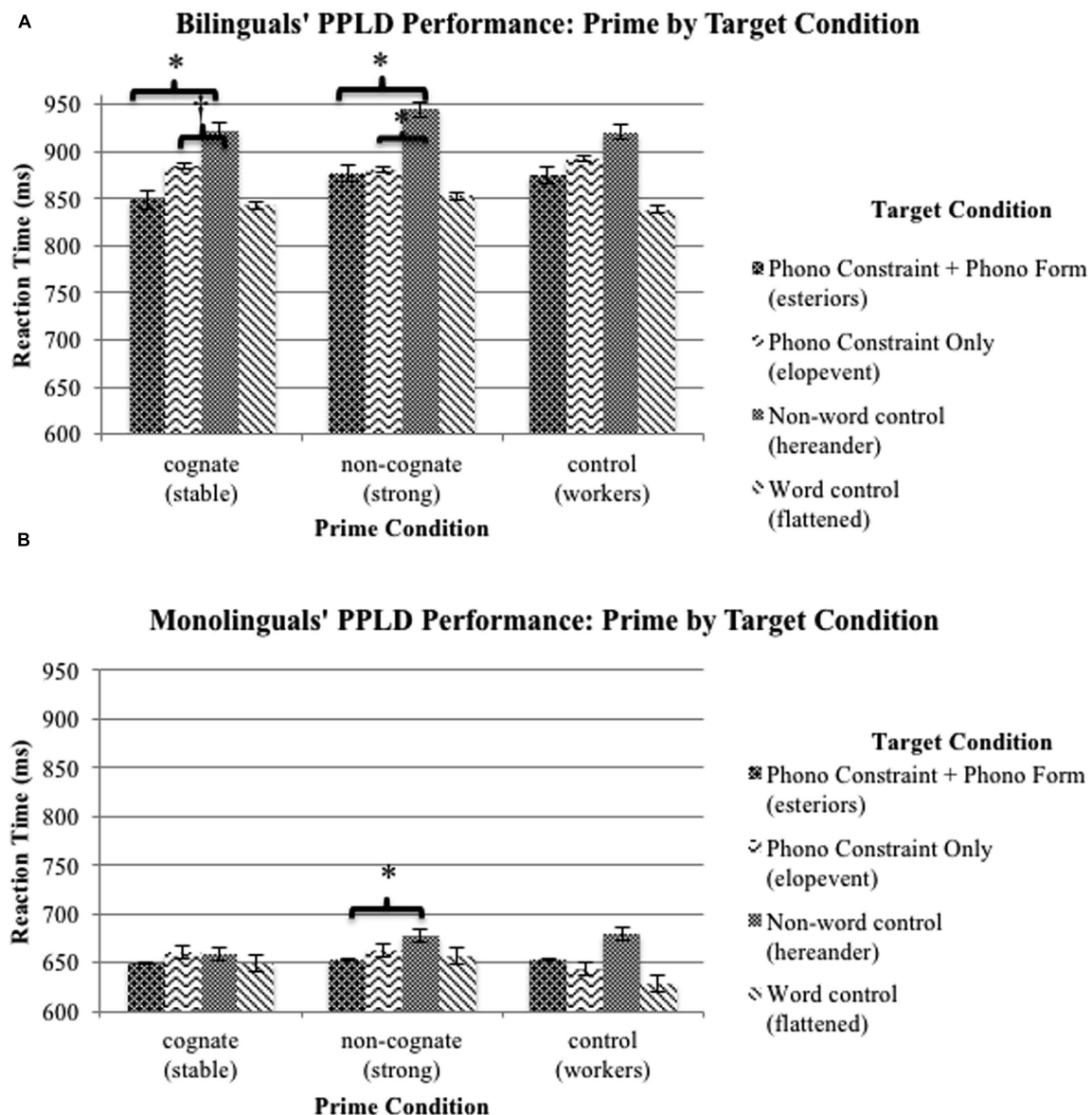


FIGURE 3 | Reaction times (RTs) on the cross-modal PPLD task for bilinguals (A) and monolinguals (B) by condition. Error bars = standard error. Differences marked for primary conditions of interest: * $p < 0.05$; † $p = 0.082$.

The results within the bilingual group demonstrate significant effects of Spanish phonotactic constraint activation during English comprehension. Bilinguals demonstrated faster reaction times, relative to control conditions, to PCF overlap non-words when primed with cognates, as well as faster reaction times to PCF overlap non-words and PC overlap non-words when primed with non-cognates.

DISCUSSION

Our goal was to explore whether bilinguals accessed phonotactic constraints from the irrelevant language (Spanish) during

English-only receptive language processing. Participants heard English words that were chosen to enhance cross-linguistic phonological activation (cognates: *stable*), that did not provide cross-linguistic phonological activation beyond the shared word onset (non-cognates: *strong*), or that were non-facilitatory of Spanish /es/ or /e/ words (controls: *workers*). Immediately after hearing the auditory prime, participants performed a lexical decision on either (1) an English-like non-word that corresponded to Spanish via phonotactic constraint (epenthesis, /e/) and form (/s/) overlap (/es/ non-words: *esteriors*), (2) PC overlap (/e/ non-words: *elopevent*), (3) on an English-like non-word that did not correspond to Spanish phonotactic constraints or form (non-word controls: *hereander*), (4) or

on a real-word control (*flattened*). Both monolinguals' and bilinguals' performance patterns were consistent with co-activation of phonologically similar representations. That is, both monolinguals and bilinguals showed facilitated responses to constraint-and-form overlap non-words. However, bilinguals displayed patterns of parallel language activation based on phonological form and/or constraint overlap, as demonstrated by significant reaction time differences to PCF overlap non-words when primed by both cognates and non-cognates and PC overlap non-words when primed with non-cognates compared to control conditions. See **Table 3** for a summary of results.

Non-target Language Phonotactic Constraint Access via Non-cognates

We aimed to tease apart PCF access in the presence (cognate primes) and absence (non-cognate primes) of previous cross-linguistic activation. With monolinguals, we expected to see a small amount of priming, as there was English phonological overlap between the prime and target conditions of interest. Critically, bilinguals but not monolinguals were found to activate the Spanish epenthesis constraint with PCF and PC overlap non-word targets when primed with English non-cognate words that had s+ phonology onsets. This finding suggests that proficient Spanish–English bilinguals may activate phonotactic constraints from their L1 when listening to English words.

Non-target Language Phonotactic Constraint Access via Cognates

There were no significant differences across the cognate prime and non-word target conditions for monolinguals. Bilinguals, however, appeared to have accessed the Spanish phonotactic constraint when primed with cognates, but that access was limited to PCF overlap non-word trials; the effect for PC overlap non-word trials was only marginally significant. This finding is consistent with previous results of bilingual parallel language activation in the presence of cognate words (e.g., Blumenfeld and Marian, 2007; Shook and Marian, 2013). Yet contrary to previous findings and expectations (e.g., Blumenfeld and Marian, 2007; Shook and Marian, 2013), cognates were found to facilitate cross-linguistic access to phonotactic constraints to a lesser extent than did non-cognates. The finding that non-cognates independently activated bilinguals' Spanish via phonotactic constraint and phonological form overlap suggests that lexico-semantic activation of the non-target language (via cognate primes) is not needed to facilitate Spanish phonotactic

constraints. Instead, phonological form overlap alone (via non-cognate primes) may consistently activate Spanish phonotactic constraints.

Taken together, the current findings suggest that Spanish–English bilinguals may activate a phonological epenthesis constraint in the non-target language (e.g., the constraint of adding an /e/ to the onset of an s+ consonant cluster) during comprehension when primed by *non-cognates*, with smaller but similar effects for *cognates*. This finding is at odds with initial predictions that a phonotactic constraint activation effect would be stronger with cognate primes, as cognate processing yields broader activation of the lexico-semantic and phonological system across both languages (Dijkstra and van Heuven, 2002; Shook and Marian, 2013). However, preliminary conclusions can be drawn from the current findings based on the cognate and non-cognate differences we observed. While it is believed that cognates, compared to non-cognates, increase co-activation of the two languages, bilinguals may need to work harder to protect from cross-linguistic competition resulting from cognates. In the current study, enhanced parallel language activation may result in an increased likelihood of intrusion from non-target language phonotactic constraints. For example, when a bilingual makes a decision on whether a string of letters forms a word, or when s/he produces a word when cross-linguistic competition (i.e., cognates) is present, s/he may emphasize language-specific plans in her response to help resolve competition. Consistently, Nip and Blumenfeld (2015) found that production of cognate sentences was associated with a greater range of speech articulator movements than non-cognate sentences in the L1 of L2 learners. Greater ranges of movement have been associated with more detailed phonological specification (Lindblom, 1990), suggesting more care in the precise articulation of the target language. Thus, across both comprehension and production, the presence of cognates may necessitate muting of phonotactic constraints from the non-target language so that bilinguals can use language-specific plans. With non-cognates, such muting is not necessary, likely due to decreased amounts of cross-linguistic competition. This preliminary conclusion is in line with the prediction that more cognitive resources may be required to inhibit the non-target language during cognate word processing (Green, 1998).

Implications for Current Accounts of Parallel Activation

The findings from this study suggest parallel activation of phonotactic constraints across two languages and are

TABLE 3 | Summary of results for bilinguals and monolinguals.

	Cognate prime Phonotactic constraint + form target: <i>stable/esteriors</i>	Cognate prime Phonotactic constraint only target: <i>stable/elopevent</i>	Non-cognate prime Phonotactic constraint + form target: <i>strong/estimagle</i>	Non-cognate prime Phonotactic constraint only target: <i>strong/encimpass</i>
Bilinguals	✓	†	✓	✓
Monolinguals	–	–	✓	–

† = Marginally significant difference, ✓ = Significant difference. Conditions of interest compared to the non-word control condition.

thus consistent with previous research demonstrating parallel activation of phonological (Marian and Spivey, 2003; Blumenfeld and Marian, 2007, 2013; Mercier et al., 2014) and lexico-semantic (e.g., Martín et al., 2010) cohorts in bilinguals during auditory and visual word processing. The current study adds to the existing bilingual language comprehension literature an additional level within cross-linguistic phonological access, the phonotactic constraint. As such, this study complements bilingual language production research that suggests bilinguals access phonotactic constraints from the non-target language (e.g., Yavas and Someillan, 2005). Furthermore, these results highlight the additional linguistic competition that bilinguals manage, relative to monolinguals, during language processing: while monolinguals demonstrated minimal interference between the primes and targets across conditions, suggesting activation of phonological representations within-language, bilinguals experienced activation from the non-target language, at the levels of both phonotactic constraint and phonological form competition.

Moreover, using the existing framework from models of bilingual language comprehension, we can extend current explanations of parallel language activation in bilinguals to incorporate the findings of the current project. Two models of bilingual language comprehension, the Bilingual Language Interaction Network for Comprehension of Speech model (BLINCS; Shook and Marian, 2013) and the Bilingual Interactive Activation + model (BIA+; Dijkstra and van Heuven, 2002), suggest that bilinguals activate both languages in parallel during single language comprehension. While both of these comprehension models do posit language co-activation based on phonology (e.g., English: *plug*, Spanish competitor: *pluma*, or *pen*), no specific claims are made about phonotactic constraint access of the non-target language.

Within the BLINCS model, bilinguals are thought to access both of their languages across various interconnected levels of processing, including phonological, phono-lexical, ortho-lexical, and semantic representations. The levels rely on a network of self-organizing maps, which provide an algorithm for learning. With activation of cross-linguistic phonological representations during comprehension, as auditory input unfolds through time, the input is first mapped onto the closest node that best matches the target (e.g., language co-activation of translation equivalents, English: *strong*/Spanish: *fuerte*), and the node is altered to become more similar to the input. Based on current findings, we can extend the BLINCS model by suggesting that nearby nodes, which include words that activate words consistent with non-target language phonotactic constraints (e.g., English: *strong*/Spanish: *edad*), might then be adapted to become more similar to the input. The space around the input, containing words following similar phonological patterns, becomes more uniform as the target word is selected. The BLINCS model also has the potential to explain the differences in processing observed across cognate and non-cognate prime conditions and non-target language phonotactic constraint access. It is possible that when bilinguals process cognates, neighboring words following the /e/ epenthesis constraint are more quickly activated than when

processing non-cognates. Over time, the cognate neighbors are suppressed as the target word is reached for selection. When processing non-cognates that activate the /e/ epenthesis constraint, neighbors also become activated, however, target word selection may take longer due to the lack of lexico-semantic overlap. Thus, stronger effects of non-target language phonotactic constraint activation may emerge when processing non-cognates.

Like the BLINCS model, the BIA+ model of bilingual written word recognition (Dijkstra and van Heuven, 2002) supports language non-selectivity (integrated bilingual lexicon) and spreading activation of cross-linguistic phonological neighbors during bilingual language comprehension. The BIA+ model states that when orthographic representations become active, associated within- and between-language phonological representations start to become activated as well. However, the model does not account for how and if phonotactic constraints from the irrelevant language are accessed, which is what was observed in the current study. As non-target language phonotactic constraints become active, so too phonological neighbors may become active that include cohorts of both languages (e.g., English and Spanish). For example, English *strong* may activate an intermediate form where the epenthesis constraint is applied, *estrong*, which may in turn co-activate Spanish words that overlap in phonological form (e.g., *estar/edad*, English: to be/age). It is thus possible that phonotactic constraint cohort members from the irrelevant language may be activated during visual word processing in addition to non-target language orthographic, and phonological cohorts. Both the BIA+ and BLINCS models can be minimally extended to provide a theoretical framework to account for parallel activation of phonotactic constraints across languages in bilinguals.

Limitations and Future Directions

The PCF overlap (/es/) non-words used in the current study could have facilitated global activation of Spanish throughout the entire task, as the non-words were Spanish-like in form. However, this was likely not the case since we provided an additional condition in which irrelevant-language phonotactic constraint access was possible, the PC overlap (/e/ non-words) condition. Including the two conditions allowed us to dissociate between phonotactic constraint and phonological form overlap with Spanish. Indeed, we found that when primed with non-cognates, bilinguals accessed the /e/ onset phonotactic constraint when making a lexical decision on the PC overlap targets. This effect was also marginally significant with cognate primes. Therefore, we can rule out that Spanish was activated only in the PCF condition, based on the evidence from the PC overlap condition. Relatedly, the finding that effects on /e/ and /es/ non-word targets were present only when directly preceded by an /sp/ and /st/ prime (and not control primes) suggests that there was no global activation of /e/ and /es/ phonology across the entire task. Finally, bilinguals, but not monolinguals, showed a significant effect for the PC condition when primed with non-cognates.

Future research is needed to further explore the possibility that Spanish–English bilinguals perceptually repair L2 auditory input

(i.e., primes such as *stable*) to have an /e/ onset, as has been shown on a Spanish-language task in Spanish monolinguals (Hallé et al., 2008). If bilinguals experienced a perceptual illusion of repairing the auditory prime to “e-stable” (/esteɪbəl/), this would also be suggestive of access to the phonotactic epenthesis constraint in the L1. While perceptual repair remains an alternative explanation to the current results, this alternative explanation is also consistent with the hypothesis of cross-linguistic activation. Thus, while the present study provides evidence that bilinguals access phonotactic rules from the non-target language during comprehension, whether the underlying mechanism(s) is constraint activation or perceptual repair remains an open question.

The contrast identified here between non-cognate and cognate words suggests that language selection mechanisms during phonotactic constraint competition also warrant further examination. For example, research might identify the time course of non-target language phonotactic constraint access (i.e., duration of L1 interference in an L2 context) during language comprehension, which will shed light on mechanisms involved with activation and suppression of non-target language phonotactic constraints. In addition, our findings showed effects of non-target language phonotactic constraint access with /es/ or /e/ onset non-word targets, not across actual English and Spanish words. We believe our results have clear implications for theoretical models of bilingual language comprehension, though stronger evidence for cross-linguistic activation of phonotactic constraints would be provided by a replication study using actual English and Spanish word targets. Moreover, varying the age of acquisition of the L2 (e.g., earlier than 5) will elucidate whether simultaneous versus sequential bilinguals experience phonotactic constraint access to a similar degree.

Finally, future studies may test different sets of language-specific phonotactic constraints to examine whether such constraints are generally accessible across languages. For example, Spanish does not permit consonant clusters at the end of words, and oftentimes native Spanish speakers reduce final consonant clusters when speaking English (e.g., *soun* for *sound*). As is the case in cross-linguistic co-activation of phonological representations (e.g., Marian and Spivey, 2003; Blumenfeld and Marian, 2007, 2013), it is possible that phonotactic constraints are especially likely to become co-activated across languages when they are specific to the dominant language. Furthermore, such constraints may become active cross-linguistically in contexts where the less dominant

language violates a phonotactic constraint in the native language.

CONCLUSION

To conclude, results from the current study demonstrate that Spanish–English bilinguals access Spanish phonotactic constraints during English *comprehension*. Moreover, bilinguals’ access to structures across both languages during spoken word comprehension is not limited specifically to phonology, but also applies to phonotactic constraints. Finally, the degree of phonological and semantic overlap across languages, as manipulated in cognate vs. non-cognate words, may modulate the extent to which cross-linguistic constraints are available, thus providing further support that the bilingual language system is highly interactive and dynamic.

AUTHOR CONTRIBUTIONS

MF, HB, and VM are responsible for the conception, design of the study, as well as for interpretation of the data, drafting the work and revising it critically for intellectual content, and final approval of the version to be published. MF and VM are responsible for data acquisition and MF is responsible for data analysis. MF, HB, and VM are in agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

ACKNOWLEDGMENTS

We would like to thank members of the *Bilingualism and Psycholinguistics Research Group* at Northwestern University for their input on this work. We would also like to thank Amanda Kellogg for recording the auditory stimuli. This project was supported by grant NICHD 1R01HD059858 to the third author.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00702>

REFERENCES

- Amengual, M. (2012). Interlingual influence in bilingual speech: cognate status effect in a continuum of bilingualism. *Bilingualism* 15, 517–530. doi: 10.1017/S1366728911000460
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The english lexicon project. *Behav. Res. Methods* 39, 445–459. doi: 10.3758/BF03193014
- Bartolotti, J., and Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cogn. Sci.* 36, 1129–1147. doi: 10.1111/j.1551-6709.2012.01243.x
- Blumenfeld, H. K., and Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: examining proficiency and lexical status using eye-tracking. *Lang. Cogn. Process.* 2, 633–660. doi: 10.1080/01690960601000746
- Blumenfeld, H. K., and Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *J. Cogn. Psychol.* 25, 547–567. doi: 10.1080/20445911.2013.812093
- Blumenfeld, H. K., and Marian, V. (2014). Cognitive control in bilinguals: advantages in stimulus–stimulus inhibition. *Bilingualism* 17, 610–629. doi: 10.1017/S1366728913000564

- Boersma, P., and Weenink, D. (2013). *Praat: Doing Phonetics by Computer [Computer Program]*. Version 5.3. Available at: <http://www.praat.org/> [accessed on 23 April, 2013].
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357
- Broersma, M., and Cutler, A. (2011). Competition dynamics of second-language listening. *Q. J. Exp. Psychol.* 64, 74–95. doi: 10.1080/17470218.2010.499174
- Carlson, M. T., Goldrick, M., Blasingame, M., and Fink, A. (2015). Navigating conflicting phonotactic constraints in bilingual speech perception. *Biling. Lang. Cogn.* 1–16. doi: 10.1017/S1366728915000334
- Darcy, I., Park, H., and Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: the relation between cognitive abilities and phonological processing. *Learn. Individ. Differ.* 40, 63–72. doi: 10.1016/j.lindif.2015.04.005
- Davis, C., Sánchez-Casas, R., García-Albea, J. E., Guasch, M., Molero, M., and Ferré, P. (2010). Masked translation priming: varying language experience and word type with Spanish English bilinguals. *Bilingualism* 13, 137–155. doi: 10.1017/S1366728909990393
- de Groot, A. M. B., Delmar, P., and Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: support for non-selective access to bilingual memory. *Q. J. Exp. Psychol.* 53A, 397–428. doi: 10.1080/713755891
- Dijkstra, T., Grainger, J., and Van Heuven, W. J. (1999). Recognition of cognates and interlingual homographs: the neglected role of phonology. *J. Mem. Lang.* 41, 496–518. doi: 10.1006/jmla.1999.2654
- Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., and Bayeen, H. (2010). How cross language similarity and task demands affect cognate recognition. *J. Mem. Lang.* 62, 284–301. doi: 10.1016/j.jml.2009.12.003
- Dijkstra, T., and van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: from identification to decision. *Bilingualism* 5, 175–197. doi: 10.1017/S1366728902003012
- Dupoux, E., Pallier, C., Kakehi, K., and Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Lang. Cogn. Process.* 16, 491–505. doi: 10.1080/01690960143000191
- Duyck, W., Assche, E., Drieghe, D., and Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: evidence for nonselective lexical access. *J. Exp. Psychol.* 33, 663–679. doi: 10.1037/0278-7393.33.4.663
- Fabra, L. R., and Romero, J. (2012). Native Catalan learners' perception and production of English words. *J. Phonet.* 40, 491–508. doi: 10.1016/j.wocn.2012.01.001
- Flege, J. E., and Davidian, R. D. (1984). Transfer and developmental processes in adult foreign language speech production. *Appl. Psycholinguist.* 5, 323–347. doi: 10.1017/S014271640000521X
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism* 1, 67–81. doi: 10.1017/S1366728998000133
- Hallé, P. A., Dominguez, A., Cuetos, F., and Segui, J. (2008). Phonological mediation in visual masked priming: evidence from phonotactic repair. *J. Exp. Psychol.* 34:177. doi: 10.1037/0096-1523.34.1.177
- Kleiner, M., Brainard, D., and Pelli D. (2007). *What's new in Psychtoolbox-3? Perception 36 ECVF Abstract Supplement*. Available at: <http://www.perceptionweb.com/abstract.cgi?id=v070821> [accessed December 16, 2013].
- Kootstra, G. J., Van Hell, J. G., and Dijkstra, T. (2012). Priming of code-switches in sentences: the role of lexical repetition, cognates, and language proficiency. *Bilingualism* 15, 797–819. doi: 10.1017/S136672891100068X
- Kroll, J. F., Bobb, S. C., Misra, M., and Guo, T. (2008). Language selection in bilingual speech: evidence for inhibitory processes. *Acta Psychol.* 128, 416–430. doi: 10.1016/j.actpsy.2008.02.001
- Linck, J. A., Hoshino, N., and Kroll, J. F. (2008). Cross-language lexical processes and inhibitory control. *Mental Lexicon* 3, 349–374. doi: 10.1075/ml.3.3.06lin
- Linck, J. A., Kroll, J. F., and Sunderman, G. (2009). Losing access to the native language while immersed in a second language: evidence for the role of inhibition in second-language learning. *Psychol. Sci.* 20, 1507–1515. doi: 10.1111/j.1467-9280.2009.02480.x
- Lindblom, B. (1990). On the notion of possible speech sound. *J. Phonet.* 18, 135–152.
- Marian, V., Bartolotti, J., Chabal, S., and Shook, A. (2012). CLEARPOND: cross-linguistic easy access resource for phonological and orthographic neighborhood densities. *PLoS ONE* 7:e43230. doi: 10.1371/journal.pone.0043230
- Marian, V., Blumenfeld, H., and Boukrina, O. (2008). Sensitivity to phonological similarity within and across languages. *J. Psycholinguist. Res.* 37, 141–170. doi: 10.1007/s10936-007-9064-9
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007)067
- Marian, V., and Spivey, M. (2003). Competing activation in bilingual language processing: within- and between-language competition. *Bilingualism* 6, 97–115. doi: 10.1017/S1366728903001068
- Martin, M. C., Macizo, P., and Bajo, T. (2010). Time course of inhibitory processes in bilingual language processing. *Br. J. Psychol.* 101, 679–693. doi: 10.1348/000712609X480571
- Mercier, J., Pivneva, I., and Titone, D. (2014). Individual differences in inhibitory control relate to bilingual spoken word processing. *Bilingualism* 17, 89–117. doi: 10.1017/S1366728913000084
- National Institutes of Health Toolbox Cognition Battery (Nih Toolbox CB) (2013). *Monographs of the Society for Research in Child Development*, Vol. 78. (Hoboken, NJ: Wiley Online Library), 1–172.
- Nip, I. S., and Blumenfeld, H. K. (2015). Proficiency and linguistic complexity influence speech motor control and performance in Spanish language learners. *J. Speech Lang. Hear. Res.* 58, 653–668. doi: 10.1044/2015_JSLHR-S-13-0299
- Parlato-Oliveira, E., Christophe, A., Hirose, Y., and Dupoux, E. (2010). Plasticity of illusory vowel perception in Brazilian-Japanese bilinguals. *J. Acoust. Soc. Am.* 127, 3738–3748. doi: 10.1121/1.3327792
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366
- PsychCorp. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Harcourt Assessment.
- Roelofs, A., and Verhoeve, K. (2006). Modeling the control of phonological encoding in bilingual speakers. *Bilingualism* 9, 167–176. doi: 10.1017/S1366728906002513
- Shook, A., and Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism* 16, 304–324. doi: 10.1017/S1366728912000466
- Siyambalapitiya, S., Chenery, H. J., and Copland, D. A. (2009). Bilingualism and aging: reversal of the cognate advantage in older bilingual adults. *Appl. Psycholinguist.* 30, 531–554. doi: 10.1017/S0142716409090262
- Van Hell, J. G., and Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychon. Bull. Rev.* 9, 780–789. doi: 10.3758/BF03196335
- Woodcock, R. W., McGrew, K. S., and Mather, N. (2001/2007). *Woodcock Johnson III Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside Publishing.
- Yavas, M., and Someillan, M. (2005). Patterns of acquisition of s/-clusters in Spanish-English bilinguals. *J. Multiling. Commun. Disord.* 3, 50–55. doi: 10.3109/02699200903362919

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Freeman, Blumenfeld and Marian. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Cognate Costs in Bilingual Speech Production: Evidence from Language Switching

Mirjam Broersma^{1,2*}, Diana Carter^{3,4} and Daniel J. Acheson^{2,5}

¹ Centre for Language Studies, Radboud University, Nijmegen, Netherlands, ² Comprehension Group, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, ³ Department of Critical Studies, University of British Columbia, Kelowna, BC, Canada, ⁴ Centre for Research on Bilingualism, Bangor University, Bangor, UK, ⁵ Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

OPEN ACCESS

Edited by:

Isabelle Darcy,
Indiana University Bloomington, USA

Reviewed by:

Elin Runnqvist,
Laboratoire Parole & Langage (CNRS),
France

Monika S. Schmid,
University of Essex, UK

*Correspondence:

Mirjam Broersma
m.broersma@let.ru.nl;
mirjam@mirjambroersma.nl

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 12 February 2016

Accepted: 12 September 2016

Published: 28 September 2016

Citation:

Broersma M, Carter D and
Acheson DJ (2016) Cognate Costs in
Bilingual Speech Production:
Evidence from Language Switching.
Front. Psychol. 7:1461.
doi: 10.3389/fpsyg.2016.01461

This study investigates cross-language lexical competition in the bilingual mental lexicon. It provides evidence for the occurrence of inhibition as well as the commonly reported facilitation during the production of cognates (words with similar phonological form and meaning in two languages) in a mixed picture naming task by highly proficient Welsh-English bilinguals. Previous studies have typically found cognate facilitation. It has previously been proposed (with respect to non-cognates) that cross-language inhibition is limited to low-proficient bilinguals; therefore, we tested highly proficient, early bilinguals. In a mixed naming experiment (i.e., picture naming with language switching), 48 highly proficient, early Welsh-English bilinguals named pictures in Welsh and English, including cognate and non-cognate targets. Participants were English-dominant, Welsh-dominant, or had equal language dominance. The results showed evidence for cognate inhibition in two ways. First, both facilitation and inhibition were found on the cognate trials themselves, compared to non-cognate controls, modulated by the participants' language dominance. The English-dominant group showed cognate *inhibition* when naming in Welsh (and no difference between cognates and controls when naming in English), and the Welsh-dominant and equal dominance groups generally showed cognate *facilitation*. Second, cognate inhibition was found as a *behavioral adaptation* effect, with slower naming for non-cognate filler words in trials *after* cognates than after non-cognate controls. This effect was consistent across all language dominance groups and both target languages, suggesting that cognate production involved cognitive control even if this was not measurable in the cognate trials themselves. Finally, the results replicated patterns of symmetrical switch costs, as commonly reported for balanced bilinguals. We propose that cognate processing might be affected by two different processes, namely competition at the lexical-semantic level and facilitation at the word form level, and that facilitation at the word form level might (sometimes) outweigh any effects of inhibition at the lemma level. In sum, this study provides evidence that cognate naming can cause costs in addition to benefits. The finding of cognate inhibition, particularly for the highly proficient bilinguals tested, provides strong evidence for the occurrence of lexical competition across languages in the bilingual mental lexicon.

Keywords: bilingual speech production, cognates, language switching, cross-language inhibition, lexical competition, behavioral adaptation

INTRODUCTION

A fascinating capacity of the human mind is the ability to cope with several languages and to use those languages to perform exactly the activity that the speaker intends. Multilingual speakers can choose to speak one language without noteworthy intrusion from the other language (Poulisse, 1999), they can translate between languages, and they can codeswitch, i.e., use several languages within one conversation or sentence. How speakers manage to select lexical items from one language rather than the other is a question that has received much attention from the domain of cognitive psychology (e.g., Jackson et al., 2001; Costa and Santesteban, 2004; Christoffels et al., 2007; Philipp et al., 2007; Verhoef et al., 2009). This paper further addresses that question by investigating the way in which cognates—words which are similar in meaning and phonological form in two languages—are processed in the bilingual mental lexicon. In particular, this paper investigates the occurrence of *inhibition* during the production of cognates by highly proficient bilinguals, which would suggest that the lexical-semantic nodes (or lemmas) of the cognates compete with each other for selection in the bilingual mental lexicon. It presents the results of a *mixed naming experiment* (i.e., picture naming with language switching) with highly proficient early Welsh-English bilinguals. We investigate naming latencies for cognates compared to non-cognate controls. Importantly, in contrast to previous studies, we also investigate the effect of cognate status on naming latencies in the *following* trial, in search of a possible *behavioral adaptation effect*. In this way, we aim to make effects of inhibition visible that might not be visible otherwise.

Despite controversy about various aspects of the bilingual word production process, there is a general consensus among psycholinguists about two things. First, a common view is that lemmas, when activated by concepts, compete for selection, for bilingual and monolingual speakers alike. A second common view is that when bilinguals speak one language, lexical representations from both languages are activated (e.g., De Bot, 1992; Green, 1998; Costa and Caramazza, 1999; Costa and Santesteban, 2004, 2006; Finkbeiner et al., 2006; Kroll et al., 2006, 2008; Abutalebi and Green, 2007; Branzi et al., 2014). There are opposing views, however, on how speakers manage to produce unilingual speech and to avoid selecting words from the unintended language, in particular with respect to the occurrence of inhibitory control. First, there are models that propose inhibitory control to be a compulsory mechanism of bilingual lexical selection. Language *non-specific* models of lexical selection (e.g., De Bot, 1992; Green, 1998; Kroll et al., 2006; Abutalebi and Green, 2007) posit that words from both languages compete for selection. Such models generally assume that this cross-language competition eventually leads to the inhibition of the non-selected words. (See however, Runnqvist et al., 2012 for a language non-specific model that does not entail inhibitory control). Second, there are models that propose that there is no inhibitory control across languages. Language-*specific* models of lexical selection (e.g., Costa and Caramazza, 1999; Finkbeiner et al., 2006), posit that, even though lexical items from both languages are activated, only those from the intended language

are considered for selection; hence, words from the two languages do not compete for selection. Third, the occurrence of inhibitory control has been proposed to depend on the proficiency of the speaker: more proficient bilinguals might not need the use of inhibitory control as they might access the lexicon in a language-specific way, whereas less proficient bilinguals might rely on cross-language inhibition to suppress words in the first language (L1) or in the dominant language when speaking in the second (L2) or less dominant language (Costa and Santesteban, 2004; Branzi et al., 2014). Others have argued, in contrast to this view, that even highly proficient bilinguals rely on inhibitory processes to avoid selection of lexical items from the unintended language (e.g., Kroll et al., 2008). Fourth, inhibitory control has been proposed to be an optional mechanism that the same bilinguals might or might not apply depending on the specifics of the task, such as the amount of preparation time and the type of distracters used (Verhoef et al., 2009; Roelofs et al., 2011).

In this paper we will use cognates to investigate inhibitory control in bilinguals. Interestingly, only few studies have suggested a role for inhibition during the production of cognates (as described below); the large majority of the experimental evidence instead points to the occurrence of *facilitation* during the processing of cognates. Advantages for cognate over non-cognate processing have been found with a variety of experimental paradigms, both in speech perception and in speech production, and in the visual and auditory modality. For example, in previous picture naming experiments, cognates were named faster than non-cognates (Costa et al., 2000; Christoffels et al., 2006, 2007; Hoshino and Kroll, 2008; Verhoef et al., 2009) and led to fewer tip-of-the-tongue experiences than non-cognates (Gollan and Acenas, 2004). Visually presented cognates are recognized faster and more accurately than non-cognates in lexical decision in the participants' L1 (Van Hell and Dijkstra, 2002) and L2 (Dijkstra et al., 1999; Lemhöfer and Dijkstra, 2004). Also in reading, masked associative priming between languages occurs for cognates but not for non-cognates (De Groot and Nas, 1991), and between-language repetition priming is larger for cognates than for non-cognates (De Groot and Nas, 1991; Gollan et al., 1997). In word association tasks, participants produce associates to cognates faster than to non-cognates both in cross-linguistic and L1-only tasks (Van Hell and De Groot, 1998; Van Hell and Dijkstra, 2002). In speech production, cognates are translated faster than non-cognates (Kroll and Stewart, 1994; Christoffels et al., 2006). Differences between cognate and non-cognate naming observed with ERPs show that cognates behave in some respect as high-frequency words (Strijkers et al., 2010); early effects, with the ERPs of cognates eliciting a smaller positivity than non-cognates at the P2 area, and later effects, with a more enhanced negativity for cognates than non-cognates in the N3 area, which are both similar to the ERP modulations for high vs. low-frequency words, have been interpreted as effects of lexical and phonological processing, respectively (Christoffels et al., 2007; Strijkers et al., 2010). This processing advantage for cognates over non-cognates is commonly ascribed to the activation of conceptual and form representations in both languages; e.g., in the case of word naming, when the activation of both lemmas spreads to the word form level, the similarity at

the word form level enhances the activation of the two word form representations, as the overlapping parts receive input from both lemmas.

Interestingly, a cognate *inhibition* effect was found in a word naming (i.e., reading aloud) language switching experiment (Filippi et al., 2014). In that study, late Italian-English bilinguals read words from a computer screen while a color cue indicated the target language (L1 Italian or L2 English). Cognates were produced more slowly than non-cognate control words. These findings could point to lexical competition between the cognates' lemmas. Further, evidence for the occurrence of inhibitory control during cognate production comes from an EEG study of bilingual picture naming (Acheson et al., 2012) which showed that bilingual speakers recruited domain-general control operations during the production of cognates. Whereas cognates were named faster than non-cognates, they were also found to induce response conflict, which showed in the form of an increased error-related negativity (ERN)-like effect, where cognates were more negative than non-cognates. Furthermore, a behavioral adaptation effect was observed, on correctly named trials, as the magnitude of the cognate facilitation effect was smaller following the naming of a cognate relative to a matched non-cognate. The authors reasoned that despite being faster to name, cognates also induced more response conflict as speakers must mediate between two very closely related pronunciations.

We propose that there might be two different processes at work during the lexical selection of cognates—competition at the lexical-semantic level, and facilitation at the word form level—the latter of which might often obscure the former. This makes it difficult to determine whether the lemmas of cognates compete for selection, as the benefit of the activation of shared word forms might outweigh the possible slowing effect of lexical competition at the lemma level¹.

Let's consider the example of the English–Welsh cognates *balloon*—*balŵn*, which are (despite their difference in spelling) pronounced the same. According to the common view described above, the speaker's intention to express a certain meaning should lead to the activation of the lemmas of both cognates. If a Welsh-English bilingual speaker wishes to speak about a balloon in Welsh, she will thus activate both the Welsh and the English lemma. According to models that do not assume the

occurrence of cross-language inhibition, even though the English “balloon” lemma is activated, it will not affect the lexical selection process, and only the “balŵn” lemma in the intended language Welsh will be considered for selection. According to inhibitory control models of lexical selection, on the other hand, the two lemmas will compete for selection. In that scenario, eventually one lemma would win the competition and inhibit its competitor; in the example, “balŵn” would be expected to win, being in the intended language, and to inhibit the lemma of “balloon.” Further, whether or not inhibition occurs might depend on the proficiency of the bilingual speaker (Costa and Santesteban, 2004; Branzi et al., 2014) or on the circumstances of the test (Verhoeef et al., 2009; Roelofs et al., 2011). Whereas in inhibitory control models of lexical selection all translation equivalents are expected to compete, it is conceivable that such competition might be stronger for cognates than for non-cognate translation pairs: It has been proposed that feedback loops from the level of phonological representations to the lemma level increase the activation of lemmas that share aspects of their form, resulting in increased lexical competition between such lemmas (Declerck and Philipp, 2015). As form overlap is maximal for cognates, so might the activation of the unintended lemma be during the lexical selection of cognates. Such feedback would enhance the activation of both the intended and the unintended lemma, and the outcome might thus either be facilitation or inhibition at the lemma level. Arguably, however, the unintended lemma might have more to gain than the intended lemma, such that the net result at the lemma level might (at least sometimes) be inhibition (compared to non-form-overlapping translation pairs where the competitor does not receive such feedback). Further, facilitation at the word form level might (sometimes) outweigh any effects of inhibition at the lemma level.

In this paper, we investigate whether we can find evidence for inhibition during bilinguals' production of cognates in a mixed picture naming task, which requires participants to switch between their languages during picture naming. Below, we describe the mixed naming paradigm, and common findings obtained with the paradigm, in some detail. Importantly, we investigate the effect of cognate status not only in the trial containing a cognate or non-cognate control word, but also in the trials immediately *following* the experimental cognates and control words. A long tradition of studies in the non-linguistic domain have shown that response conflict in the *preceding* trial can modulate performance in the following trial (Gratton et al., 1992), e.g., in the Eriksen Flanker task (Eriksen and Eriksen, 1974). This has been explained as the result of conflict monitoring (Botvinick et al., 2001; Yeung et al., 2004), as described below in more detail. If the production of cognates thus attracts inhibitory control, we would predict to find longer naming latencies after cognates than after controls in the non-cognate filler words in the next trial.

This study investigates the occurrence of cognate inhibition in highly proficient, early bilinguals. As mentioned above, the level of proficiency that speakers have in their two languages has been found to affect the occurrence of cross-language inhibition. In previous naming studies, both blocked by language and mixed, unbalanced bilinguals with a lower level of proficiency in one

¹This argument is similar in nature to one that has been made for the possible competition between non-cognate translation equivalents during picture-word interference experiments. Picture naming is slowed by the presence of a semantically related competitor word (compared to an unrelated control), also when the competitor is in a language other than the one that is used for naming; when the competitor word is the translation equivalent of the picture name, however, picture naming is not slowed but rather speeded (Costa and Caramazza, 1999; Costa et al., 1999; Hermans, 2004; Hermans et al., 1998). This is somewhat surprising, because, as Abutalebi and Green (2007) observed, translation equivalents are arguably the strongest possible lexical competitors and should be expected to create strong interference. The speeding effect has been ascribed, however, to the fact that providing the translation equivalent of the picture name will contribute to the recognition of the picture, and therefore facilitate picture naming, an advantage which might outweigh any possible competition at the lemma level (Abutalebi and Green, 2007; Hermans, 2004). This line of reasoning is similar to our argument about cognates that a processing advantage at one level (in the case of cognates: at the word form level) might outweigh any possible competition at the lexical-semantic level.

of their languages have been found to show clearer signs of cross-language inhibition than highly proficient bilinguals, for whom some studies even have found no evidence of cross-language inhibition at all (Costa and Santesteban, 2004; Branzi et al., 2014). The present study investigates whether evidence for cognate inhibition can nevertheless be found for such highly proficient, early bilinguals, which would provide the strongest evidence for the occurrence of lexical competition across languages in the bilingual mental lexicon. To this end, we tested highly proficient, fluent Welsh-English bilinguals, who were bilingual from childhood, and who lived in Wales in the United Kingdom, a highly bilingual environment. Below, the linguistic situation in Wales is described in more detail.

Further, we will assess whether a possible cognate inhibition effect depends on language dominance. Bilingual speakers' language dominance has been shown to affect the strength of cognate facilitation effects in each language. Cognate facilitation in picture naming tests is typically much larger in the bilingual speakers' non-dominant language than in their dominant language, where it is often entirely absent (Costa et al., 2000; Christoffels et al., 2006, 2007; Gollan et al., 2007; Ivanova and Costa, 2008; Verhoef et al., 2009; Strijkers et al., 2010; Poarch and Van Hell, 2012). Similar effects of proficiency, with stronger cognate facilitation effects for less proficient languages than for more proficient languages, have been found with other experimental tasks as well (e.g., Van Hell and Dijkstra, 2002). We therefore assess whether any cognate costs might depend on language dominance, and differs between the two languages, analogous to findings for cognate facilitation in picture naming. Note that, as our participants are highly proficient in both languages, we do not have a priori expectations about the shape that such effects might take.

In summary, we investigate whether there is evidence for inhibitory control during the production of cognates as compared to non-cognate control words during picture naming in a mixed naming task by Welsh-English bilinguals. Such cognate inhibition would be evidence for cross-language lexical competition in highly proficient bilinguals. In order to investigate the occurrence of cognate inhibition, we assess the naming latencies of the cognates vs. non-cognate control words themselves (where the typically reported pattern is cognate *facilitation*), as well as the naming latencies of non-cognate filler words in the immediately following trial. Shorter naming latencies for cognates than for non-cognate controls would point to cognate facilitation, in line with the most commonly reported pattern. Longer naming latencies for cognates than for non-cognate controls, on the other hand, and longer naming latencies for non-cognate filler words after a cognate than after a non-cognate control, will be taken as evidence for cognate inhibition. We hypothesize that—possibly in addition to cognate facilitation—we will find evidence for cognate inhibition, on the cognate trials themselves and/or on the following trials. We also hypothesize that the participants' language dominance might affect the direction of the cognate effect, with some groups showing cognate facilitation, and others cognate inhibition.

Common Findings in Mixed Naming Experiments

Mixed naming, or language switching, is one of the variations of the classical task switching paradigm (for a review see, Monsell, 2003). Task switching experiments involve two competing tasks, both indicated by an arbitrary cue; in language switching experiments, the tasks are picture or number naming in language A and language B. Requiring participants to perform two tasks within the same experiment (or block) commonly leads to slower and less accurate responses than when only one task is involved. This so called *mixing cost* is found in language switching (Christoffels et al., 2007) as well as in other task switching experiments (for a review see, Los, 1996).

Further, responses are generally slower and less accurate on switch trials (where the task to be performed differs from that in the previous trial) compared to non-switch trials, in language switching (Meuter and Allport, 1999; Jackson et al., 2001; Costa and Santesteban, 2004; Christoffels et al., 2007; Verhoef et al., 2009, 2010) and in other task switching experiments (Allport et al., 1994; Rogers and Monsell, 1995; Rubinstein et al., 2001). This *switch cost* has been proposed to reflect task-set reconfiguration, i.e., disengaging from the old task and engaging in the new task (Rogers and Monsell, 1995) or task set inertia, i.e., the interference from the previous task with the new task (Allport et al., 1994; Altmann and Gray, 2008).

In mixed naming studies, an *asymmetric switch cost* for switches into the speaker's L1 and L2 is often found (Meuter and Allport, 1999; Costa and Santesteban, 2004; Campbell, 2005; Philipp et al., 2007; Wang et al., 2007; Verhoef et al., 2009). Somewhat counterintuitively, switching into the L1 entails larger costs than switching into the L2, which is often interpreted as evidence that producing words in the weaker L2 requires strong inhibition of the L1, which needs to be overcome before a switch into the L1 can take place, while producing words in the L1 does not require (as much) inhibition of the L2 (Meuter and Allport, 1999; Campbell, 2005; but see, Costa and Santesteban, 2004), an interpretation which is also supported by ERP evidence (Jackson et al., 2001; Verhoef et al., 2009). Asymmetric switch costs are not always found, and vary with the preparation interval before the switch (Verhoef et al., 2009), characteristics of the stimuli like the script in which numeral stimuli are presented (Campbell, 2005), and with the speaker's proficiency, with smaller or no asymmetries for more balanced bilinguals (Costa and Santesteban, 2004; Costa et al., 2006).

Behavioral Adaptation Effects

This study addresses the engagement of control in bilinguals by focusing on a behavioral phenomenon that is well-established in the cognitive control literature that has, to date, received little attention in bilingualism research: behavioral adaptation effects. Adaptation effects refer to behavioral modulation following the detection of conflict or errors and are thought to be hallmarks of the recruitment of a cognitive control mechanism (Botvinick et al., 2001). One such adaptation effect, post-error slowing and accuracy improvements, occurs after an explicit error is made (Rabbitt, 1966; Laming, 1979). More relevant to the present

investigation, however, are studies showing adaptation following *correct* performance on trials with high amounts of conflict, such as in the Eriksen Flanker task (Eriksen and Eriksen, 1974). In the Flanker Task, people respond with a left or right button press to a stimulus that is flanked by congruent (e.g., < < <) or incongruent (e.g., < > <) stimuli, corresponding, respectively, to low and high conflict situations. Within this task (and similar tasks such as the Simon (1969) and Stroop (1935) tasks), people are slower and less accurate for incongruent than for congruent stimuli. Importantly, the magnitude of this congruency effect is modulated by the presence of conflict in the *preceding* trial: it is smaller following a high conflict, incongruent trial than following a low conflict, congruent trial (Gratton et al., 1992). Although some researchers have accounted for these adaptation effects in the Flanker Task in terms of stimulus repetition (e.g., Mayr et al., 2003; Nieuwenhuis et al., 2006), such effects are also present in the Stroop and Simon tasks; the fact that the effect generalizes to other tasks that induce high amounts of conflict and when stimulus repetition has been controlled suggests that common mechanisms for mediating conflict may exist (e.g., Stürmer et al., 2002; Kerns et al., 2004).

An explanation of these results is specified at both the computational and neural level within the *conflict monitoring hypothesis* (e.g., Botvinick et al., 2001; Yeung et al., 2004). According to this hypothesis, a region of the medial prefrontal cortex, the dorsal ACC, serves as a detector of response conflict. The ACC, in turn, sends a signal to the DLPFC, a region which maintains current task goals and resolves conflict by sending biasing signals to task-relevant representations, thus focusing attention on the relevant rather than on the irrelevant, conflicting information. Evidence for this model comes from a number of neuroimaging studies that have shown the involvement of the ACC during high conflict trials and the recruitment of the DLPFC during conflict adaptation (Kerns et al., 2004). Thus, the conflict monitoring hypothesis provides a well-established framework in which the detection of conflict leads to the recruitment of cognitive control operations, which in turn bias the activation of task-relevant representations over task-irrelevant ones.

To date, the study of conflict adaptation effects and the subsequent recruitment of control has typically been limited to tasks that use very simple manual responding, and with the exception of the Stroop task, do not use language. In the present study we address adaptation effects after cognate naming in a picture naming language-switching task, which involves more processing steps and a wider range of motor effectors than is typically employed in the cognitive control literature. We investigate whether evidence for cross-language inhibition shows as a behavioral adaptation effect which can be measured in the trial after the crucial cognate or non-cognate control word.

The Linguistic Situation in Wales

We tested fluent, early bilinguals in English and Welsh (or *Cymraeg*), living in Bangor, Wales (UK). Wales has been officially bilingual since 1993, when the Welsh Language Act declared Welsh and English to be equal in the public sector. Wales has a stable bilingual community (Mueller Gathercole and Thomas,

2009). Both Welsh and English are present in all aspects of daily life, including the media, literature, government documents, and on signs. At the societal level (i.e., as opposed to the individual level), English is the dominant language and Welsh a minority language. There are no monolingual Welsh speakers; yet, a large number of speakers are native(-like) in both Welsh and English (Thomas and Gathercole, 2005). In the region of Gwynedd, which encompasses Bangor, 69% of the population speaks both English and Welsh, according to the 2001 UK census (Thomas and Gathercole, 2005).

Although most schools teach through the medium of English, Welsh has been a compulsory subject in primary and secondary school since 1999. In some homes, both languages are spoken; in others, only Welsh or only English is spoken. Children growing up in families where only one language is spoken are likely to overhear the other language at least occasionally. For children who have not been exposed to either Welsh or English at home, going to school is often the first systematic exposure to that language (Mueller Gathercole and Thomas, 2009). Children start attending school between the age of 4 and 5 (in the month of September after turning 4). By the age of 4½ the majority of children are acquiring both languages (Mueller Gathercole and Thomas, 2009).

While English is a West-Germanic language, Welsh is a Brythonic language, from the Celtic branch of the Indo-European language family. Due to their linguistic distance, the cognates that they share are not derived from a common root. The vast majority or possibly all Welsh-English cognates (set aside proper nouns, i.e., person or place names) consist of English borrowings into Welsh; the only possible counterexample that we are aware of is the English word *penguin*, which might be derived from Welsh *pen gwyn* (*white head*; see, e.g., Klein, 1986). E.g., in the 460,000 word Siarad corpus of Welsh-English bilingual conversations (Deuchar et al., 2014), all cognates are borrowings from English into Welsh.

METHODS

Participants

Forty-eight Welsh-English bilinguals (36 female; mean age: 25.2, range: 19–49, *SD*: 7.5) were recruited as paid volunteers among students and staff of Bangor University, Wales (UK). All reported to be balanced bilinguals, to be fluent and highly proficient in both languages, and to have started acquiring both languages before the age of seven. All were born and raised in Wales, which is highly bilingual, and lived in Wales at the time of testing.

Seventeen participants reported feeling (if only slightly) more dominant in Welsh, 17 in English, and 14 reported dominance to be equal for both languages or to be situation-dependent. According to self-report, 24 participants were exposed to Welsh from birth and to English from a mean age of 4.6 (*SD*: 2.2; which coincides with the age at which children start attending school), 11 participants to English from birth and to Welsh from a mean age of 4.6 (*SD*: 2.7), and 13 participants to both languages from birth. Self-reported language dominance and L1 were moderately correlated, $r_{(46)} = 0.56, p < 0.001$.

Materials

Stimulus words and their corresponding pictures were selected from the International Picture Naming Project (IPNP) database (Székely et al., 2004). First, 36 Cognates and 36 non-cognate Controls were selected (all of which were also originally from Snodgrass and Vanderwart, 1980; Appendix A) and grouped into pairs of one Cognate and one Control each. Cognates were phonologically identical (e.g., English: *shark* /ΣA:k/—Welsh: *siarc* /ΣA:k/), or slightly different due to differences between the English and Welsh phoneme inventories (e.g., English: *bus* /bʌs/—Welsh: *bws* /bYs/). Controls did not overlap in word form in the two languages. Eighteen Cognates and paired Controls were monosyllabic and 18 disyllabic in English and Welsh. Cognates and Controls were matched on the following 32 potentially relevant characteristics.

First, 26 variables were taken from the IPNP database. Pictures from the IPNP have been extensively tested in several languages, and 26 variables are provided based on prior studies, containing information in four categories: “Error coding” (percentage of valid, invalid, and missing responses), “Name agreement” (number of alternative names and seven measures of response agreement), “Reaction times” (seven measures), and “Features of the dominant response and picture characteristics” (nine measures including estimates of objective visual complexity, conceptual complexity, length in syllables and in characters, presence or absence of initial frication, lexical frequency, age of acquisition, word complexity). Here, the values based on a study with adult native speakers of English (Székely et al., 2003) were used.

Second, we assessed the length in phonemes in English and Welsh, the number of syllables in Welsh, and the lexical frequency in Welsh using the natural logarithm of the summed frequencies in the CEG lexical database of written Welsh (Ellis et al., 2001) and the Siarad corpus of Welsh-English bilingual conversations (Deuchar et al., 2014).

Third, in an online control experiment, estimates of subjective goodness of the match between the Welsh word and the corresponding pictures, and of subjective age of acquisition of the Welsh words were obtained. Six participants in the main experiment took part in the control experiment, after doing the main experiment, on a separate day. On each trial they were presented with a written Welsh word and the corresponding picture. In the first block they rated on a 7-point scale how well the picture depicted the word. In the second block, they indicated how old they thought they were when they first heard or read the word.

Paired sample *t*-tests showed no differences between Cognates and Controls on any of the 32 variables described above; Cognates and Controls were thus well-matched. Finally, as fillers, pictures were selected of 159 non-cognates and 18 cognates, and 10 practice items, all of one to four syllables long.

Design

The experimental Cognate/Control pairs were distributed over two lists, with equal numbers of mono-, and disyllabic items,

and presentation was counterbalanced across participants such that each participant saw either the Cognate or the Control of every pair. Each participant thus saw 18 experimental Cognates and 18 experimental Controls, as well as all fillers (totaling 195 stimuli). Items were presented in a semi-random order, such that Cognates and Controls were preceded by at least two non-cognate filler words; the immediately preceding filler was the same for matched Cognates and Controls. For all stimuli, target language was counterbalanced across participants, such that half of the participants were required to name the item in English and the other half in Welsh. Each stimulus list contained a total of 101 trials in one language and 94 in the other, and 22 language switches. The position (i.e., trial number) of language switches was the same in all lists. Language switches never occurred on an experimental Cognate or Control, or on the immediately preceding filler word, but could only occur in trials following an experimental Cognate or Control. A blue vs. red picture background indicated whether Welsh or English was the target language (counterbalanced across participants).

As predictability of the upcoming task (Poljac et al., 2009) and language (Declerck et al., 2015) makes switching easier, the proportion of cognates and language switches was kept low, and their distribution was varied: Each list of 195 items contained 27 cognates (18 of which were experimental items), occurring at unpredictable intervals with 2–17 words between two cognates, and 22 language switches, also at unpredictable intervals, with 5–17 words between two switches. To avoid priming specific lexical candidates (Kroll et al., 2006), picture names were not trained beforehand, and no pictures were repeated during the experiment.

Procedure

Participants were tested one at a time in a sound proof booth, seated in front of a computer and a microphone. They received written instructions in both English and Welsh to name pictures as fast as possible, and to press a response button on the computer after they had finished speaking. They were asked not to use articles in their response. They were instructed about the color cues indicating the language, and as a reminder there were labels below the screen with the words “English” and “Welsh,” in both languages, printed in the appropriate colors. The experiment started with a practice part.

Pictures were presented one at a time on the computer screen. The pictures were black line-drawings on a blue or red background. The picture stayed on the screen until the participant pressed the response button. At 600 ms after button press, the next picture appeared on the screen. Audio recordings of the entire experiment were made, and the onset of each picture presentation on the screen was marked in the recording. The experiment was controlled with Nijmegen Experiment Set-Up software.

Data Processing

The onset of each verbal response was labeled manually to obtain greater accuracy than with automatic extraction, with

the speech editor Praat. Naming latencies were calculated as the duration between the onset of picture presentation and the onset of the verbal response. For each response, the response language was coded (as cognate, Welsh or English; note that for cognates it was not—and by definition cannot be—determined whether the response was English or Welsh), and whether the response consisted of a single word, without article, without errors (i.e., completely and correctly pronounced) or repairs, and whether it matched the intended picture name.

In 92.5% of all trials, participants responded in the correct language. In 93.3% of the trials, participants gave a single-word response without errors or repairs. None of the responses to experimental items or items directly preceding them formed a Welsh-English false friend (i.e., a word with the same form but different meaning). Given the very low proportion of errors, only naming latencies were analyzed. Data analysis was conducted on experimental Cognates and Controls (to test for a cognate effect), and on the subset of 54 filler words (all of which were non-cognates) that occurred immediately after the experimental Cognates and Controls, both in switch and non-switch condition (to test for post-Cognate slowing, and for the occurrence of switch costs commonly reported with the mixed naming paradigm).

Three pairs of experimental items were removed from analysis (see Appendix A), because the non-cognate filler occurring *before* the Cognate/Control sometimes received a cognate response. For the remaining stimuli, there was still no difference between Cognates and Controls on any of the 32 stimulus characteristics.

Responses were only included in the analysis if they (1) consisted of a single word without errors or repairs, (2) were given in the intended language, (3) matched the intended picture name, and (4) had a naming latency ≤ 3000 ms. This resulted in the removal of 415 experimental trials (26%).

Data Analysis

The data was analyzed using linear mixed effects models with crossed random effects for participants and items, using the lmer package (Bates, 2005) in R version 2.15.2 (R Development Core Team, 2009).

Two different series of analyses were performed. The first series analyzed the naming latencies of the experimental Cognate and Control words to assess the occurrence of cognate inhibition. The second series analyzed the naming latencies of the 54 non-cognate filler items following the experimental Cognate and Control words, first to assess cognate inhibition, and second to assess whether the present data adhere to the commonly reported pattern of switch costs, and whether such switch costs were symmetrical between the two languages as to be expected for our highly proficient, early bilinguals (Costa and Santesteban, 2004; Costa et al., 2006).

In the analysis of the experimental Cognates and Controls, the variables of primary interest were the fixed effects of Cognate Condition (Cognate and Control), Target Language (English and Welsh), and each participants' self-reported Language Dominance (English-dominant, Welsh-dominant, and

equal dominance). In order to avoid collinearity in the data and to maximize the likelihood of model convergence, the factors Cognate Condition and Target Language were mean-centered prior to analysis (Baayen, 2008). Cognates and English trials were coded as -1 , and Controls and Welsh trials as $+1$. Thus, negative coefficients correspond to slower naming times for Cognates and English. Self-reported Language Dominance was coded categorically, with "Both" serving as the control group. Three- and two-way interactions among these variables were included in the analysis. In addition to these variables of primary interest, the analysis also included the natural log frequency, number of syllables, and average self-reported age of acquisition of each word, all in the language relevant in that trial. To control for spillover effects from naming earlier words, naming latencies of the preceding filler trial were also included.

The analysis of the filler items included the same fixed effects of Cognate Condition (now pertaining to the *preceding* trial), Target Language, and Language Dominance, as well as a fixed effect of Language Switching (Switch and Non-Switch), and interaction terms. Switch trials were coded as -1 and Non-Switch trials as $+1$. The analysis also contained naming latencies of the preceding trial. The analysis now did *not* include the natural log frequency, number of syllables, and average self-reported age of acquisition of the items, as all comparisons in the analysis of the filler items were within-items.

In order to determine which variables to include in the model, a forward selection procedure was used in which each of the variables was entered into the analysis individually, followed by interaction terms, and improvements in model fit were assessed through likelihood ratio tests (Baayen et al., 2008). Analyses included main effects of each of the fixed effects, as well as random intercepts for participants and items. Effects that did not improve model fit were excluded from analyses. The models reported correspond to the best fit models based on this procedure. In addition to the factors of interest in the study (Cognate Condition, Target Language, Language Dominance and, for the fillers only, Language Switching), the only other variables that added significantly to the models were the number of syllables for the Cognates and Controls, and the naming latencies of the preceding trial. Thus, lexical frequency and average self-reported age of acquisition of the words did not affect the outcomes significantly.

As the inclusion of random slopes did not improve model fit for any of the variables, random slopes were not included in the analysis; thus p -values for each predictor were estimated using resampling techniques available with the pvals.fnc function of the languageR package (Baayen et al., 2008). Further, due to some positive skewing in the naming latencies, analyses were performed on log naming latencies; note however, that performing the analysis on raw naming latencies led to similar results (not reported). Finally, all the analyses were also performed with self-reported L1 instead of self-reported Language Dominance, yielding similar results (not reported).

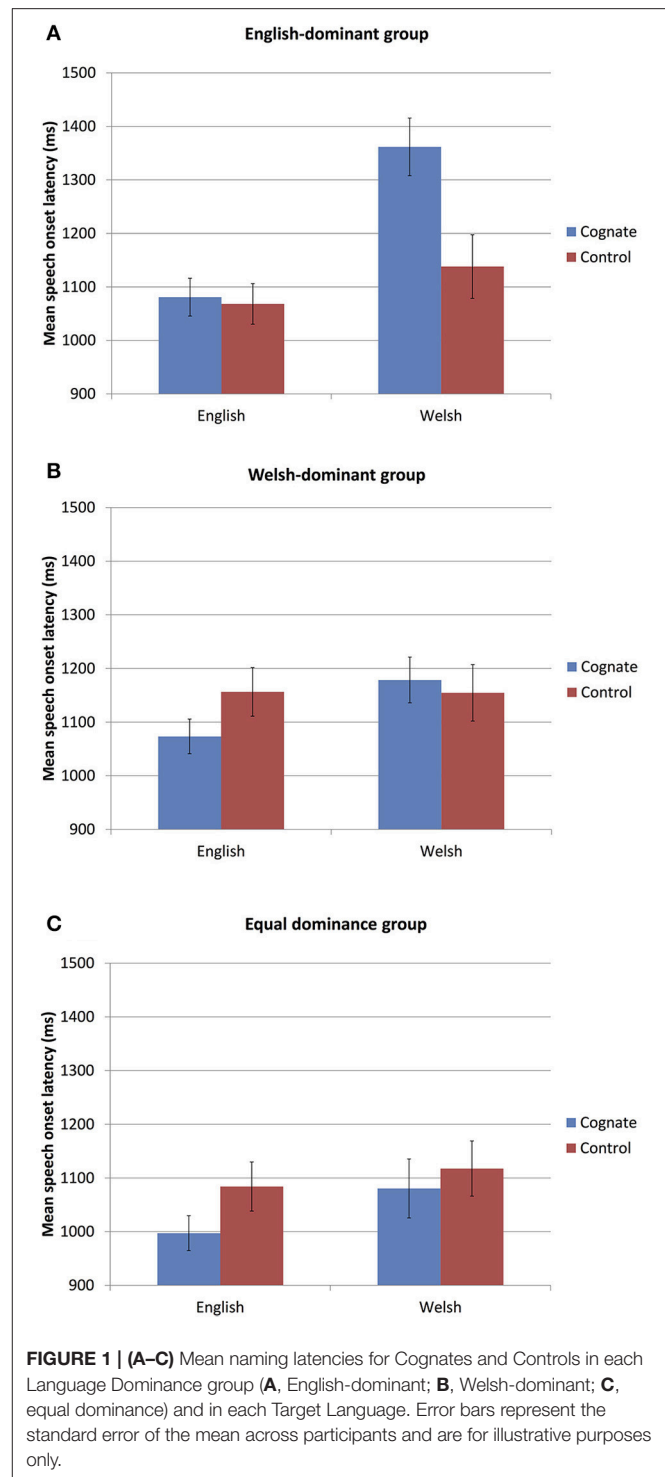
RESULTS

Cognates vs. Controls

First, we compare the Cognates and non-cognate Controls. We hypothesized (1) that in addition to the commonly reported cognate facilitation, we might find evidence for cognate inhibition, and (2) that the participants' language dominance might affect the direction of the cognate effect, with some groups showing cognate facilitation (i.e., shorter naming latencies for Cognates compared to non-cognate Controls), and others cognate inhibition (i.e., longer naming latencies for Cognates compared to non-cognate Controls). Indeed, **Figure 1** shows that language dominance affected the direction of the cognate effect. Whereas the Welsh-dominant and the equal dominance groups show (a tendency toward) *cognate facilitation* in most conditions, the English-dominant group shows no difference between cognates and controls when naming in English and, importantly, *cognate inhibition* when naming in Welsh. Results of the best-fitting mixed effects model are presented in **Table 1**. Importantly, as **Table 1** shows, there were two significant interactions, between Cognate Condition and Language Dominance, and between Cognate Condition and Target Language.

Following up on the two two-way interactions, separate mixed-effects models were estimated for the effect of Cognate Condition for each Language Dominance group and each Target Language separately (**Table 2**). The analyses showed that for the English-dominant group, when the target language was English, naming latencies for Cognates and Controls were not significantly different (mean difference = 0.01 s, $SD = 0.39$), whereas when the target language was Welsh, naming latencies were significantly longer for Cognates than for Controls (mean difference = 0.22 s, $SD = 0.50$). For the Welsh-dominant group, naming latencies were not significantly different for Cognates and Controls neither when the target language was English (mean difference = -0.08 s, $SD = 0.40$) nor when it was Welsh (mean difference = 0.02 s, $SD = 0.48$). For the equal dominance group, naming latencies were significantly shorter for Cognates than for Controls when naming in English (mean difference = -0.09 s, $SD = 0.37$) but not in Welsh (mean difference = -0.03 s, $SD = 0.46$).

Further, separate mixed-effects models were estimated for the effects of Cognate Condition and Target Language for each Language Dominance group (**Table 3**). They show that for the English-dominant group, naming latencies were shorter in English than in Welsh (mean difference = -0.20 s, $SD = 0.45$), which is in line with a greater proficiency in English than in Welsh (e.g., Meuter, 2005). For the other two groups, naming latencies in the two languages were not significantly different (Welsh-dominant: mean difference = -0.05 s, $SD = 0.44$; equal dominance: mean difference = -0.07 s, $SD = 0.40$). Those analyses also show that, collapsed over Target Language, naming latencies were significantly longer for Cognates than for Controls for the English-dominant group (mean difference = 0.11 s, $SD = 0.44$), and significantly shorter for Cognates than for Controls for the equal dominance group (mean difference = -0.07 s, $SD = 0.41$); there was no statistically significant difference between Cognates and Controls for the



Welsh-dominant group (mean difference = -0.03 s, $SD = 0.44$).

With respect to the control variables, the omnibus analysis (**Table 1**) revealed main effects of Number of Syllables, and of naming latency of the Preceding Trial, showing that, as expected, people were slower to initiate speech when words had more

TABLE 1 | Results of the best-fitting linear mixed effects model predicting log response times for Cognates vs. Controls.

	Sum of squares	Mean square	Num DF	Den DF	F	Pr > F
Cognate condition	0.03	0.03	1	1120.88	0.29	0.587
Target language	1.17	1.17	1	1144.22	11.76	0.001
Language dominance	0.21	0.11	2	43.14	1.08	0.350
Number of syllables	0.94	0.94	1	1118.41	9.46	0.002
Response time preceding trial	0.67	0.67	1	1158.73	6.72	0.010
Cognate condition * target language	0.43	0.43	1	1156.19	4.32	0.038
Cognate condition * language dominance	0.91	0.46	2	1120.62	4.60	0.010
Target language * language dominance	0.31	0.15	2	1139.40	1.54	0.215
Cognate condition * target language * language dominance	0.02	0.01	2	1156.67	0.10	0.903

TABLE 2 | Results of the best-fitting linear mixed effects model predicting log response times for Cognates vs. Controls, for each language dominance group and each Target Language separately.

	B	CI	Std. Error	p
ENGLISH-DOMINANT, TARGET LANGUAGE ENGLISH				
Intercept	-0.14	-0.30 – 0.03	0.08	0.099
Cognate condition	-0.01	-0.05 – 0.03	0.02	0.541
Number of syllables	0.02	-0.06 – 0.11	0.04	0.564
Response time preceding trial	0.10	0.01 – 0.19	0.05	0.035
ENGLISH-DOMINANT, TARGET LANGUAGE WELSH				
Intercept	0.20	-0.04 – 0.44	0.12	0.112
Cognate condition	-0.08	-0.14 – -0.02	0.03	0.006
Number of syllables	-0.03	-0.15 – 0.10	0.06	0.683
Response time preceding trial	0	-0.10 – 0.10	0.05	0.969
WELSH-DOMINANT, TARGET LANGUAGE ENGLISH				
Intercept	-0.18	-0.35 – -0.02	0.08	0.035
Cognate condition	0.02	-0.02 – 0.06	0.02	0.331
Number of syllables	0.14	0.06 – 0.23	0.04	0.002
Response time preceding trial	0.01	-0.07 – 0.10	0.04	0.740
WELSH-DOMINANT, TARGET LANGUAGE WELSH				
Intercept	0.03	-0.19 – 0.24	0.11	0.816
Cognate condition	-0.02	-0.07 – 0.02	0.02	0.350
Number of syllables	-0.05	-0.16 – 0.06	0.06	0.349
Response time preceding trial	0.09	0.00 – 0.18	0.05	0.046
EQUAL DOMINANCE, TARGET LANGUAGE ENGLISH				
Intercept	-0.28	-0.45 – -0.11	0.09	0.002
Cognate condition	0.05	0.00 – 0.09	0.02	0.033
Number of syllables	0.09	-0.00 – 0.18	0.05	0.060
Response time preceding trial	0.12	0.03 – 0.22	0.05	0.012
EQUAL DOMINANCE, TARGET LANGUAGE WELSH				
Intercept	-0.21	-0.42 – 0.01	0.11	0.060
Cognate condition	0.01	-0.04 – 0.07	0.03	0.613
Number of syllables	0.17	0.06 – 0.28	0.06	0.003
Response time preceding trial	0	-0.10 – 0.10	0.05	0.985

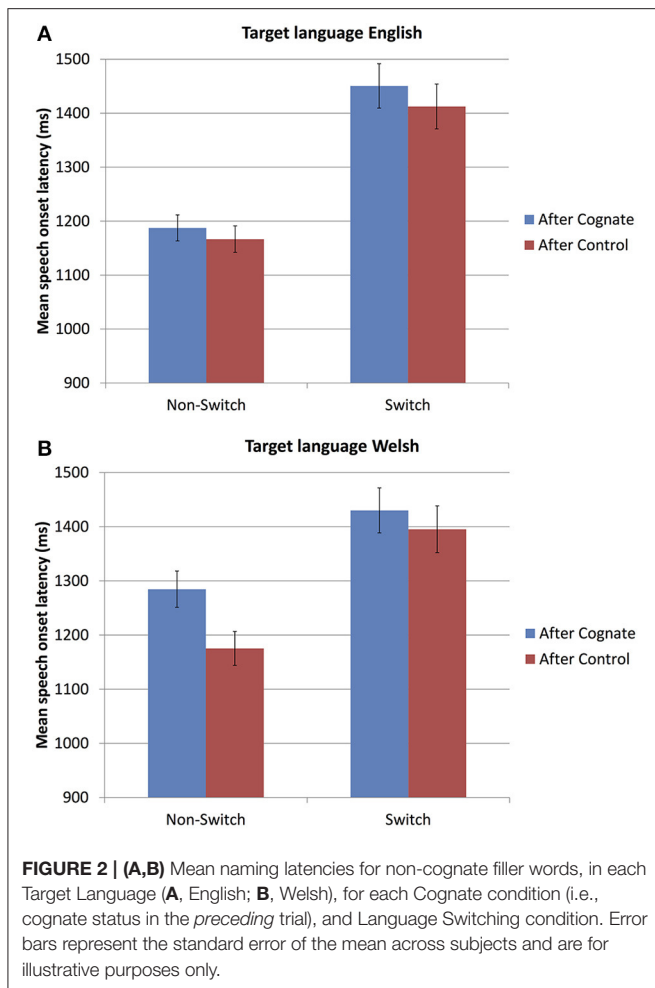
syllables, and when they were slower on the preceding trial. Two additional control variables were explored but not retained in the final models because they did not affect the outcomes significantly, namely: lexical frequency and average self-reported age of acquisition of the words.

TABLE 3 | Results of the best-fitting linear mixed effects model predicting log response times for Cognates vs. Controls, for each language dominance group separately.

	B	CI	Std. Error	P
ENGLISH-DOMINANT				
Intercept	0.03	-0.11 – 0.17	0.07	0.682
Cognate condition	-0.04	-0.07 – -0.01	0.02	0.022
Target language	0.06	0.03 – 0.10	0.02	<0.001
Number of syllables	0	-0.07 – 0.06	0.03	0.928
Response time preceding trial	0.05	-0.02 – 0.12	0.03	0.158
WELSH-DOMINANT				
Intercept	-0.09	-0.23 – 0.05	0.07	0.198
Cognate condition	0	-0.03 – 0.03	0.01	0.971
Target language	0.02	-0.01 – 0.05	0.02	0.138
Number of syllables	0.06	-0.00 – 0.13	0.03	0.065
Response time preceding trial	0.05	-0.01 – 0.11	0.03	0.091
EQUAL DOMINANCE				
Intercept	-0.23	-0.37 – -0.10	0.07	0.001
Cognate condition	0.03	0.00 – 0.07	0.02	0.046
Target language	0.02	-0.01 – 0.06	0.02	0.211
Number of syllables	0.12	0.05 – 0.19	0.03	0.002
Response time preceding trial	0.06	-0.01 – 0.13	0.04	0.101

Filler Items after Cognates vs. Controls

Non-cognate filler items were first analyzed to ascertain whether the data showed switch costs, as typically reported for mixed naming experiments (Meuter and Allport, 1999; Jackson et al., 2001; Costa and Santesteban, 2004; Christoffels et al., 2007; Verhoeve et al., 2009, 2010), and whether those switch costs were symmetrical as expected (Costa and Santesteban, 2004; Costa et al., 2006). Indeed, **Figure 2** shows the expected switch costs, with longer naming latencies in switch than in non-switch trials. Further, **Figure 2** shows that these switch costs are not asymmetrical; rather, the size of the switch costs is similar in English and Welsh, which is consistent with previous findings for highly proficient bilinguals (Costa and Santesteban, 2004; Costa et al., 2006). Results of the best fitting mixed effects model are presented in **Table 4**. Indeed, as **Figure 2** suggests, pictures were named significantly more slowly in Switch than in Non-Switch trials (mean difference = 0.22 s, *SD* = 0.48), and there was no interaction between Language Switch



and Target Language, confirming that switch costs were not asymmetrical.

Crucially, in line with our hypothesis, **Figure 2** also shows that naming latencies for the non-cognate filler items were longer when the *preceding* trial was a Cognate than when it was a non-cognate Control. Note that this is not an artifact of the naming latency of the preceding trial: As expected, pictures were named more slowly when the previous trial was named more slowly (**Table 4**); in addition, however, even though the analysis factored out the naming latency of the preceding trial, pictures were named significantly more slowly after a Cognate than after a Control (mean difference = 0.05 s, $SD = 0.49$).

Importantly, cognate status in the preceding trial affected naming latencies irrespective of Language Dominance: There were no significant two- or three-way interactions with Language Dominance, and no main effect of Language Dominance (**Table 4**). Indeed, the pattern of slower naming after a Cognate than after a Control was present across all three Language Dominance groups (mean difference, English-dominant: 0.01 s; Welsh-dominant: 0.05 s; equal dominance: 0.09 s). It was thus carried by all groups—not only by the English-dominant group (that exhibited cognate inhibition on the Cognate vs. Control trials themselves), but also by the other two groups (that showed

either cognate facilitation, or no difference between Cognates and Controls on those trials). This points to the recruitment of cognitive control during cognate naming, even if the Cognate and Control trials themselves do not reveal it, as we hypothesized.

DISCUSSION

As hypothesized, this study has provided evidence for inhibition as well as the more commonly reported facilitation during the production of cognates compared to non-cognate control words in a mixed picture naming task by highly proficient Welsh-English bilinguals. First, facilitation and inhibition were found on the cognate and control trials themselves. As hypothesized, the participants' language dominance affected the direction of this cognate effect, with the English-dominant group showing cognate *inhibition* when naming in Welsh (and no difference between cognates and controls when naming in English), and the Welsh-dominant and equal dominance groups generally showing a pattern of cognate *facilitation*. Second, cognate inhibition was found as a *behavioral adaptation* effect, with non-cognate filler words being named more slowly after cognates than after non-cognate controls.

Interestingly, this behavioral adaptation effect was found consistently across all language dominance groups and both target languages. Thus, in contrast to the experimental items themselves, where naming latencies were longer for cognates than for controls only for the Welsh-English bilinguals and only when naming in Welsh, cognate inhibition as shown in the next trial was a more general phenomenon. This suggests that cognate production might require the recruitment of cognitive control, even if this is not measurable in the cognate trials themselves. This finding is reminiscent of effects of response conflict in the non-linguistic domain, where performance is modulated by the presence of response conflict in the preceding trial (Gratton et al., 1992), e.g., in the Eriksen Flanker task (Eriksen and Eriksen, 1974), which has been interpreted as a result of conflict monitoring (Botvinick et al., 2001; Yeung et al., 2004). Note that the finding that filler words were named more slowly after cognates than after control words cannot be an artifact of slower responses to cognates than controls. First, if it was an artifact it should be limited to the English-dominant group and to the fillers following on a Welsh cognate; the effect is, however, independent of language dominance group and of target language. Second, the experimental and statistical methodology employed here makes that interpretation unlikely². We thus conclude that the slower naming of fillers after cognates than after control words is independent of the differences in naming latencies between

²First, the self-paced trial timing of the experiment aimed to limit the spillover effect of naming latencies of one trial into the next. With fixed rather than self-paced trial timing, there would have been a confound (Monsell, 2003). Instead, participants pressed a button after finishing naming of each picture, and the next picture was presented 600 ms after their button press. Thus, the delay between finishing naming of the cognate or control and the onset of the next trial was the same in both conditions. Second, naming latencies of cognates and controls were included in the analysis of the following fillers, which enabled us to separate the two effects. Results showed that even when the spillover effects from cognates and controls into the following trials were taken into account, the effect of cognate condition on the naming of the following fillers remained robust.

TABLE 4 | Results of the best-fitting linear mixed effects model predicting log response times for fillers.

	Sum of squares	Mean square	Num DF	Den DF	F	Pr > F
Cognate condition	0.68	0.68	1	1043.90	7.40	0.007
Language switching condition	1.80	1.80	1	34.66	19.76	0.000
Target language	0.21	0.21	1	1043.63	2.34	0.127
Language dominance	0.39	0.19	2	47.16	2.13	0.130
Number of syllables	0.85	0.85	1	34.50	9.27	0.004
Response time preceding trial	1.90	1.90	1	1070.75	20.79	0.000
Cognate condition * language switching condition	0.02	0.02	1	1045.84	0.18	0.671
Cognate condition * language dominance	0.04	0.02	2	1037.74	0.21	0.811
Language switching condition * language dominance	0.23	0.11	2	1033.49	1.25	0.288
Cognate condition * language switching condition * language dominance	0.02	0.01	2	1037.11	0.10	0.904

cognates and controls themselves, and that it might result from increased cognitive control during the production of cognates. This interpretation is in line with Acheson et al. (2012). In their bilingual picture naming study, they found that bilingual speakers named cognates faster than non-cognates. Yet, they also found an increased ERN-like effect, indicating increased response conflict for cognates compared to non-cognates. In addition, they found a behavioral adaptation effect, with the magnitude of the cognate facilitation effect being smaller after the naming of a cognate than after the naming of a matched non-cognate. Acheson et al. (2012) conclude that even though the cognates were named faster than non-cognates, they must have induced more response conflict than the non-cognates because of the two highly similar pronunciations that the speakers had to mediate between.

On the cognate trials themselves, the Welsh-dominant and equal dominance groups showed the typically reported pattern of cognate facilitation. The English-dominant group, when naming in Welsh, showed cognate inhibition rather than facilitation. This finding is uncommon, as there is a large literature reporting advantages for the processing of cognates over non-cognates, in various bilingual populations, and using a wide range of experimental paradigms involving both speech production and perception (De Groot and Nas, 1991; Kroll and Stewart, 1994; Gollan et al., 1997; Van Hell and De Groot, 1998; Dijkstra et al., 1999; Van Hell and Dijkstra, 2002; Lemhöfer and Dijkstra, 2004; Christoffels et al., 2006, 2007). In the realm of picture naming, previous (mostly monolingual) experiments have also reported either faster naming for cognates than for non-cognates (Costa et al., 2000; Christoffels et al., 2006, 2007; Hoshino and Kroll, 2008; Verhoef et al., 2009) or the absence of any difference, in the case of highly proficient bilinguals (e.g., Costa et al., 2000; Ivanova and Costa, 2008; Strijkers et al., 2010)—but no cognate inhibition. The only exception that we are aware of is a mixed word naming experiment (Filippi et al., 2014), which showed that late bilinguals were slower to produce cognates than non-cognate control words.

The explanation that is generally offered for cognate facilitation is that the similarity at the semantic as well as the form level leads to enhanced activation of the word form representations, as the lemmas of both cognates contribute to the

activation of the shared word forms. It has also been proposed, on the other hand, that even small amounts of phonological overlap can lead to increased lexical competition between the lemmas of the words that share aspect of their form, as a result of feedback from the word form to the lemma level (Declerck and Philipp, 2015). We have proposed that during the production of cognates, such feedback to both the intended and the unintended lemma might cause either facilitation or inhibition. We have further proposed that cognate processing might thus be affected by two different processes, namely competition at the lexical-semantic level and facilitation at the word form level, and that facilitation at the word form level might (sometimes) outweigh any effects of inhibition at the lemma level. The results of the present study, showing both facilitation and inhibition, could stem from the interplay between those two processes. The finding of facilitation on the cognate trials and inhibition as a behavioral adaptation on the next trial *within the same participants* is also in line with such an account. If there are indeed two different processes at work during the lexical selection of cognates—competition at the lexical-semantic level, and facilitation at the word form level—the latter of which might often obscure the former, this could contribute to the explanation of why some studies have found cognate facilitation and others cognate inhibition (namely Filippi et al., 2014, and the present study): in many studies the benefit of the activation of shared word forms might have outweighed—and thus obscured—the possible slowing effect of lexical competition at the lemma level.

Why then is it that we find cognate inhibition (on the cognate trials themselves) for the English-dominant group and facilitation for the Welsh-dominant and the equal dominance groups? And why is it that the English-dominant bilinguals showed cognate inhibition (again, on the cognate trials themselves) when naming in Welsh, but not in English? While the answer remains speculative, it calls in mind two proposals that have been put forward for the occurrence of cross-language inhibitory control. First, recall that it has been proposed that the occurrence of inhibitory control depends on the speakers' language dominance, such that bilingual speakers might depend on cross-language inhibition to suppress words in their dominant language when speaking in their less dominant language, but not vice versa (Costa and Santesteban, 2004; Branzi et al.,

2014). This is in line with the English-dominant bilinguals in the present study showing cognate inhibition when naming in Welsh, their non-dominant language, but not in English, their dominant language. Second, it has been proposed that the use of inhibitory control depends on the specifics of the task, including preparation time and the type of distracters used (Verhoef et al., 2009; Roelofs et al., 2011). Such details might extend to the nature of the cognates used in the experiment. An explanation for the occurrence of cognate inhibition in Welsh but not in English may be related to the origin of the cognates. The cognates used in this experiment, and possibly all Welsh-English cognates except for proper nouns, as explained in the Introduction, were borrowings from English into Welsh. This might explain why cognate costs were found (on the cognate trials themselves) in Welsh for the English-dominant participants, but not in English for the Welsh-dominant (and/or equal dominance) participants. Naming a cognate in Welsh, even though firmly established as a Welsh word, might require overcoming the prepotent response of naming it in the language of origin, which might entail more lexical competition than naming it in English.

Another possible explanation for the difference between our findings and those from previous studies is in the experimental paradigm and methodological details. In the present study we investigated the production of cognates in a *mixed* picture naming task, which requires participants to switch between their languages during picture naming. The combination of cognates and the mixed naming paradigm is rather uncommon; most previous studies involving cognate naming have used a monolingual or blocked-language design (Costa et al., 2000; Gollan and Acenas, 2004; Christoffels et al., 2006; Hoshino and Kroll, 2008), which does not require words from the other language to be active (but see, Christoffels et al., 2007; Verhoef et al., 2009). The mixed naming task (e.g., Meuter and Allport, 1999; Costa et al., 2006; Gollan and Ferreira, 2009; Poarch and Van Hell, 2012), in contrast, requires that lexical representations from both languages are activated and considered for selection. Under such circumstances, cross-language lexical competition can be expected to be stronger than if only one language is needed for the task at hand (e.g., Green, 1998). The contrast between the present results and those found with other experimental paradigms is in line with the claim of Kroll et al. (2006) that the occurrence of parallel activation and cross-language competition is contingent on task demands.

We are aware of two previous studies that also included cognates in a mixed picture naming experiment. Those studies showed cognate facilitation rather than inhibition (Christoffels et al., 2007; Verhoef et al., 2009). There are two major differences in the methodology of those studies compared to the present study which may have contributed to the difference in outcomes. First, the experiment in Verhoef et al. (2009) was specifically designed to enable participants to inhibit responses in the non-target language, by presenting the language cues prior to the pictures. Thus, cognate costs should not be expected. Second, in both studies (Christoffels et al., 2007; Verhoef et al., 2009), the same picture names were repeated extensively during the experiment, thus priming the lexical candidates, which has been suggested to affect the occurrence of lexical competition (Kroll

et al., 2006). In the present study, language cues were provided simultaneously with the pictures such that lemmas from both languages would be active during lexical selection, and picture names were never repeated during the experiment, which may have optimized the possibility of finding a cognate inhibition effect.

The present results are in line with those found with a mixed word naming (i.e., reading aloud) experiment, which also found a cognate inhibition effect (Filippi et al., 2014), despite the differences between the picture naming and word naming tasks, and the cognitive processes involved: The word naming task and the picture naming task are known to require different cognitive processes (e.g., Mousikou and Rastle, 2015). E.g., the role of semantic information is assumed to be smaller in reading than in picture naming; reading aloud does not *necessarily* involve the retrieval of semantic information, which is indispensable for the picture naming task, and could under some circumstances be performed purely by converting graphemes to phonemes (Riès et al., 2012; Valente et al., 2016). The present study thus shows that the cognate inhibition effect as found by Filippi et al. (2014) is not limited to the word naming task.

While this study presents a novel finding with respect to the occurrence of cognate costs, the other patterns in the data are fully in line with those in previous studies. First, this study replicates the typical switch costs, with responses being slower on switch trials than on non-switch trials, that have been found in language switching (Meuter and Allport, 1999; Jackson et al., 2001; Costa and Santesteban, 2004; Christoffels et al., 2007; Verhoef et al., 2009, 2010) as well as in non-linguistic task switching experiments (Allport et al., 1994; Rogers and Monsell, 1995; Rubinstein et al., 2001). In the present study, switch trials were also slower than non-switch trials. Second, this study replicates the finding that switch costs are symmetrical for highly proficient bilinguals: whereas less proficient bilinguals are commonly found to show asymmetric switch costs, with switching into the stronger language entailing larger costs than switching into the weaker language (Meuter and Allport, 1999; Costa and Santesteban, 2004; Campbell, 2005; Philipp et al., 2007; Wang et al., 2007; Verhoef et al., 2009), there were no asymmetric switch costs in the present study, in line with previous findings for highly proficient bilinguals (Costa and Santesteban, 2004; Costa et al., 2006). We aimed to test highly proficient bilinguals, and the results suggest that the participants fit that description indeed.

In summary, this study shows evidence that cognate naming can cause costs rather than benefits, showing both as inhibition *during* cognate production and as a behavioral adaptation effect *after* cognate production. It provides evidence for cross-language lexical competition, supporting models of lexical selection that allow for inhibitory control (e.g., De Bot, 1992; Green, 1998; Abutalebi and Green, 2007). It has been proposed that cross-language inhibition might only occur for speakers with low proficiency in one of the languages (Costa and Santesteban, 2004; Branzi et al., 2014). Others have argued that words from both languages can compete in highly proficient speakers as well (e.g., Kroll et al., 2008). The present results support the latter view, by providing evidence for cross-language inhibition during cognate production in highly proficient, early Welsh-English

bilinguals. The finding of cognate inhibition, particularly for these highly proficient bilinguals, thus provides strong evidence for the occurrence of lexical competition across languages in the bilingual mental lexicon.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Radboud University's Ethics Assessment Committee (EAC) Humanities³ and the Radboud University's code of academic integrity and conduct⁴ that adhere to European regulations.

AUTHOR CONTRIBUTIONS

All authors have been substantially involved in (aspects of) the development of the research plan, experimental design, data

³<http://www.ru.nl/eac-humanities/>

⁴<http://www.ru.nl/english/about-us/working-radboud/integrity-conduct/>

REFERENCES

- Abutalebi, J., and Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguist.* 20, 242–275. doi: 10.1016/j.neuroling.2006.10.003
- Acheson, D. J., Ganushchak, L. Y., Christoffels, I. K., and Hagoort, P. (2012). Conflict monitoring in speech production: physiological evidence from bilingual picture naming. *Brain Lang.* 123, 131–136. doi: 10.1016/j.bandl.2012.08.008
- Allport, A., Styles, E. A., and Hsieh, S. (1994). "Shifting intentional set: Exploring the dynamic control of tasks," in *Attention and Performance XV*, eds C. Umiltà and M. Moscovitch (Cambridge, MA: MIT), 421–452.
- Altmann, E. M., and Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychol. Rev.* 115, 602–639. doi: 10.1037/0033-295X.115.3.602
- Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: CUP.
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Bates, D. M. (2005). Fitting linear mixed models in R. *R News* 5, 27–30.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652. doi: 10.1037/0033-295X.108.3.624
- Branzi, F. M., Martin, C. D., Abutalebi, J., and Costa, A. (2014). The after-effects of bilingual language production. *Neuropsychologia* 52, 102–116. doi: 10.1016/j.neuropsychologia.2013.09.022
- Campbell, J. I. D. (2005). Asymmetrical language switching costs in Chinese-English bilinguals' number naming and simple arithmetic. *Bilingualism: Lang. Cognit.* 8, 85–91. doi: 10.1017/S136672890400207X
- Christoffels, I. K., De Groot, A. M. B., and Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: the role of expertise and language proficiency. *J. Mem. Lang.* 54, 324–345. doi: 10.1016/j.jml.2005.12.004
- Christoffels, I. K., Firk, C., and Schiller, N. O. (2007). Bilingual language control: an event-related brain potential study. *Brain Res.* 1147, 192–208. doi: 10.1016/j.brainres.2007.01.137
- Costa, A., and Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Bilingualism: Lang. Cognit.* 2, 231–244. doi: 10.1017/S1366728999000334
- Costa, A., and Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *J. Mem. Lang.* 50, 491–511. doi: 10.1016/j.jml.2004.02.002
- Costa, A., and Santesteban, M. (2006). The control of speech production by bilingual speakers: Introductory remarks. *Bilingualism: Lang. Cognit.* 9, 115–117. doi: 10.1017/S1366728906002471
- Costa, A., Caramazza, A., and Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *J. Exp. Psychol.* 26, 1283–1296. doi: 10.1037/0278-7393.26.5.1283
- Costa, A., Miozzo, M., and Caramazza, A. (1999). Lexical selection in bilinguals: do words in the bilingual's two lexicons compete for selection? *J. Mem. Lang.* 41, 365–397. doi: 10.1006/jmla.1999.2651
- Costa, A., Santesteban, M., and Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 1057–1074. doi: 10.1037/0278-7393.32.5.1057
- De Bot, K. (1992). A bilingual production model: levelt's 'Speaking' model adapted. *Appl. Linguist.* 13, 1–24.
- De Groot, A. M. B., and Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *J. Mem. Lang.* 30, 90–123. doi: 10.1016/0749-596X(91)90012-9
- Declerck, M., and Philipp, A. M. (2015). The unusual suspect: Influence of phonological overlap on language control. *Bilingualism: Lang. Cognit.* 18, 726–736. doi: 10.1017/S1366728914000236
- Declerck, M., Koch, I., and Philipp, A. M. (2015). The minimum requirements of language control: evidence from sequential predictability effects in language switching. *J. Exp. Psychol.* 41, 377–394. doi: 10.1037/xlm0000021
- Deuchar, M., Davies, P., Herring, J., Parafita Couto, M. C., and Carter, D. (2014). "Bilingual language use," in *Advances in the Study of Bilingualism*, eds E. Thomas and I. Mennen (Bristol: Multilingual Matters), 93–110.
- Dijkstra, T., Grainger, J., and Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homophones: the neglected role of phonology. *J. Mem. Lang.* 41, 496–518. doi: 10.1006/jmla.1999.2654
- Ellis, N. C., O'Dochartaigh, C., Hicks, W., Morgan, M., and Laporte, N. (2001). Cronfa Electroneg o Gymraeg (CEG): a 1 million word lexical database and frequency count for Welsh. Available online at: www.bangor.ac.uk/canolfanbedwyr/ceg.php.en
- Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16, 143–149. doi: 10.3758/BF03203267
- Filippi, R., Karaminis, T., and Thomas, M. S. C. (2014). Language switching in bilingual production: empirical data and computational modelling.

collection, data analysis, interpretation, and/or writing of the paper.

ACKNOWLEDGMENTS

This research was supported by a Small Research Grant from the British Academy awarded to the first and second author (grant number 101421), and by a Vidi grant from the Netherlands Organization for Scientific Research (NWO) awarded to the first author (grant number 276-89-006). We thank the Max Planck Institute for Psycholinguistics and the Centre for Research on Bilingualism, Bangor University, for their generous support, and for facilitating several research visits of the first and second author to Bangor and Nijmegen, respectively. We thank Peredur Davies, Myfyr Prys, and Alberto Rosignoli for advice on stimulus selection, and Willemijn van den Berg, Lies Cuijpers, Thomas Kuijpers, and Denise Moerel for measuring naming latencies. Special thanks to Jonathan Stammers for invaluable help during several stages of this study.

- Bilingualism: Lang. Cognit.* 17, 294–315. doi: 10.1017/S1366728913000485
- Finkbeiner, M., Gollan, T. H., and Caramazza, A. (2006). Lexical access in bilingual speakers: what's the (hard) problem? *Bilingualism: Lang. Cognit.* 9, 153–166. doi: 10.1017/S1366728906002501
- Gollan, T. H., and Acenas, L.-A. R. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English and Tagalog-English bilinguals. *J. Exp. Psychol.* 30, 246–269. doi: 10.1037/0278-7393.30.1.246
- Gollan, T. H., and Ferreira, V. S. (2009). Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *J. Exp. Psychol.* 35, 640–665. doi: 10.1037/a0014981
- Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., and Jernigan, T. L. (2007). The bilingual effect on Boston Naming Test performance. *J. Int. Neuropsychol. Soc.* 13, 197–208. doi: 10.1017/s1355617707070038
- Gollan, T. H., Forster, K. I., and Frost, R. (1997). Translation priming with different scripts: masked priming with cognates and noncognates in Hebrew-English bilinguals. *J. Exp. Psychol.* 23, 1122–1139. doi: 10.1037/0278-7393.23.5.1122
- Gratton, G., Coles, M. G. H., and Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *J. Exp. Psychol.* 121, 480–506. doi: 10.1037/0096-3445.121.4.480
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Lang. Cognit.* 1, 67–81. doi: 10.1017/S1366728998000133
- Hermans, D. (2004). Between-language identity effects in picture-word interference tasks: a challenge for language-nonspecific or language-specific models of lexical access? *Int. J. Bilingualism* 8, 115–125. doi: 10.1177/13670069040080020101
- Hermans, D., Bongaerts, T., De Bot, K., and Schreuder, R. (1998). Producing words in a foreign language: can speakers prevent interference from their first language? *Bilingualism: Lang. Cognit.* 1, 213–229. doi: 10.1017/S1366728998000364
- Hoshino, N., and Kroll, J. F. (2008). Cognate effects in picture naming: does cross-language activation survive a change of script? *Cognition* 106, 501–511. doi: 10.1016/j.cognition.2007.02.001
- Ivanova, I., and Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychol.* 127, 277–288. doi: 10.1016/j.actpsy.2007.06.003
- Jackson, G. M., Swainson, R., Cunningham, R., and Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism: Lang. Cognit.* 4, 169–178. doi: 10.1017/s1366728901000268
- Kerns, J. G., Cohen, J. D., MacDonald, A. W. I., Cho, R. Y., Stenger, V. A., and Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science* 302, 1023–1026. doi: 10.1126/science.1089910
- Klein, E. (1986). *Klein's Comprehensive Etymological Dictionary of the English Language, 6th Edn.* Amsterdam: Elsevier.
- Kroll, J. F., and Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *J. Mem. Lang.* 33, 149–174. doi: 10.1006/jmla.1994.1008
- Kroll, J. F., Bobb, S. C., and Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Lang. Cognit.* 9, 119–135. doi: 10.1017/S1366728906002483
- Kroll, J. F., Bobb, S. C., Misra, M., and Guo, T. (2008). Language selection in bilingual speech: evidence for inhibitory processes. *Acta Psychol.* 128, 416–430. doi: 10.1016/j.actpsy.2008.02.001
- Laming, D. (1979). Choice reaction performance following an error. *Acta Psychol.* 43, 199–224. doi: 10.1016/0001-6918(79)90026-X
- Lemhöfer, K., and Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: effects of code similarity in language-specific and generalized lexical decision. *Mem. Cogn.* 32, 533–550. doi: 10.3758/BF03195845
- Los, S. A. (1996). On the origin of mixing costs: exploring information processing in pure and mixed blocks of trials. *Acta Psychol.* 94, 145–188. doi: 10.1016/0001-6918(95)00050-X
- Mayr, U., Awh, E., and Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nat. Rev. Neurosci.* 6, 450–452. doi: 10.1038/nn1051
- Meuter, R. F. I. (2005). “Language selection in bilinguals: mechanisms and processes” in *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. F. Kroll and A. M. B. De Groot (Oxford: Oxford University Press), 349–370.
- Meuter, R. F. I., and Allport, A. (1999). Bilingual language switching in naming: asymmetrical costs of language selection. *J. Mem. Lang.* 40, 25–40. doi: 10.1006/jmla.1998.2602
- Monsell, S. (2003). Task switching. *Trends Cogn. Sci.* 7, 134–140. doi: 10.1016/S1364-6613(03)00028-7
- Mousikou, P., and Rastle, K. (2015). Lexical frequency effects on articulation: a comparison of picture naming and reading aloud. *Front. Psychol.* 6:1571. doi: 10.3389/fpsyg.2015.01571
- Mueller Gathercole, V. C., and Thomas, E. M. (2009). Bilingual first-language development: dominant language takeover, threatened minority language takeover. *Bilingualism: Lang. Cognit.* 12, 213–237. doi: 10.1017/S1366728909004015
- Nieuwenhuis, S., Stins, J. F., Posthuma, D., Polderman, T. J. C., Boomsma, D. I., and De Geus, E. J. (2006). Accounting for sequential trial effects in the flanker task: conflict adaptation or associative priming? *Mem. Cogn.* 34, 1260–1272. doi: 10.3758/BF03193270
- Philipp, A. M., Gade, M., and Koch, I. (2007). Inhibitory processes in language switching: evidence from switching language-defined response sets. *Eur. J. Cogn. Psychol.* 19, 395–416. doi: 10.1080/09541440600758812
- Poarch, G. J., and Van Hell, J. G. (2012). Cross-language activation in children's speech production: evidence from second language learners, bilinguals, and trilinguals. *J. Exp. Child Psychol.* 111, 419–438. doi: 10.1016/j.jecp.2011.09.008
- Poljac, E., Koch, I., and Bekkering, H. (2009). Dissociating restart cost and mixing cost in task switching. *Psychol. Res.* 73, 407–416. doi: 10.1007/s00426-008-0151-9
- Poulisse, N. (1999). *Slips Of The Tongue: Speech Errors In First And Second Language Production*. Amsterdam: John Benjamins.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <http://www.R-project.org>
- Rabbitt, P. M. A. (1966). Errors and error-correction in choice-response tasks. *J. Exp. Psychol.* 71, 264–272. doi: 10.1037/h0022853
- Riès, S., Legou, T., Burle, B., Alario, F. X., and Malfait, N. (2012). Why does picture naming take longer than word reading? The contribution of articulatory processes. *Psychol. Bull. Rev.* 19, 955–961. doi: 10.3758/s13423-012-0287-x
- Roelofs, A., Piai, V., and Garrido Rodriguez, G. (2011). Attentional inhibition in bilingual naming performance: evidence from delta-plot analyses. *Front. Psychol.* 2, 184. doi: 10.3389/fpsyg.2011.00184
- Rogers, R. D., and Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *J. Exp. Psychol.* 124, 207–231. doi: 10.1037/0096-3445.124.2.207
- Rubinstein, J. S., Meyer, D. E., and Evans, J. E. (2001). Executive control of cognitive processes in task switching. *J. Exp. Psychol.* 27, 763–797. doi: 10.1037/0096-1523.27.4.763
- Runnqvist, E., Strijkers, K., Alario, F.-X., and Costa, A. (2012). Cumulative semantic interference is blind to language: implications for models of bilingual speech production. *J. Mem. Lang.* 66, 850–869. doi: 10.1016/j.jml.2012.02.007
- Simon, J. R. (1969). Reactions toward the source of stimulation. *J. Exp. Psychol.* 81, 174–176. doi: 10.1037/h0027448
- Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, familiarity and visual complexity. *J. Exp. Psychol.* 6, 174–215. doi: 10.1037/0278-7393.6.2.174
- Strijkers, K., Costa, A., and Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cereb. Cortex* 20, 912–928. doi: 10.1093/cercor/bhp153
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651
- Stürmer, B., Leuthold, H., Soetens, E., Schröter, H., and Sommer, W. (2002). Control over location-based response activation in the Simon task: behavioral and electrophysiological evidence. *J. Exp. Psychol.* 28, 1345–1363. doi: 10.1037/0096-1523.28.6.1345
- Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., et al. (2003). Timed picture naming: extended norms and validation against previous studies. *Behav. Res. Methods Instrum. Comp.* 35, 621–633. doi: 10.3758/BF03195542
- Székely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., et al. (2004). A new on-line resource for psycholinguistic studies. *J. Mem. Lang.* 51, 247–250. doi: 10.1016/j.jml.2004.03.002

- Thomas, E., and Gathercole, V. (2005). "Minority language survival: obsolescence or survival for Welsh in the face of English dominance?" in *Proceedings of the 4th International Symposium on Bilingualism* (Somerville, MA: Cascadia).
- Valente, A., Pinet, S., Alario, F. X., and Laganaro, M. (2016). "When" does picture naming take longer than word reading? *Front. Psychol.* 7:31. doi: 10.3389/fpsyg.2016.00031
- Van Hell, J., and De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: effects of concreteness and cognate status in word association. *Bilingualism: Lang. Cognit.* 1, 193–211. doi: 10.1017/S1366728998000352
- Van Hell, J., and Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychon. Bull. Rev.* 9, 780–789. doi: 10.3758/BF03196335
- Verhoef, K., Roelofs, A., and Chwilla, D. J. (2009). Role of inhibition in language switching: evidence from event-related brain potentials in overt picture naming. *Cognition* 110, 84–99. doi: 10.1016/j.cognition.2008.10.013
- Verhoef, K., Roelofs, A., and Chwilla, D. J. (2010). Electrophysiological evidence for endogenous control in switching between languages in overt picture naming. *J. Cogn. Neurosci.* 22, 1832–1843. doi: 10.1162/jocn.2009.21291
- Wang, Y., Xue, G., Chen, C., Xue, F., and Dong, Q. (2007). Neural bases of asymmetric language switching in second-language learners: an ER-fMRI study. *Neuroimage* 35, 862–870. doi: 10.1016/j.neuroimage.2006.09.054
- Yeung, N., Botvinick, M. M., and Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol. Rev.* 111, 931–959. doi: 10.1037/0033-295X.111.4.931

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Broersma, Carter and Acheson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

TABLE A | Experimental stimuli.

Cognate		Control	
English	Welsh	English	Welsh
MONOSYLLABIC			
Bus	Bws	Foot	Troed
Car	Car	Bench	Mainc
Cat	Cath	Nose	Trwyn
Clock	Cloc	Egg	Wy
Clown	Clown	Bear	Arth
Corn	Corn	Snake	Neidr
Desk	Desg	Saw	Llif
Drum	Drwm	Shirt	Crys
Fan	Ffan	Leg	Coes
Fork	Fforc	Goat	Gafr
Globe	Glôb	Door	Drws
Hat	Het	Wolf	Blaidd
Lamp	Lamp	Arm	Braich
Shark	Siarc	Rope	Rhaff
Tie	Tei	Hand	Llaw
Train	Trên	Cow	Buwch
Truck	Tryc	Ear	Clust
Watch	Wats	Spoon	Llwy
DISYLLABIC			
Balloon	Balŵn	Hammer	Morthwyl
Button	Botwm	Horseshoe	Pedol
Cactus	Cactws	Ladder	Ysgol
Camel	Camel	Window	Ffenest
Dolphin	Dolffin	Sandwich	Brechdan
Giraffe	Jiráff	Beetle	Chwilen
Guitar	Gitâr	Shower	Cawod
Igloo	Iglw	Rooster	Ceiliog
Monkey	Mwnci	Saddle	Cyfrwy
Parrot	Parot	Lobster	Cimwch
Pencil	Pensil	Mountain	Mynydd
Penguin	Pengwin	Whistle	Chwiban
Piano	Piano	Turtle	Crwban
Rocket	Rocket	Lizard	Madfall
Scorpion	Sgorpion	Kettle	Tegell
*Toaster	Tostiwr	Flower	Blodyn
*Tractor	Tractor	Barrel	Casgen
*Yoyo	Yoyo	Eagle	Eryr

**Item sets removed from analysis.*



Effects of the Native Language on the Learning of Fundamental Frequency in Second-Language Speech Segmentation

Annie Tremblay^{1*}, Mirjam Broersma², Caitlin E. Coughlin¹ and Jiyoun Choi³

¹ University of Kansas, Lawrence, KS, USA, ² Radboud University Nijmegen, Nijmegen, Netherlands, ³ Hanyang University, Seoul, South Korea

OPEN ACCESS

Edited by:

Marcela Pena,
Catholic University of Chile, Chile

Reviewed by:

Gorka Elordieta,
University of the Basque Country,
Spain
Hsiao-Lan Wang,
National Taiwan Normal University,
Taiwan

*Correspondence:

Annie Tremblay
atrembla@ku.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 29 January 2016

Accepted: 14 June 2016

Published: 29 June 2016

Citation:

Tremblay A, Broersma M,
Coughlin CE and Choi J (2016)
Effects of the Native Language on
the Learning of Fundamental
Frequency in Second-Language
Speech Segmentation.
Front. Psychol. 7:985.
doi: 10.3389/fpsyg.2016.00985

This study investigates whether the learning of prosodic cues to word boundaries in speech segmentation is more difficult if the native and second/foreign languages (L1 and L2) have similar (though non-identical) prosodies than if they have markedly different prosodies (Prosodic-Learning Interference Hypothesis). It does so by comparing French, Korean, and English listeners' use of fundamental-frequency (F0) rise as a cue to word-final boundaries in French. F0 rise signals phrase-final boundaries in French and Korean but word-initial boundaries in English. Korean-speaking and English-speaking L2 learners of French, who were matched in their French proficiency and French experience, and native French listeners completed a visual-world eye-tracking experiment in which they recognized words whose final boundary was or was not cued by an increase in F0. The results showed that Korean listeners had greater difficulty using F0 rise as a cue to word-final boundaries in French than French and English listeners. This suggests that L1–L2 prosodic similarity can make the learning of an L2 segmentation cue difficult, in line with the proposed Prosodic-Learning Interference Hypothesis. We consider mechanisms that may underlie this difficulty and discuss the implications of our findings for understanding listeners' phonological encoding of L2 words.

Keywords: second language, speech segmentation, prosody, eye tracking, French

INTRODUCTION

The segmentation of continuous speech into individual words is a particularly challenging task for non-native listeners, in that cues to word boundaries differ across languages. The cues that may be useful for segmenting the native language (L1) are often inefficient or even misleading for segmenting a second/foreign language (L2). Whether or not non-native listeners can learn to use segmentation cues has been shown to depend in part on the similarity between the L1 and the L2 (e.g., Weber and Cutler, 2006; Al-jasser, 2008; Tremblay et al., 2012; Tremblay and Spinelli, 2014). Unclear, however, is *how* L2 learning is shaped by the degree of similarity between the L1 and the L2. Most existing L2 speech segmentation studies have focused on L1–L2 pairings that differed drastically in how segmentation cues signal word boundaries (e.g., French–English, Japanese–English; Cutler et al., 1992; Cutler and Otake, 1994; Tremblay et al., 2012). It remains to be determined whether segmentation cues such as prosody are more difficult to learn if the L1 and

L2 prosodies pattern in non-identical but similar ways (henceforth, ‘similar[ly]’) in how they signal word boundaries than if they are drastically different. Assessing whether L1–L2 similarity hurts the learning of L2 segmentation cues may in turn shed important light on the cognitive mechanisms that underlie such learning and on L2 learners’ phonological encoding of L2 words.

The present study tests whether the learning of a new segmentation cue is more difficult if the L1 and L2 prosodic systems are similar than if they are markedly different. We will refer to this as the *Prosodic-Learning Interference Hypothesis*. For this hypothesis, similarity is operationalized as a given prosodic cue (e.g., fundamental frequency [F0] rise) signaling the *same word boundary* in both the L1 and the L2 (e.g., F0 rise signals word-final boundaries in both languages). For learning to take place, the L1 and L2 prosodic systems need by definition not to be identical. Hence, the L1 and L2 prosodic systems will be considered similar, though not identical, if a given prosodic cue signals the same word boundary in the L1 and L2 prosodic systems but does so differently (e.g., the alignment of the word-final F0 rise differs between the L1 and the L2). In contrast, the L1 and L2 prosodic systems will be considered different if a given prosodic cue signals *different word boundaries* in the L1 and the L2 (e.g., F0 rise signals word-initial boundaries in the L1 but word-final boundaries in the L2).

Upon initial inspection, the existing literature on non-native speech segmentation appears to suggest that the use of L1 cues is beneficial to L2 speech segmentation when the L1 and L2 pattern similarly. For example, Murty et al. (2007) have shown that listeners whose L1 is Telugu, a Dravidian language that resembles Japanese in its mora-timed rhythm, segment Japanese words similarly to native Japanese listeners, whereas listeners from non-mora-timed L1s (French and English) had not been found to do so (Otake et al., 1993; Cutler and Otake, 1994). Similarly, Kim et al. (2008) have found that listeners whose L1 is Korean, a syllable-timed language, segment French words similarly to native French listeners, whereas listeners from non-syllable-timed L1s (English, Dutch, and Japanese) had not been found to do so (Cutler et al., 1983, 1986; Otake et al., 1996; Cutler, 1997). However, given the difficulty in quantifying rhythmic similarity across languages, the actual degree of similarity between Telugu and Japanese and between Korean and French remains unclear.¹

Prosody, specifically F0 information, may provide a better test case for assessing how the learning of L2 segmentation cues is shaped by the degree of similarity between the L1 and the L2, in that F0 can be measured relatively independently of the segmental content of languages, thus facilitating direct

comparisons across languages.² There are good reasons to hypothesize that the learning of F0 cues may be more difficult if the L1 and L2 prosodic systems are similar than if they are completely different. First, L2 learners may perceive the F0 movement in the L1 and the L2 as identical and thus not readjust their use of segmentation cues. This perceptual assimilation would be similar *in spirit* to Best and Tyler’s (2007) Perceptual Assimilation Model of L2 speech perception (PAM-L2; see also Best, 1995) and to Flege’s (1995) Speech Learning Model (SLM), where L2 learners do not accurately perceive or produce L2 phonemes as a result of assimilation to L1 phonemes. Second, L2 learners may only readjust their use of F0 cues if these unadjusted cues do result in parsing errors, namely in the greater activation of L2 competitor words over L2 target words. In other words, parsing failure may be necessary to trigger L2 learning (for such a proposal, see Carroll, 2004).

The present study tests the Prosodic-Learning Interference Hypothesis by examining how Korean- and English-speaking L2 learners of French use F0 rise to locate phrase-final (thus, also word-final) boundaries in French.³ In French, the last non-reduced syllable of the last content word of the accentual phrase (AP) receives a pitch accent in non-utterance-final position, and the first or second syllable of the first content word in the AP can optionally receive a phrase accent (e.g., Jun and Fougeron, 2000, 2002; Welby, 2006). For example, in [*un gentil chaton*]_{AP} ‘a nice kitty,’ a phrase accent can be aligned with the first syllable of *gentil* and a pitch accent is aligned with the last syllable of *chaton*. The basic underlying tonal pattern of the AP in French is L(HL)H*, where H represents a high phrase accent, H* represents a high pitch accent, and L represents low tones (e.g., Jun and Fougeron, 2002; Welby, 2006). The predominant acoustic cues to (non-utterance-final) pitch accents in French are a rise in F0 and lengthening, whereas the predominant cue to phrase accents is an F0 rise (Welby, 2006). Whereas the F0 in pitch accents rises until the end of the AP-final syllable, the F0 in phrase accents is usually lower, flatter, and more variable in its slope and alignment earlier in the AP. Lengthening and F0 rise aligned with the right edge of the AP-final syllable are thus reliable cues to word-final boundaries in AP-final position in French, whereas a flatter F0 rise earlier in the AP can cue word-initial boundaries (e.g., Christophe et al., 2004; Bagou and Frauenfelder, 2006; Welby, 2007; Spinelli et al., 2007, 2010).

Previous studies have shown that native French listeners locate word-final boundaries at the offset of both lengthened syllables (e.g., Banel and Bacri, 1994; Bagou et al., 2002) and syllables with an F0 rise (e.g., Bagou et al., 2002; Bagou and Frauenfelder,

¹ Existing acoustic metrics of rhythm such as the normalized Pairwise Variability Index (*nPVI*; Low et al., 2000) and the proportion of speech time dedicated to vocalic intervals (%V; Ramus et al., 1999) have not compared these languages, and even if they had, inconsistencies in how these metrics map different languages in a rhythmic space (e.g., Grabe and Low, 2002 vs. Ramus, 2002) would likely make it difficult to draw firm conclusions on the degree of similarity between the above language pairs. The fact that these metrics additionally reflect syllable structure differences among languages could also make these comparisons difficult (e.g., Korean and French have different syllable structures).

² We do not seek to claim that F0 plays a more important role than other prosodic cues (e.g., duration) in speech segmentation (in French or across languages). Ultimately, F0 is only one of the prosodic cues that contribute to signaling word boundaries, and it is only one of the cues through which the prosodic system of the language is realized. We focus on F0 because it provides an easier and clearer test of the hypothesis that the learning of L2 segmentation cues is shaped by the degree of similarity between the L1 and the L2 (particularly for the language pairs selected for this study).

³ F0 rise does not cue word-final boundaries within phrases in French. Throughout the paper, we will refer to the use of F0 cues to word-final boundaries in French with the understanding that such cues occur in phrase-final position and thus signal word-final boundaries only in phrase-final position.

2006). Christophe et al. (2004) provided further evidence that phrase-final prosodic boundaries (and pitch accents) mediate lexical access in French. They found that monosyllabic words (e.g., *chat* [ʃa] 'cat') were recognized more slowly when they were temporarily ambiguous with a competitor word created segmentally between the monosyllabic word and the first syllable of the following word (e.g., *chagrin* [ʃagʁɛ̃] 'heartache' in [*dun chat grinceux*]_{AP} [dɛ̃ʃagʁɛ̃sɔ̃] 'of a cranky cat') than when they were not temporarily ambiguous with such a competitor (e.g., [*dun chat drogué*]_{AP} [dɛ̃ʃadʁɔʒe] 'of a drugged cat'; [ʃadʁɔ] is not a French word); however, if the monosyllabic word was at an AP-final boundary and thus received a pitch accent (e.g., [*le gros chat*]_{AP} [gʁɔʃa] *grimpeait aux arbres*]_{AP} [lɛgʁɔʃa gʁɛ̃pɛʁɔzɑ̃bʁɛ̃] 'the big cat was climbing trees'), the target word was no longer recognized more slowly when it was temporarily ambiguous with a phonemic competitor than when it was not (e.g., [*le gros chat*]_{AP} [*dressait l'oreille*]_{AP} [lɛgʁɔʃadʁɛ̃sɛlɔʁɛ̃] 'the big cat was sticking up his ears'; [ʃadʁɛ̃] is not a French word). These findings suggest that phrase-final boundaries, marked with a pitch accent and thus realized with both lengthening and an F0 rise, act as filter and constrain lexical access (see also Michelas and D'Imperio, 2010). In an artificial-language segmentation study, Tyler and Cutler (2009) also showed that French listeners independently use F0 and duration cues to word-final boundaries.

Korean is similar to French in that prominence is also at the level of the AP. In the Seoul dialect, the basic underlying tonal pattern of the AP is (LH)LH or (HH)LH, with the first tone being H if the first sound is tense or aspirated and L otherwise (e.g., Jun, 1995, 1998, 2000; Beckman and Jun, 1996). For example, in [jɔŋman-ine-nin]_{AP} 'youngman-family-topic,' the first H is "loosely aligned" with the second syllable of the phrase and the second H is aligned with the final syllable of the phrase (Jun, 1998, pp. 195, 196). Thus, like French, Korean has an H tone on the AP-final (and thus word-final) syllable, which can cue word-final boundaries in that AP-final position. However, unlike French, the phrase-final F0 rise in Korean peaks before the syllable offset and begins decreasing thereafter such that it is already low in the next syllable, whereas in French the F0 begins decreasing *after* the accented syllable (cf. Jun, 2000, p. 21, vs. Jun and Fougeron, 2002, p. 163). Korean also differs from French in that lengthening does not consistently cue AP-final boundaries in Korean (cf. Oh, 1998 and Cho and Keating, 2001, vs. Jun, 1993; Chung et al., 1996); however, syllables at the end of the intonational phrase (IP) are consistently lengthened in both Korean (e.g., Jun, 1993, 1995, 1998, 2000; Cho and Keating, 2001) and French (e.g., Jun and Fougeron, 2000, 2002). In that sense, French and Korean are similar but not identical in how they cue word-final boundaries.

Like French listeners, Korean listeners use prosodic cues to phrase-final accents for locating word-final boundaries in continuous speech. In an artificial-language segmentation study, Kim et al. (2012) showed that Korean listeners use both F0 and lengthening as cues to word-final boundaries. Similarly, in word-spotting experiments, Kim and Cho (2009) demonstrated that Korean listeners recognized (prototypical) LH words more easily when these words were preceded by a syllable containing an H tone than when they were preceded by an L tone; however, the same was not true of (atypical) HL words that were preceded

by a syllable containing an L tone. Kim and Cho (2009) further showed that the L tone at the onset of the target disyllabic words was not helpful for segmentation if it was not preceded by an H tone, suggesting that it is the contrast in F0 tones that enhances Korean listeners' segmentation of Korean speech, but only if H is in word-final position. Kim and Cho (2009) also showed that Korean listeners benefited from lengthening at least under some circumstances.⁴

English differs from both French and Korean in that prominence is lexical rather than phrasal, and pitch accents are aligned with stressed syllables and they are not necessarily phrase-final (e.g., Beckman and Elam, 1997). Statistically, stress tends to be word-initial rather than word-final, especially in nouns (e.g., Cutler and Carter, 1987; Clopper, 2002). Stress in accented words thus provides a somewhat reliable cue to word-initial boundaries in English (e.g., Cutler and Butterfield, 1992; McQueen et al., 1994; Mattys, 2004). The primary prosodic correlates of stressed syllables in accented English words are F0 rise, increased duration, and greater intensity (e.g., Lieberman, 1960; Beckman, 1986), and the importance of each depends in part on the location of the accented syllable in the word (e.g., Tremblay and Owens, 2010) and on the location of the word in the phrase (e.g., Tyler and Cutler, 2009). It is thus the case that English is quite different from French in how prosodic cues signal word boundaries.

English listeners tend to parse accented syllables as word-initial. This was shown in a variety of experimental paradigms (e.g., juncture perception task: Cutler and Butterfield, 1992; word-spotting tasks: McQueen et al., 1994; cross-modal priming tasks: Mattys, 2004; Mattys et al., 2005). However, because stress is strongly correlated with vowel quality in English, English listeners make limited use of prosodic cues to stress in the absence of segmental cues to stress (e.g., Cutler and Clifton, 1984; Cutler, 1986; Small et al., 1988; Fear et al., 1995; Cooper et al., 2002). When English listeners do use prosodic cues to word boundaries, they associate F0 rise with word-initial boundaries (Tyler and Cutler, 2009). This is indeed what we should expect given the statistical tendency for stress to occur word-initially. Interestingly, English listeners also appear to associate lengthening to word-final boundaries (Tyler and Cutler, 2009), suggesting that different prosodic cues can signal different word boundaries in English. Tyler and Cutler (2009) attribute the facilitative effect of word-final lengthening to the phrase- (and thus word-) final lengthening that occurs in English and many other languages (see also Vaissière, 1983; Hayes, 1995).

The similarities and differences among French, Korean, and English allow us to test whether the learning of a new segmentation cue is more difficult if the L1 and L2 prosodic systems are similar than if they are markedly different. French and Korean pattern similarly in that their prosody is phrasal, and for words in AP-final position, word-final boundaries are cued by an F0 rise; yet, they differ in that the AP-final F0 peak is aligned differently in the two languages (earlier in Korean,

⁴They did if the following word had atypical prosody (e.g., HL) but not if it had prototypical prosody (e.g., LH). The authors attribute the asymmetrical effect of lengthening to a segmental confound in their design.

later in French). In contrast, English differs from both French and Korean in that prominence is lexical and F0 rise signals word-initial rather than word-final boundaries. If the learning of a new segmentation cue is more difficult when the L1 and L2 prosodic systems are similar than when they are different, Korean L2 learners of French should have more difficulty in using F0 cues to word-final boundaries in French than *both* native French listeners and English L2 learners of French.

In a word-monitoring experiment, Tremblay et al. (2012) examined French and English listeners' use of F0 and duration cues to word-final boundaries in French. In an adaptation of Christophe et al. (2004), they asked native French listeners and mid- and high-proficiency English L2 learners of French to monitor disyllabic words that were not in the stimuli but that were created phonemically between a monosyllabic noun and the first syllable of the following word (e.g., *chalet* 'cabin' in *chat lépreux* 'grumpy cat'). In the across-AP condition, the monosyllabic word in the stimuli (e.g., *chat*) received a pitch accent, and thus the disyllabic word to be monitored (e.g., *chalet*) crossed an AP boundary (e.g., *[[Le chat]_{AP} [lépreux et légendaire]_{AP}]_{PP} s'endort doucement* 'The leprous and legendary cat is slowly falling asleep'); in the within-AP condition, the monosyllabic word in the stimuli (e.g., *chat*) was not accented, and thus the disyllabic word to be monitored (e.g., *chalet*) was located within an AP (e.g., *[[Le chat lépreux]_{AP}]_{PP} s'endort doucement* 'The leprous cat is slowly falling asleep'). If prosody constrained lexical access, participants should show fewer detections of the disyllabic word to be monitored (i.e., fewer false alarms) in the across-AP condition than in the within-AP condition. Experiment 1 used natural stimuli; in Experiment 2, stimuli were resynthesized such that the F0 was swapped between the across-AP and within-AP conditions, thus making it possible to examine the effect of F0 cues independently of duration cues. Different participants at similar proficiencies completed Experiments 1 and 2.

The results of Experiment 1 showed that the high-level L2 learners and native listeners, but not the mid-level L2 learners, had fewer false alarms in the across-AP condition than in the within-AP condition, indicating that sufficiently advanced English L2 learners of French could parse accented syllables as word-final. However, the results of Experiment 2 showed that only the native listeners were able to use F0 cues to word-final boundaries. These results suggest that unlike French listeners, English listeners were not able to use F0 rises as a cue to word-final boundaries in French; they could only use duration as a cue to word-final boundaries, but only if they were sufficiently proficient in French (for details, see Tremblay et al., 2012).

The present study uses the same stimuli as those used in Experiment 2 of Tremblay et al. (2012), but in a visual-world eye-tracking experiment, thus shedding light on the time course of activation of target and competitor words as listeners hear F0 cues to word-final boundaries. We examine the segmentation of French speech by native French listeners and by *both* English and Korean L2 learners of French, with the L2 listeners being matched not only in their French proficiency, but also in *all* their language background information. Thus, if any difference is found between the L2 groups, such a difference could only be attributed to the

participants' L1. The use of eye tracking will allow us to determine not only if Korean L2 learners of French have more difficulty than English L2 learners of French in using F0 cues to word-final boundaries in French, but also if English L2 learners of French can in fact learn to use F0 cues to word-final boundaries in French, something that was not found in Tremblay et al. (2012).

MATERIALS AND METHODS

Ethics Statement

The study was approved by the Human Subjects Committee of the University of Kansas, Lawrence. Participants read and signed a written consent form. No vulnerable population was involved.

Participants

Twenty-five native French listeners (mean age: 26.4, SD: 4.6), 16 English L2 learners of French (mean age: 23.9, SD: 0.9), and 16 Korean L2 learners of French (mean age: 23.3, SD: 8.2) participated in this study. The English listeners were undergraduate or graduate students at a Midwestern university in the US who either majored in French or identified themselves as having functional proficiency in French. The Korean listeners were undergraduate students majoring in French or in French-Korean translation at a university in Seoul, Korea.⁵ All participants had normal or corrected-to-normal vision, and no participants reported any hearing impairment. All participants received monetary compensation or course credit in exchange for their time.

The L2 learners filled out a language background questionnaire and completed a cloze test that would assess their proficiency in French (Tremblay, 2011). Their language background information and proficiency scores are summarized in Table 1. The English and Korean listeners were matched in both their experience with French and their proficiency in French.⁶ One-way ANOVAs with L1 as between-group variable did not reveal significant differences between the two groups on any of the language background variables or on the proficiency scores ($p > 0.1$).

All Korean listeners also had some knowledge of English. On a scale from 1 to 4 (1 = beginner, 2 = intermediate, 3 = advanced, 4 = near-native), they rated their English-listening skills as

⁵Among the Korean listeners, 2 were native speakers of the Chungcheong dialect, 2 were native speakers of the Kyungsang dialect, and 1 was a native speaker of the Gangwon dialect; all other Korean listeners were native speakers of the Seoul dialect. Kyungsang Korean differs from Seoul Korean in that it has lexical pitch accents, with some accents ending with an L tone rather than an H tone (e.g., Lee and Zhang, 2014). However, recent research suggests that young Kyungsang Korean speakers are in the process of losing these different pitch accents, possibly due to the close contact with and influence of Seoul Korean (Lee and Jongman, 2015). To ensure that our two Kyungsang Korean speakers did not drive any of the results, we ran our statistical analyses (presented further below) both with and without these two speakers. The variables that reached significance were exactly the same in the two different analyses. We therefore kept these two speakers in our analyses.

⁶More Korean and English L2 learners of French were tested, but a subset of each group was selected so that they would match in both their proficiency in and experience with French.

TABLE 1 | L2 learners' language background information and proficiency scores.

	AFE ^a	YrsInstr ^b	MthsRes ^c	%Use ^d	Cloze
English L2 learners of French (<i>n</i> = 16)	16.8 (4.3)	6.2 (3.2)	14.0 (23.5)	13.7 (10.2)	23.3 (8.2)
Korean L2 learners of French (<i>n</i> = 16)	18.8 (2.1)	5.7 (2.2)	14.3 (15.8)	12.5 (10.6)	21.7 (5.7)

Mean (standard deviation). ^aAge of first exposure to French. ^bNumber of years of formal instruction on French. ^cMonths of residence in a French-speaking environment. ^dPercent weekly use of French.

similar to their French-listening skills (English: mean: 2.6, SD: 0.7; French: mean: 2.4, SD: 0.7; $t < |1|$).

Materials

All stimuli came from Tremblay et al. (2012). Participants heard sentences in which a competitor word was created segmentally between a monosyllabic target word and the first syllable of the disyllabic adjective following it (e.g., *chalet* 'cabin' in *chat lépreux* 'leprous cat'). In the across-AP condition, the monosyllabic target word received a pitch accent, and the disyllabic competitor word crossed an AP boundary (e.g., *[[Le chat]_{AP}][lépreux et légendaire]_{AP}]* *s'endort doucement* 'The leprous and legendary cat is slowly falling asleep'). The first AP contained an LH* tonal pattern, with the L tone belonging to either a phrase-initial accent or a pitch accent and the H* tone belonging to a pitch accent. In the within-AP condition, the pitch accent instead fell on the last syllable of the post-nominal adjective (e.g., *[[Le chat lépreux]_{AP}]* *s'endort doucement* 'The leprous cat is slowly falling asleep'). The AP in this condition contained an LLH* tonal pattern, with the first L tone belonging to a phrase-initial accent and the LH* tones belonging to a pitch accent.

The auditory stimuli were recorded by a female native speaker of French from Bordeaux (France) using a Marantz PMD 750 solid-state recorder and head-mounted condenser microphone. The speaker was trained to produce the stimuli such that an H* tone would fall on the monosyllabic noun in the across-AP condition but on the last syllable of the post-nominal adjective in the within-AP and control conditions. In both experimental conditions, the peak F0 of the H* tone was aligned with the AP-final boundary. The H* tone produced on the monosyllabic noun in the across-AP condition was not followed by a pause so that the disyllabic competitor word could be erroneously detected.

Next, the F0 contours of the across-AP and within-AP conditions were resynthesized such that the F0 of the first four syllables was swapped between the two experimental conditions. The first four syllables of the resynthesized across-AP sentences thus contained the F0 contour of the corresponding syllables in the within-AP condition, and the first four syllables of the resynthesized within-AP sentences contained the F0 contour of the corresponding syllables in the across-AP condition. This manipulation, which made it possible to examine the effect of F0 rise independently of duration, resulted in four conditions: (i) an across-AP condition with F0 rise (natural); (ii) an across-AP condition without F0 rise (F0 rise removed); (iii), a within-AP condition with F0 rise (F0 rise added); and (iv) a within-AP condition without F0 rise (natural).

The experimental stimuli were resynthesized using close-copy stylization (e.g., de Pijper, 1983). The first four syllables of the experimental items were divided into 20 segments each, and

the average F0 of each segment was extracted. The existing pitch points in each segment were then dragged vertically using the Pitch Synchronous Overlap-Add (PSOLA) method in Praat (Boersma and Weenink, 2004) so that they would approximate the value of the extracted average in the corresponding segment of the opposite condition. After the initial resynthesis, the pitch contour of the natural and resynthesized conditions were closely examined, and resynthesized contours that were judged not to be sufficiently similar to the natural contours of the opposite condition were altered so that they would better approximate them. Once the contours were judged to be satisfactory, a stop Hann-band filter from 500 to 1,000 Hz with a smoothing of 100 Hz was applied to all the stimuli to mask the occasionally robotic sound that resulted from the F0 manipulation. This filter did not significantly affect the segmental quality of the stimuli. **Figure 1** shows an example of natural and resynthesized stimulus in the across-AP and within-AP conditions (adapted from Figure 4 of Tremblay et al., 2012).

Acoustic analyses of the first two syllables in the stimuli (e.g., *le chat*) performed in Praat (Boersma and Weenink, 2004) are reported in Tremblay et al. (2012). In brief, these analyses revealed that the prosodic cue manipulation was successful, with the resynthesized monosyllabic word (e.g., *chat*) having a significantly different F0 in the across-AP and within-AP sentences.

The experiment included a total of 32 experimental stimuli randomly interspersed with the 69 filler stimuli, 8 of which were used in the practice session. The participants were assigned to one of four lists and saw each experimental item in only one condition (total: 8 items per condition; for the complete list of experimental items, see Tremblay et al., 2012).

Participants saw four words on the computer display and clicked on the word they thought they heard. In the experimental stimuli, the display included the target (monosyllabic) word (e.g., *chat*), the competitor (disyllabic) word (e.g., *chalet*), and two distracter words. To ensure that the participants would not be biased in their fixations toward the target and competitor words (given their segmental overlap), the distracter words also overlapped together in their segmental content. These distracter words were either both monosyllabic (e.g., *clé* 'key' and *craie* 'chalk'; 6 items), both disyllabic (e.g., *chemin* 'path' and *cheval* 'horse'; 6 items), or one of each (e.g., *prince* 'prince' and *principe* 'principle'; 20 items), and they did not overlap segmentally or semantically with the target and competitor words. Since the words across the four prosodic conditions are identical, L2 learners' familiarity with the words in the display cannot explain any prosodic effect that we may find (for discussion, see Tremblay et al., 2012).

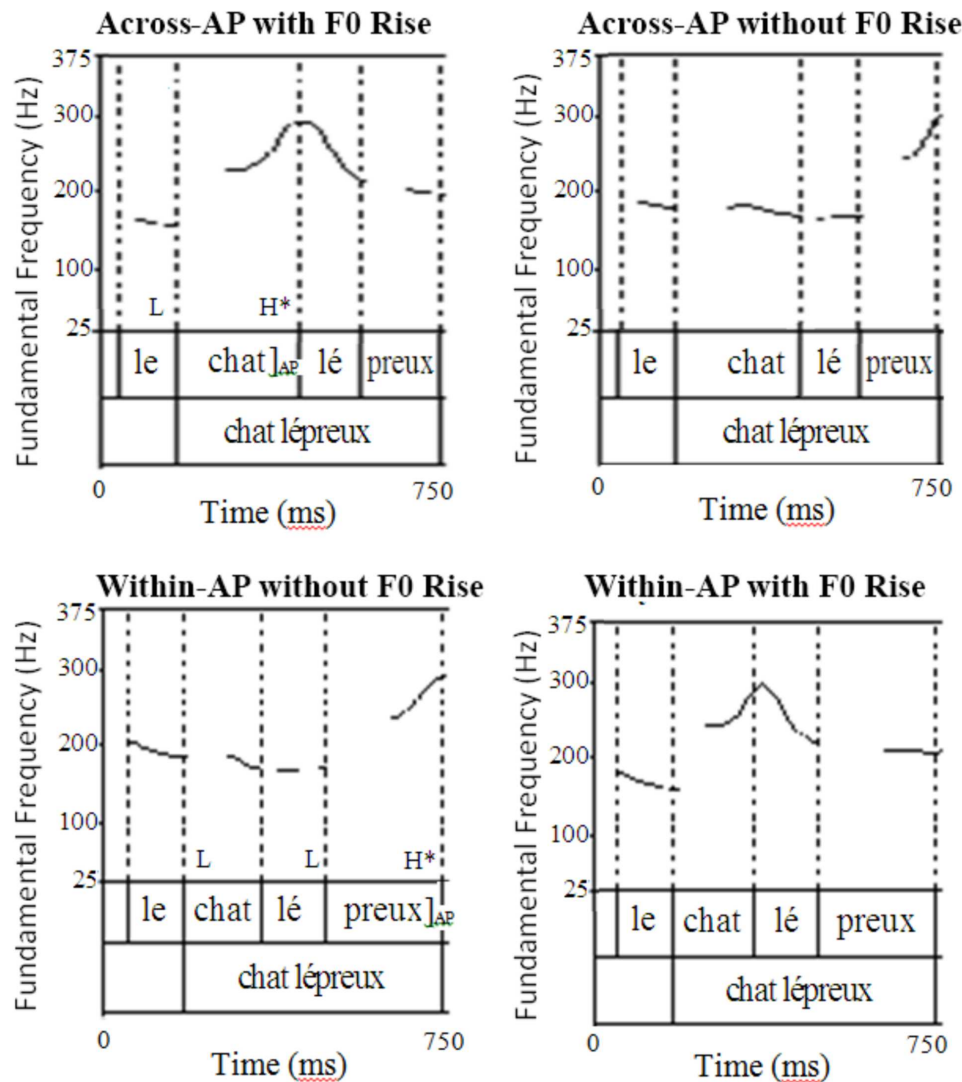


FIGURE 1 | Example of natural and resynthesized stimuli (adapted from Figure 4 of Tremblay et al., 2012).

All words in the visual display were presented orthographically (for a validation of this method, see Huetig and McQueen, 2007; McQueen and Viebahn, 2007). It was decided to present the words orthographically rather than with images, first because not all the experimental words were easily imageable, and second to facilitate the task with L2 learners, who may not have equal familiarity with all the words in the experiment. Since prosody is independent of word spelling in French, this characteristic of our experimental design does not pose any concern.

Procedures

The eye-tracking experiment was designed and compiled with Experiment Builder software (SR Research), and the participants' eye movements were recorded with an Eyelink eye tracker (SR Research) at a sampling rate of either 250 Hz or 1,000 Hz, depending on the location of the data collection. An

ASIO-compatible sound card was used on the display computer to ensure that the audio timing would be accurate.

The experiment began with a calibration of the eye tracker using the participants' right eye. If the eye tracker could not be successfully calibrated with the participant's right eye, his/her left eye was instead used. This initial calibration was followed by a practice session (8 trials) and by the main experiment (93 trials). In each trial, the participants saw four orthographic words in a (non-displayed) 2×2 grid for 4,000 ms. The words then disappeared and a fixation cross appeared in the middle of the screen for 500 ms. As the fixation cross disappeared, the four words reappeared on the screen in their original position and the auditory stimulus was heard (synchronously) over headphones. The participants were instructed to click on the target word with the mouse as soon as they heard the target word in the stimulus. The participants' eye movements were measured from the onset of the target word (e.g., the onset of *chat*). The trial

ended with the participants' response, with an inter-trial interval of 1,000 ms.

The 32 experimental and 61 filler trials were pseudo-randomized and presented in four blocks (23 trials per block, except for one block that contained 24 trials). Each block contained 8 experimental trials (2 from each condition). Both the order of the experimental and filler trials within a block and the order of blocks were randomized across participants. The participants took a break after completing the second block. The eye tracker was calibrated at the beginning of each block and whenever it was necessary during the experiment. The participants completed the experiment in approximately 15–20 minutes.

Data Analysis and Predictions

Experimental trials that received distracter responses (rather than target or competitor responses) or for which eye movements could not reliably be tracked were excluded from the analyses. This resulted in the exclusion of 6.4% of all trials (2.7% for French listeners, 1.5% for Korean listeners, and 2.2% for English listeners). For the remaining trials, we analyzed the participants' eye movements in each of the four regions of interest (corresponding to the four orthographic words on the screen).

Proportions of fixations to the target, competitor, and distracter words were extracted in 8-ms time windows from the onset of the target word to 1,500 ms after the target word. To better capture any effect of lexical competition due to the manipulated F0 cues, statistical analyses were conducted on the *difference* between target and competitor fixations (i.e., competitor fixations were subtracted from target fixations). This difference factors out any difference in the speed with which participants begin to fixate both target and competitor words (over distracter words), thus making the data more comparable between native listeners and L2 learners.

Listeners' fixation differences were modeled using growth curve analysis (GCA; Mirman, 2014). GCAs are similar to mixed-effects models (for discussion, see Baayen, 2008), but they also include time coefficients, thus enabling researchers to model participants' fixations over time. GCA is ideal for analyzing participants' proportions of fixations as the speech signal unfolds, because they can model cross-over effects in fixations that cannot always be captured in traditional time-window analyses of eye-tracking data. For example, if Fixation Line A is 10% higher than Fixation Line B from 200 to 300 ms but 10% lower than Fixation Line B from 400 to 500 ms (with the two lines intersecting at 350 ms), a time window analysis that averages fixations from 200 to 500 ms would likely show no difference between the two lines, when in fact the directionality of the effect evidenced by the two lines reversed half way through the time window. GCAs can thus model subtle changes in the curvilinear patterns of eye fixations over time, capturing differences in the slope and curvature of the fixation lines. GCAs also have the advantage of not requiring (potentially arbitrary) decisions regarding the critical time window for the statistical analysis.

GCAs include orthogonal time coefficients, the fixed variables of interest, and random variables. The time coefficients model the shape of the proportions of fixations over time. In a visual-world

eye-tracking paradigm, the difference between participants' target and competitor fixations typically takes the form of an 's'-shaped (i.e., cubic) line, with fixations initially being flat (and sometimes decreasing depending on the degree of competition), then increasing in a steady slope, and finally leveling off. The analysis in this study thus includes linear, quadratic, and cubic time coefficients. The time coefficients are centered, and they are made orthogonal prior to entering them in the analyses because these time coefficients would otherwise be highly correlated, which would make the model unstable and the results difficult to interpret. The fixed variables in GCAs are those of the experimental design.

The results of GCAs are interpreted as follows: For the researcher to be able to conclude that a manipulation of the speech signal resulting in two different conditions has an effect on participants' fixations, the GCA must show an *interaction* between this manipulation and at least one of the time coefficients. Such an interaction indicates that as the speech signal unfolds over time, the shape of participants' fixation line changes differently for the two conditions. Finding only an effect of experimental variable and no interaction between it and any of the time coefficients indicates that fixation proportions are higher or lower in one condition than in another, but the shape of participants' fixation lines is similar across the two conditions. Hence, such an effect could not be attributed to any manipulation of the speech signal (i.e., such an effect would be better interpreted as a baseline effect). The GCAs in the present study included two fixed variables: whether or not the word-final boundary was signaled by an F0 rise (within-participant), and the native language of the participants (between-participant), with native French listeners being compared to English and Korean L2 learners of French in a first analysis and with the L2 groups being compared to each other in a second analysis. Because the test items in the across-AP and within-AP conditions differ in their duration, they cannot be compared directly in a GCA analysis of listeners' proportions of fixations. Hence, we examined the effects of F0 rise and L1 separately for the across-AP and within-AP conditions.

Like mixed-effects models, GCAs can also include crossed random variables. The GCAs proposed by Mirman (2014) include participant as random intercept and the orthogonal time coefficients as random slopes for the participant variable, thus allowing the analysis to model a line of a different shape for each participant. Such an analysis is ideal to capture between-participant variability in their fixations over time.⁷

The GCAs were run using the lme4 package in R (Bates et al., 2015). The initial analysis included the three time coefficients, the presence or absence of a word-final F0 rise, listeners' L1, and all interactions as fixed effects, and it included participant as random intercept and the three time coefficients as random slope for the participant variable. The fixed effects other than the three time

⁷Given the complexity of this analysis and the much larger dataset it runs on (as compared to those used in typical mixed-effect models), adding additional random variables to the analysis requires significant computing power, with each analysis taking several hours to run. For this reason, in the present study, only participant is used as random intercept. The above fixed variables are not added as random slopes, because the analyses often fail to converge.

coefficients were then removed from the model one at a time, and model comparisons were run in pairwise fashion to determine if the more complex model accounted for significantly more of the variance, as determined by log-likelihood ratio tests. We report the simplest model including the three time coefficients that accounted for significantly more of the variance than simpler models. If the best model yielded significant interactions involving L1, follow-up GCAs were conducted separately for each of the L1. The alpha level of these subsequent models was adjusted manually using the Bonferroni correction.

If the presence of a word-final F0 rise enhances speech segmentation, the GCAs should yield both an effect of F0 rise (with larger difference between target and competitor fixations in the condition with an F0 rise than in the condition without such a rise) and an interaction between this F0 rise and at least one of the three time coefficients, indicating that the differential fixation lines in the condition with vs. without an F0 rise have different shapes as the speech signal unfolds. If participants' L1 modulates their ability to use this F0 rise, the GCAs should yield three-way interactions between the presence and absence of a word-final F0 rise, participants' L1, and at least one of the time coefficients.

RESULTS

French, English, and Korean listeners' proportions of target, competitor, and distracter fixations in the across-AP and within-AP conditions are presented in the figures in Appendix A in the supplemental data.

Across-AP Condition

All Listeners

Recall that the across-AP condition naturally contained an F0 rise, and F0 was resynthesized such that it would be flat. The best GCA on the difference between listeners' proportions of target and competitor fixations in the across-AP condition included all simple effects and all interactions. The results of this GCA and the interpretation of the GCA coefficients can be found in

Appendix B in the supplemental data (Table B1). The modeled differences between target and competitor fixations (henceforth, differential fixations) are illustrated in **Figure 2**.

Among other effects, the GCA yielded significant three-way interactions between F0, L1, and the time coefficients. In order to understand the directionality of these three-way interactions, subsequent GCAs were performed on the differential proportions of fixations separately for each L1 group. For French and English listeners' differential fixations, these subsequent GCAs with all simple effects and all interactions had the best fit. For Korean listeners' differential fixations, the best GCA included F0 and the time coefficients, but no interaction between them. The results of these subsequent GCAs are presented in **Table 2**. For each group, the baseline is the difference between the proportions of target and competitor fixations in the condition *with* an F0 rise (i.e., the natural speech condition). Because the time coefficients were made orthogonal, any effect of a fixed variable (e.g., F0, L1) is to be interpreted on the averaged differential fixations over time (Mirman, 2014).

French Listeners

For the GCA on French listeners' data, the significant positive t value for the quadratic time coefficient indicates that French listeners' differential fixation line in the condition with an F0 rise had a convex shape. The significant negative t value for F0 means that French listeners had a lower differential proportion of fixations in the condition without an F0 rise than in the condition with an F0 rise. Crucially, the significant positive t values for the interaction between F0 and the linear and quadratic time coefficients indicate that French listeners had a differential fixation line that had more of an ascending slope and was more convex in the condition without an F0 rise than in the condition with an F0 rise. Furthermore, the significant negative t value for the interaction between F0 and the cubic time coefficient means that the French listeners' differential fixation line in the condition without an F0 rise had more of a canonical 's' shape than their differential fixation line in the condition with an F0 rise.

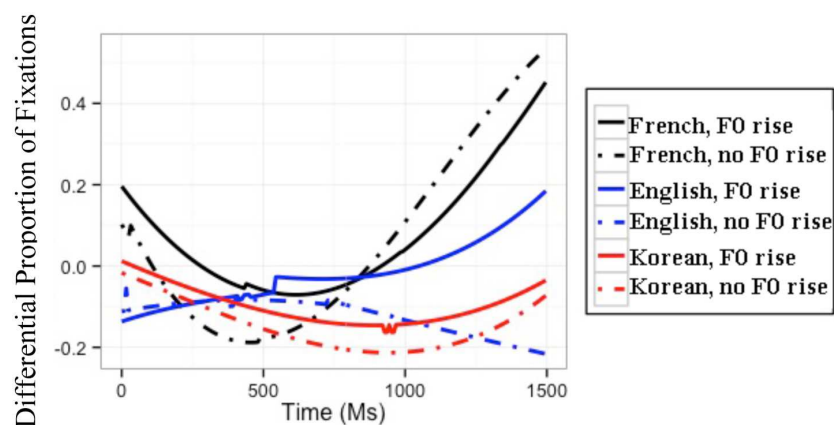


FIGURE 2 | Modeled difference between proportions of target and competitor fixations for the three L1 groups in the across-AP condition. The difference is calculated as the fixations to targets minus those to competitors; hence, positive values reflect more fixations to targets than to competitors.

These results can be observed in the modeled differential proportions of fixations in **Figure 2**: In the absence of an F0 rise, French listeners showed lower differential proportions of fixations, thus more lexical competition, during the first 750 ms *post* target-word onset, after which fixations became more similar between the two F0 conditions. The absence of an F0 rise thus modulates French listeners' fixations early on in the word recognition process, making it more difficult to locate the word-final boundary and resulting in increased lexical competition.

English Listeners

For the GCA on English listeners' data, the significant negative *t* value for F0 indicates that English listeners had a lower differential proportion of fixations in the condition without an F0 rise than in the condition with an F0 rise. Importantly, the significant negative *t* values for the interaction between F0 and the linear and quadratic time coefficients mean that English listeners' differential fixation line had more of a descending slope and more of a concave shape in the condition without an F0 rise than in the condition with an F0 rise.

These results can also be seen in the modeled differential proportions of fixations in **Figure 2**: English listeners' differential proportions of fixations in the two F0 conditions were similar up until 500 ms *post* target-word onset, after which English listeners showed lower differential proportions of fixations, thus more lexical competition, in the condition without an F0 rise than in the condition with an F0 rise. The absence of F0 rise thus modulates English listeners' fixations later on in the word recognition process. In other words, English listeners could incorporate the use of F0 cues to word-final boundaries in the segmentation of French speech (unlike the results of Tremblay et al., 2012), but did so later than French listeners.

Korean Listeners

Finally, for the GCA on Korean listeners' data, the significant negative *t* value for the intercept means that Korean listeners' differential proportion of fixations in the condition with an F0 rise was lower than 0, and the significant negative *t* value for F0 indicates that Korean listeners had a lower differential proportion of fixations in the condition without an F0 rise than in the condition with an F0 rise. Importantly, the interaction between F0 and time did not make it to the model, indicating that Korean listeners' use of F0 did not change as a function of time.

These results are illustrated in the modeled differential proportions of fixations in **Figure 2**: Although Korean listeners showed an effect of F0 in the predicted direction, this effect of F0 did not change as the speech signal unfolded. The effect of F0 can therefore not be attributed to Korean listeners' intake of the speech signal.

L2 Listeners

To ascertain whether Korean listeners differed significantly from English listeners in their differential proportions of fixations, an additional GCA was run only on the L2 learners' data in the across-AP condition, with the English group as baseline. The model with the best fit included all simple effects and all

interactions. The results of this GCA and the interpretation of the GCA coefficients can be found in Appendix B in the supplemental data (Table B2). In brief, this GCA revealed significant three-way interactions between L1, F0, and the linear and quadratic time coefficients, indicating that English and Korean listeners differed from each other in the effect of F0 as a function of time: English listeners' differential fixation lines for the two F0 conditions differed in their quadratic shape; Korean listeners did not show this effect (cf. **Table 2**).

Within-AP Condition

All Listeners

Recall that the within-AP condition naturally did *not* contain an F0 rise, and the flat F0 was resynthesized such that the target word would end with an F0 rise. The best GCA on the difference

TABLE 2 | Growth curve analyses on the difference between listeners' target and competitor fixations in the across-AP condition separately for French, English, and Korean listeners.

Group	Variable	Estimate (SE)	<i>t</i>
French	(Intercept)	0.094	2.566
	Time		
	Linear	0.783	1.513
	Quadratic	1.856	3.912 **
	Cubic	−0.214	< 1
	F0	−0.016	−2.580 *
	Time × F0		
	Linear	1.247	14.928 ***
	Quadratic	0.202	2.432 *
	Cubic	−0.492	−5.891 ***
English	(Intercept)	0.087	< 1
	Time		
	Linear	−0.902	< 1
	Quadratic	1.523	1.179
	Cubic	−0.252	< 1
	F0	−0.099	−11.931 ***
	Time × F0		
	Linear	−1.501	−13.211 ***
	Quadratic	−0.566	−4.992 ***
	Cubic	−0.116	−1.021
Korean	(Intercept)	−0.097	−3.486 **
	Time		
	Linear	−0.354	< 1
	Quadratic	0.588	2.214
	Cubic	0.100	< 1
	F0	−0.051	−7.482 ***

$\alpha = 0.0167$; * = $p < 0.0167$; ** = $p < 0.003$; *** = $p < 0.0003$. French model: $n = 25$; 9,236 observations. English model: $n = 16$; 5,833; observations. Korean model: $n = 16$; 6,012 observations.

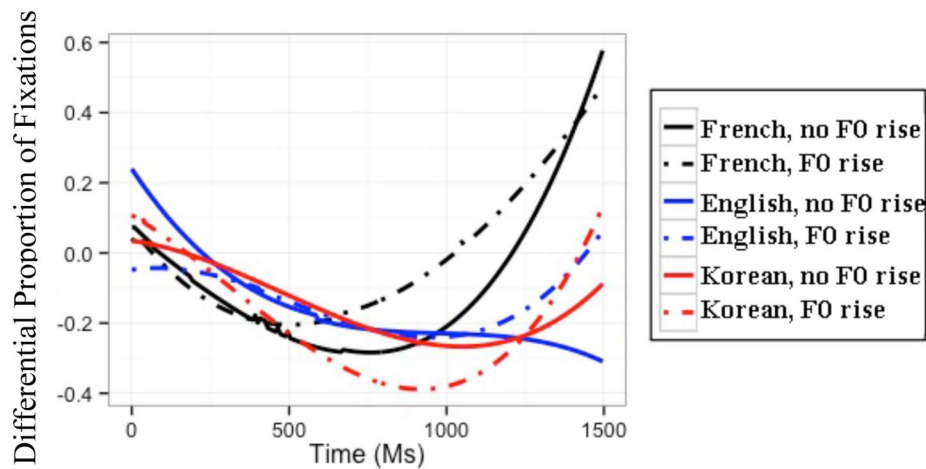


FIGURE 3 | Modeled difference between proportions of target and competitor fixations for the three L1 groups in the within-AP condition. The difference is calculated as the fixations to targets minus those to competitors; hence, positive values reflect more fixations to targets than to competitors.

between listeners' proportions of target and competitor fixations in the within-AP condition included all simple effects and all interactions. The results of this GCA and the interpretation of the GCA coefficients can be found in Appendix B in the supplemental data (Table B3). The modeled differences between target and competitor fixations are illustrated in **Figure 3**.

Among other effects, the GCA yielded significant three-way interactions between F0, L1, and the time coefficients. Again, in order to understand the directionality of these three-way interactions, subsequent GCAs were run on the differential proportions of fixations separately for each L1 group. For French, English, and Korean listeners' differential fixations, these subsequent GCAs with all simple effects and all interactions had the best fit. The results of these subsequent GCAs are presented in **Table 3**. For each group, the baseline is the difference between the proportions of target and competitor fixations in the condition *without* an F0 rise (i.e., the natural speech condition). Again, because the time coefficients were made orthogonal, any effect of a fixed variable (e.g., F0, L1) is to be interpreted on the averaged differential fixations over time (Mirman, 2014).

French Listeners

For the GCA on French listeners' data, the significant positive t value for the quadratic time coefficient indicates that French listeners' differential fixation line in the condition without an F0 rise had a convex shape. The significant positive t value for F0 means that French listeners had a higher differential proportion of fixations in the condition with an F0 rise than in the condition without an F0 rise. Crucially, the significant positive t value for the interaction between F0 and the linear time coefficient indicates that French listeners had a differential fixation line that had more of an ascending slope in the condition with an F0 rise than in the condition without an F0 rise. Furthermore, the significant negative t values for the interaction between F0 and the quadratic and cubic time coefficients mean that French listeners had a differential fixation line that was less convex and

had more of a canonical 's' shape in the condition with an F0 rise than in the condition without an F0 rise.

These results can be seen in the modeled differential proportions of fixations in **Figure 3**: From 500 ms *post* target-word onset, French listeners showed higher differential proportions of fixations, thus less lexical competition, in the presence of an F0 rise than in the absence of an F0 rise. The presence of an F0 rise thus modulates French listeners' fixations later on in the word recognition process, making it easier to locate the word-final boundary and resulting in decreased lexical competition.

English Listeners

For the GCA on English listeners' data, the significant negative t value for the intercept means that English listeners' differential proportion of fixations in the condition without F0 rise was below 0. The significant negative t value for the linear time coefficient indicates that English listeners' differential proportion of fixations in the condition without F0 rise had a descending slope. Importantly, the significant positive t values for the interaction between F0 and the linear and cubic time coefficients mean that English listeners' differential fixation line had more of an ascending slope and more of a reversed 's' shape in the condition with an F0 rise than in the condition without an F0 rise.

These results are illustrated in the modeled differential proportions of fixations in **Figure 3**: From the target-word onset until 300 ms, English listeners' differential proportions of fixations in the two F0 conditions were divergent, with lower fixations in the condition with an F0 rise than in the condition without an F0 rise; at approximately 300 ms *post* target-word onset, the two differential fixation lines converged, and they diverged again shortly after 1,000 ms, with English listeners showing higher differential proportions of fixations in the condition with an F0 rise than in the condition without an F0 rise. The early difference between the two F0 conditions cannot be

TABLE 3 | Growth curve analyses on the difference between listeners' target and competitor fixations in the within-AP condition separately for French, English, and Korean listeners.

Group	Variable	Estimate (SE)	t
French	(Intercept)	−0.075	−1.891
	Time		
	Linear	1.102	2.426
	Quadratic	2.559	7.888 ***
	Cubic	0.551	2.056
	F0	0.075	12.289 ***
	Time × F0		
	Linear	0.538	6.407 ***
	Quadratic	−0.714	−8.515 ***
	Cubic	−0.694	−8.267 ***
English	(Intercept)	−0.152	−3.574 **
	Time		
	Linear	−1.619	−3.695 **
	Quadratic	0.795	2.538
	Cubic	−0.434	−1.688
	F0	0.016	2.144
	Time × F0		
	Linear	1.285	12.501 ***
	Quadratic	0.162	1.587
	Cubic	0.899	8.760 ***
Korean	(Intercept)	−0.151	−5.682 ***
	Time		
	Linear	−1.033	−2.400
	Quadratic	0.772	2.366
	Cubic	0.367	1.493
	F0	−0.046	−6.084 ***
	Time × F0		
	Linear	0.275	2.672 *
	Quadratic	1.199	11.649 ***
	Cubic	0.207	2.009

$\alpha = 0.0167$; * = $p < 0.0167$; ** = $p < 0.003$; *** = $p < 0.0003$. French model: $n = 25$; 9,188 observations. English model: $n = 16$; 5,887 observations. Korean model: $n = 16$; 6,013 observations.

attributed to English listeners' processing of the speech signal, in that it is present from the very beginning of the target word. The late divergence in the expected direction, however, confirms that English listeners could eventually incorporate the use of F0 cues to word-final boundaries in the segmentation of French speech (unlike the results of Tremblay et al., 2012).

Korean Listeners

Last but not least, for the GCA on Korean listeners' data, the significant negative t value for the intercept means that Korean listeners' differential proportion of fixations in the condition

without an F0 rise was below 0. The significant negative t value for F0 indicates that Korean listeners had a *lower* differential proportion of fixations in the condition with an F0 rise than in the condition without an F0 rise. The effect of F0 is thus in the wrong direction. The positive t values for the interactions between F0 and the linear and quadratic time coefficients mean that Korean listeners' differential fixation line was more ascending and more convex in the condition with an F0 rise than in the condition without an F0 rise.

These results can be observed in the modeled differential proportions of fixations in **Figure 3**: Korean listeners showed an effect of F0 in the wrong direction, showing a lower differential proportion of fixation (and thus more competition) in the condition with an F0 rise than in the condition without an F0 rise. This effect of F0 lasted from approximately 200 to 1,250 ms, and reversed thereafter. It is thus possible that Korean listeners were eventually able to use F0 cues to word-final boundaries in the within-AP condition of this experiment. What is clear from these results, however, is that at best they showed great difficulty in using this F0 rise.

L2 Listeners

Again, to ascertain whether Korean listeners differed from English listeners in their differential proportions of fixations, an additional GCA was run only on the L2 learners' data in the within-AP condition, with the English group as baseline. The model with the best fit included all simple effects and all interactions. The results of this GCA and the interpretation of the GCA coefficients can be found in Appendix B in the supplemental data (Table B4). In brief, this GCA revealed a significant three-way interactions between L1, F0, and the linear, quadratic, and cubic time coefficients, indicating that English and Korean listeners differed from each other in the effect of F0 as a function of time: English listeners' differential fixation lines for the two F0 conditions had similar quadratic shapes and ultimately diverged in favor of the condition with an F0 rise; by contrast, Korean listeners' differential fixation lines differed in their quadratic shape and diverged earlier in favor of the condition without an F0 rise (cf. **Table 3**).

DISCUSSION

The present study investigated whether the learning of prosodic cues to word boundaries in speech segmentation is more difficult if the L1 and L2 have similar (though non-identical) prosodies than if they have markedly different prosodies. It did so by focusing on French, English, and Korean listeners' use of F0 rise as a cue to word-final boundaries in French. French and Korean pattern similarly in that word-final boundaries in AP-final position are cued by an F0 rise; yet, they differ in that the AP-final F0 peak is aligned differently in the two languages (earlier in Korean, later in French). English differs from both French and Korean in that F0 rise signals word-initial rather than word-final boundaries. Similarity between the L1 and L2 prosodic systems was hypothesized to make the learning of F0 cues to word-final boundaries difficult. Hence, it was

predicted that Korean L2 learners of French would have more difficulty in using F0 cues to word-final boundaries in French than *both* native French listeners and English L2 learners of French.

The results of the eye-tracking experiment showed that F0 cues modulated native French listeners' differential proportions of fixations (i.e., the difference between their proportion of target fixations and their proportion of competitor fixations), with the flattening of the F0 rise resulting in a fixation line that is *lower*, more ascending, more convex, and more 's'-shaped than in the condition where the F0 rise was naturally present (across-AP), and with the addition of an F0 rise resulting in a fixation line that is *higher* and less convex (though also more ascending and 's'-shaped) than in the condition where the F0 was naturally flat (within-AP). The different directionality of the F0 effect in the across-AP and within-AP conditions and the interaction between these effects and the time coefficients provide evidence that native French listeners used the F0 rise to locate word-final boundaries in continuous French speech, and they add to the existing literature showing that prosodic cues to word-final boundaries constrain lexical access in native French listeners (e.g., Christophe et al., 2004; Michelas and D'Imperio, 2010; Tremblay et al., 2012).

The eye-tracking results also revealed that English L2 learners of French showed evidence of *ultimately* integrating F0 cues in the word recognition process, with the flattening of the F0 rise resulting in a fixation line that was *lower*, more descending, and more concave than in the condition where the F0 rise was naturally present (across-AP). Unlike native French listeners, native English listeners did not show an overall effect of F0 in their fixations in the within-AP condition; however, F0 cues modulated the shape of their differential fixations, with the addition of an F0 rise resulting in a fixation line that was more ascending, and more reversed-'s'-shaped than in the condition where the F0 was naturally flat (within-AP), and with fixations ultimately being numerically higher in the condition with an F0 rise than in the condition without an F0 rise. These results are novel, in that they suggest that English L2 learners of French can, in fact, use F0 rise as a cue to word-final boundaries in French; as far as we know, this study is the first to report such findings. Since the English listeners in this study were somewhat less proficient than the high-proficiency English listeners in Tremblay et al. (2012) (who did not show any effect of F0), the divergent findings between the studies are likely due to the different methodologies employed in the two studies, with eye tracking providing a precise window into the time course of lexical processing and thus capturing the use of cues that may otherwise have a weaker effect in a word-monitoring task.

Finally, the results of the eye-tracking experiment showed that Korean L2 learners of French either did not use the F0 cues in the speech signal (across-AP condition) or did so but in the wrong direction (within-AP condition): The flattening of the F0 rise resulted in a fixation line that was *lower* than in the condition where the F0 rise was naturally present (across-AP), but this effect of F0 did not change as a function of time, and as such,

cannot be attributed to the speech signal;⁸ and the addition of an F0 rise resulted in a fixation line that was *lower* (not higher), more ascending, and more convex than in the condition where the F0 was naturally flat (within-AP). Since the directionality of the F0 effect numerically reverses toward the end of the word recognition process, it is possible that Korean L2 learners of French eventually become able to integrate F0 cues in a target-like manner in their speech segmentation. Overall, however, their pattern of results suggests that they experience great difficulty using F0 rise as a cue to word-final boundaries in French, a finding that is also novel.

These results suggest that the similarity between the L1–L2 prosodic systems in the use of F0 cues makes the learning of L2 segmentation cues difficult for L2 learners, in line with the proposed Prosodic-Learning Interference Hypothesis.⁹ We suspect that Korean listeners' difficulty using the F0 rise as a cue to word-final boundaries in French stems from the different alignments of the AP-final F0 rise in French and in Korean. When hearing an F0 rise in French, Korean L2 learners of French must adjust the timing with which they anticipate a phrase-final (thus, word-final) boundary. If Korean listeners parse French the way they parse Korean, they might wait until the F0 begins lowering to anticipate a word-final boundary; at that point in time, it will already be too late for them to make use of this F0 information in French, as the next word will have already begun. This may explain why Korean listeners had difficulty using the F0 rise in French. If anything, the results in the within-AP condition suggested that Korean listeners initially interpreted this F0 rise as signaling a *word-initial* boundary in French. The late alignment of the F0 rise may thus have been perceived by Korean listeners as being located on the first syllable of the adjective following the monosyllabic word (e.g., *chat lé-*), thus resulting in more lexical competition from the disyllabic word (e.g., *chalet*) in the presence of an F0 rise than in the absence of such a rise.

We believe that the prosodic similarity between French and Korean poses a learnability problem for Korean L2 learners of French and in turn results in speech segmentation difficulties. From the effect of L1 on processing alone, English L2 learners of French should have *more* difficulty in using F0 rise to locate word-final boundaries in French than Korean L2 learners of French, as F0 rise signals *word-initial* rather than phrase-final boundaries in

⁸It is unclear why this effect of F0 emerges since it cannot be attributed to the speech signal.

⁹One concern that might be raised with this study is the fact that the target and competitor words in the eye-tracking experiment were presented orthographically, and Korean uses a different orthographic system from French and English, thus possibly making the task more difficult for Korean L2 learners of French. This concern is unlikely to explain the present results, however. First, Korean speakers learn the roman alphabet from the age of 6 when they begin learning English. Second, the participants had 4,000 ms to preview the orthographic words before they heard the target word. This should have given them plenty of time to orthographically decode the words on the screen. Third, the dependent variable in the statistical analysis was the difference between the proportions of target fixations and the proportions of competitor fixations. This difference closely reflects the amount of lexical activation of both the target and competitor words, and thus factors out overall speed differences due to the decoding of the orthographic words (i.e., such speed differences would affect fixations to both target and competitor words, not one over the other). The use of orthography in this experiment is thus very unlikely to explain why Korean listeners had more difficulty than English listeners in using the F0 rise as a cue to word-final boundaries in French.

English. Yet, English L2 learners of French were ultimately able to integrate F0 cues to word-final boundaries in a target-like manner to segment French speech, both in non-AP-final (within-AP condition) and AP-final (across-AP condition) positions. Since our L2 groups were matched in both their French proficiency and French experience, the observed difference between the two L2 groups suggest that the prosodic similarity between French and Korean may pose a learnability problem for Korean listeners, consistent with the Prosodic-Learning Interference Hypothesis.

Similarity between L1-L2 prosodic systems may hurt L2 learning for two reasons: L2 listeners may perceptually assimilate L2 prosodic cues to L1 prosodic cues and/or they may not experience parsing failure as a result of not using L2 prosodic cues. First, Korean L2 learners of French may perceive the F0 rise in French as similar to that in Korean. As a result, they may not readjust their use of segmentation cues. Similar perceptual assimilations have been reported for the perception of segments (for a discussion of PAM-L2 and SLM, see Best and Tyler, 2007 and Flege, 1995, respectively). However, since F0 cues are unlikely to be perceived categorically (at least in French and Korean), the exact process underlying the assimilation of F0 cues in the L1 and L2 would likely be different from that postulated for L1 and L2 segments. Second, L2 learners may not readjust their use of segmentation cues if these unadjusted cues do not cause parsing failure (i.e., if they do not result in the greater activation of L2 competitor words over L2 target words; for such a proposal, see Carroll, 2004). The present results do not adjudicate between these two types of mechanisms, but they pave the way for further research to try to tease them apart.

The main contribution of this study is in demonstrating that learning to segment speech in the L2 is difficult if a particular prosodic cue signals the same word boundary in the L1 and L2 but does so differently. We have provided evidence that Korean L2 learners of French, unlike English L2 learners of French matched to them in French proficiency and French experience, have great difficulty learning to use F0 rise to locate word-final boundaries in French, a result which we hypothesize is due to the different alignments of the AP-final F0 peak in Korean (earlier) and French (later). To the best of our knowledge, this study is the first to report that similarity between the L1 and L2 can hurt L2 learning in the domain of sentential prosody. Further research should examine how differences in this AP-final F0-peak alignment impact Korean listeners' segmentation of French speech. Our findings also raise the question of whether L2 learning is similarly impacted by subtle prosodic differences that manifest themselves differently (e.g., cues that have different alignments vs. different strengths) or that are used to express a categorical distinction in one language but not the other. Answering these questions would make an important contribution to the understanding of how non-native listeners become (or do not become) able to segment speech successfully in an L2.

The findings of this study also raise questions about the mechanisms underlying L2 learners' encoding of prosodic cues. Prosodic information in French is, at least to a large degree, independent from segmental information: The same words can be realized very differently depending on their position in the AP. This makes it unlikely that native French listeners would encode the prosody of each exemplar French word they hear in their lexical representations. Computing the prosody of the utterance independently of its segmental content, with listeners aligning words with the prosodic constituents of the utterance, may be a more efficient strategy. Korean L2 learners of French, for whom prominence is also phrasal, may also compute the prosody of the French utterance somewhat (but perhaps not completely) independently of its segmental content (since the tonal pattern of the AP in Korean is partly influenced by the AP-initial segment), but with the alignment of the prosodic constituents (signaled by prosodic cues such as F0 rise) being slightly off and thus resulting in speech segmentation difficulties. In contrast, English L2 learners of French, for whom prominence is both lexical and phrasal, may begin the learning of French words by encoding a great deal of prosodic details for each exemplar, and only later become less reliant on such lexical encoding. Importantly, even if English listeners were to adopt a different strategy from French listeners at their onset of learning French, ultimately they showed the right pattern of fixations when segmenting French, albeit late in the word recognition process. Further research should shed light on the precise mechanisms that underlie listeners' encoding of prosodic information, whether this encoding varies across languages, and if so, how these differences affect the L2 learning of prosody.

AUTHOR CONTRIBUTIONS

AT: Stimuli creation, experimental design, data collection, data analysis, writing of manuscript. MB: Data analysis, writing of manuscript. CC: Stimuli creation, experimental design, data collection, data analysis, writing of manuscript. JC: Data collection, writing of manuscript.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grant no. BCS-1423905 awarded to the first author. Support for this research also comes from a *Language Learning* small research grant awarded to the first author.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00985>

REFERENCES

- Al-jasser, F. (2008). The effect of teaching english phonotactics on the lexical segmentation of english as a foreign language. *System* 36, 94–106. doi: 10.1016/j.system.2007.12.002
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge, MA: Cambridge University Press.
- Bagou, O., Fougeron, C., and Frauenfelder, U. H. (2002). "Contribution of prosody to the segmentation and storage of 'words' in the acquisition of a new mini-language," in *Proceedings of Speech Prosody 2002 Conference*, eds B. Bel and I. Marlien (Aix-en-Provence: Laboratoire Parole et Langage), 59–62.
- Bagou, O., and Frauenfelder, U. H. (2006). "Stratégie de segmentation prosodique: rôle des prééminences initiales et finales dans l'acquisition d'une langue artificielle," in *Proceedings of the XXVIèmes Journées d'Étude sur la Parole*, Dinart, 571–574.
- Banel, M.-H., and Bacri, N. (1994). On metrical patterns and lexical parsing in French. *Speech Commun.* 15, 115–126. doi: 10.1016/0167-6393(94)90046-9
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). *lme4: Linear Mixed-Effects Models Using 'Eigen' and S4. R package version 1.1-10*. Available at: <http://CRAN.R-project.org/package=lme4>
- Beckman, M. E. (1986). *Stress and Non-Stress Accent*. Dordrecht: Foris.
- Beckman, M. E., and Elam, G. A. (1997). *Guidelines for ToBI Labeling (Version 3.0)*. Available at: http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/
- Beckman, M. E., and Jun, S.-A. (1996). *K-ToBI (Korean ToBI) Labeling Conventions, Manuscript*. Ohio State University & UCLA. Available at: <http://www.linguistics.ucla.edu/people/jun/sunah.htm>
- Best, C. T. (1995). "A direct realist perspective on cross-language speech perception," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, ed. W. Strange (Timonium, MD: York Press), 167–200.
- Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: commonalities and complementarities," in *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production*, eds M. J. Munro and O.-S. Bohn (Amsterdam: John Benjamins), 13–34.
- Boersma, P., and Weenink, D. (2004). *Praat: Doing Phonetics by Computer (Version 5.2.11) [Computer program]*. Available at: www.praat.org
- Carroll, S. E. (2004). Segmentation: learning how to 'hear words' in the L2 speech stream. *Trans. Philol. Soc.* 102, 227–254. doi: 10.1111/j.0079-1636.2004.00136.x
- Cho, T., and Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *J. Phon.* 29, 155–190. doi: 10.1006/jpho.2001.0131
- Christophe, A., Peperkamp, S., Pallier, C., Block, E., and Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *J. Mem. Lang.* 51, 523–547. doi: 10.1016/j.jml.2004.07.001
- Chung, K., Chang, S., Choi, J., Nam, S., Lee, M., Chung, S., et al. (1996). *A Study of Korean Prosody and Discourse for the Development of Speech Synthesis/Recognition System*. Daejeon: KAIST Artificial Intelligence Research Center.
- Clopper, C. (2002). *Frequency of Stress Patterns in English: A Computational Analysis*. IULC Working Papers Online, 2. Available at: <https://www.indiana.edu/~iulcw/pdfs/02-clopper02.pdf>
- Cooper, N., Cutler, A., and Wales, R. (2002). Constraints of lexical stress on lexical access in english: evidence from native and non-native listeners. *Lang. Speech* 45, 207–228. doi: 10.1177/00238309020450030101
- Cutler, A. (1986). Forbear is a homophone: lexical prosody does not constrain lexical access. *Lang. Speech* 29, 201–220.
- Cutler, A. (1997). The syllable's role in the segmentation of stress languages. *Lang. Cogn. Process.* 12, 839–845. doi: 10.1080/016909697386718
- Cutler, A., and Butterfield, S. (1992). Rhythmic cues to speech segmentation: evidence from juncture misperception. *J. Mem. Lang.* 31, 218–236. doi: 10.1016/0749-596X(92)90012-M
- Cutler, A., and Carter, D. M. (1987). The predominance of strong initial syllables in the english vocabulary. *Comput. Speech Lang.* 2, 133–142. doi: 10.1016/0885-2308(87)90004-0
- Cutler, A., and Clifton, C. E. Jr. (1984). "The use of prosodic information in word recognition," in *Attention and Performance X: Control of Language Processes*, eds H. Bouma and D. G. Bouwhuis (Hillsdale, NJ: Erlbaum), 183–196.
- Cutler, A., Mehler, J., Norris, D., and Segui, J. (1983). A language specific comprehension strategy. *Nature* 304, 159–160. doi: 10.1038/304159a0
- Cutler, A., Mehler, J., Norris, D., and Segui, J. (1986). The syllable's differing role in the segmentation of french and english. *J. Mem. Lang.* 25, 385–400. doi: 10.1016/0749-596X(86)90033-1
- Cutler, A., Mehler, J., Norris, D., and Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cogn. Psychol.* 24, 381–410. doi: 10.1016/0010-0285(92)90012-Q
- Cutler, A., and Otake, T. (1994). Mora or phoneme? evidence for language-specific listening. *J. Mem. Lang.* 33, 824–844. doi: 10.1006/jmla.1994.1039
- de Pijper, J. R. (1983). *Modeling British English intonation*. Dordrecht: Foris.
- Fear, B. D., Cutler, A., and Butterfield, S. (1995). The strong/weak syllable distinction in english. *J. Acoust. Soc. Am.* 97, 1893–1904. doi: 10.1121/1.412063
- Flege, J. E. (1995). "Second-language speech learning: theory, findings, and problems," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, ed. W. Strange (Timonium, MD: York Press), 233–272.
- Grabe, E., and Low, E. L. (2002). "Durational variability in speech and the rhythm class hypothesis," in *Papers in Laboratory Phonology Vol. 7*, eds N. Warner and C. Gussenhoven (Berlin: Mouton de Gruyter), 515–546.
- Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago, IL: The University of Chicago Press.
- Huetting, F., and McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *J. Mem. Lang.* 57, 460–482. doi: 10.1016/j.jml.2007.02.001
- Jun, S.-A. (1993). *The Phonetics and Phonology of Korean Prosody*. Ph.D. dissertation, The Ohio State University, Columbus.
- Jun, S.-A. (1995). "Asymmetrical prosodic effects on the laryngeal gesture in Korean," in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, eds B. Connell and A. Arvaniti (Cambridge, MA: Cambridge University Press), 235–253.
- Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15, 189–226. doi: 10.1017/S0952675798003571
- Jun, S.-A. (2000). *K-ToBI (Korean ToBI) Labelling Conventions (version 3.1)*. Available at: <http://www.linguistics.ucla.edu/people/jun/ktobi/K-tobi.html>
- Jun, S.-A., and Fougeron, C. (2000). "A phonological model of French intonation," in *Intonation: Analysis, Modelling, and Technology*, ed. A. Botinis (Dordrecht: Kluwer Academic Publishers), 209–242.
- Jun, S.-A., and Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus* 14, 147–172. doi: 10.1515/prbs.2002.002
- Kim, J., Davis, C., and Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Lang. Speech* 51, 343–359. doi: 10.1177/0023830908099069
- Kim, S., Broersma, M., and Cho, T. (2012). The use of prosodic cues in learning new words in an unfamiliar language. *Stud. Second Lang. Acquis.* 34, 415–444. doi: 10.1017/S0272263112000137
- Kim, S., and Cho, T. (2009). The use of phrase-level prosodic information in lexical segmentation: evidence from word-spotting experiments in Korean. *J. Acoust. Soc. Am.* 125, 3373–3386. doi: 10.1121/1.3097777
- Lee, H., and Jongman, A. (2015). Acoustic evidence for diachronic sound change in Korean prosody: a comparative study of the Seoul and South Kyungsang dialects. *J. Phon.* 50, 15–33. doi: 10.1016/j.wocn.2015.01.003
- Lee, H., and Zhang, J. (2014). The nominal pitch accent system of South Kyungsang Korean. *J. East Asian Linguist.* 23, 71–111. doi: 10.1007/s10831-013-9119-x
- Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *J. Acoust. Soc. Am.* 32, 451–454. doi: 10.1121/1.1908095
- Low, E. L., Grabe, E., and Nolan, F. (2000). Quantitative characterisations of speech rhythm: 'Syllable-timing' in Singapore English. *Lang. Speech* 43, 377–401. doi: 10.1177/00238309000430040301
- Mattys, S., White, L., and Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *J. Exp. Psychol. Gen.* 134, 477–500. doi: 10.1037/0096-3445.134.4.477
- Mattys, S. L. (2004). Stress versus coarticulation: toward an integrated approach to explicit speech segmentation. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 397–408.
- McQueen, J. M., Norris, D., and Cutler, A. (1994). Competition in spoken word recognition: spotting words in other words. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 621–638.

- McQueen, J. M., and Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Q. J. Exp. Psychol.* 60, 661–671. doi: 10.1080/17470210601183890
- Michelas, A., and D'Imperio, M. (2010). "Accentual phrase boundaries and lexical access in French," in *Proceedings of the Speech Prosody Conference*, Chicago, IL.
- Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Boca Raton, FL: Taylor & Francis.
- Murty, L., Otake, T., and Cutler, A. (2007). Perceptual tests of rhythmic similarity: I. Mora rhythm. *Lang. Speech* 50, 77–99. doi: 10.1177/00238309070500010401
- Oh, M. (1998). The prosodic analysis of intervocalic tense consonant lengthening in Korean. *Jpn. Korean Linguist.* 8, 317–330.
- Otake, T., Hatano, G., Cutler, A., and Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *J. Mem. Lang.* 32, 258–278. doi: 10.1006/jmla.1993.1014
- Otake, T., Hatano, G., and Yoneyama, K. (1996). "Speech segmentation by Japanese listeners," in *Phonological Structure and Language Processing: Cross-Linguistic Studies*, eds T. Otake and A. Cutler (Berlin: Mouton de Gruyter), 183–201.
- Ramus, F. (2002). "Acoustic correlates of linguistic rhythm: perspectives," in *Proceedings of Speech Prosody 2002*, Aix-en-Provence, 115–120.
- Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265–292. doi: 10.1016/S0010-0277(00)00101-3
- Small, L. H., Simon, S. D., and Goldberg, J. S. (1988). Lexical stress and lexical access: homographs versus nonhomographs. *Percept. Psychophys.* 44, 272–280. doi: 10.3758/BF03206295
- Spinelli, E., Grimaud, N., Meunier, F., and Welby, P. (2010). An intonational cue to segmentation in phonemically identical sequences. *Attent. Percept. Psychophys.* 72, 775–787. doi: 10.3758/APP.72.3.775
- Spinelli, E., Welby, P., and Schaeffgen, A.-L. (2007). Fine-grained access to targets and competitors in phonemically identical spoken sequences: the case of French elision. *Lang. Cogn. Process.* 22, 828–859. doi: 10.1080/01690960601076472
- Tremblay, A. (2011). Proficiency assessment in second language acquisition research: "Clozing" the gap. *Stud. Second Lang. Acquis.* 33, 339–372. doi: 10.1017/S0272263111000015
- Tremblay, A., Coughlin, C., Bahler, C., and Gaillard, S. (2012). Differential contribution of prosodic cues in the native and non-native segmentation of French speech. *Lab. Phonol.* 3, 385–423. doi: 10.1515/lp-2012-0018
- Tremblay, A., and Owens, N. (2010). The role of acoustic cues in the development of (non-)target-like second-language prosodic representations. *Can. J. Linguist.* 55, 85–114. doi: 10.1353/cjl.0.0067
- Tremblay, A., and Spinelli, E. (2014). English listeners' use of distributional and acoustic-phonetic cues to liaison in French: evidence from eye movements. *Lang. Speech* 57, 310–337. doi: 10.1177/0023830913504569
- Tyler, M. D., and Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *J. Acoust. Soc. Am.* 126, 367–376. doi: 10.1121/1.3129127
- Vaissière, J. (1983). "Language-independent prosodic features," in *Prosody: Models and Measurements*, eds A. Cutler and R. D. Ladd (Berlin: Springer), 53–66.
- Weber, A., and Cutler, A. (2006). First-language phonotactics in second-language listening. *J. Acoust. Soc. Am.* 119, 597–607. doi: 10.1121/1.2141003
- Welby, P. (2006). French intonational structure: evidence from tonal alignment. *J. Phon.* 34, 343–371. doi: 10.1016/j.wocn.2005.09.001
- Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Commun.* 49, 28–48. doi: 10.1016/j.specom.2006.10.005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Tremblay, Broersma, Coughlin and Choi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Early Prosodic Acquisition in Bilingual Infants: The Case of the Perceptual Trochaic Bias

Ranka Bijeljac-Babic^{1,2*}, Barbara Höhle³ and Thierry Nazzi^{1*}

¹ Laboratoire Psychologie de la Perception, Université Paris Descartes – CNRS, Paris, France, ² Université de Poitiers, Poitiers, France, ³ Universität Potsdam, Potsdam, Germany

OPEN ACCESS

Edited by:

Isabelle Darcy,
Indiana University Bloomington, USA

Reviewed by:

LouAnn Gerken,
University of Arizona, USA
Mariapaola D'Imperio,
Aix-Marseille Université et Laboratoire
Parole et Langage, France

*Correspondence:

Ranka Bijeljac-Babic
ranka.bijeljac-
babic@parisdescartes.fr;
Thierry Nazzi
thierry.nazzi@parisdescartes.fr

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 18 November 2015

Accepted: 03 February 2016

Published: 23 February 2016

Citation:

Bijeljac-Babic R, Höhle B and Nazzi T
(2016) Early Prosodic Acquisition
in Bilingual Infants: The Case of the
Perceptual Trochaic Bias.
Front. Psychol. 7:210.
doi: 10.3389/fpsyg.2016.00210

Infants start learning the prosodic properties of their native language before 12 months, as shown by the emergence of a trochaic bias in English-learning infants between 6 and 9 months (Juszyk et al., 1993), and in German-learning infants between 4 and 6 months (Höhle et al., 2009, 2014), while French-learning infants do not show a bias at 6 months (Höhle et al., 2009). This language-specific emergence of a trochaic bias is supported by the fact that English and German are languages with trochaic predominance in their lexicons, while French is a language with phrase-final lengthening but lacking lexical stress. We explored the emergence of a trochaic bias in bilingual French/German infants, to study whether the developmental trajectory would be similar to monolingual infants and whether amount of relative exposure to the two languages has an impact on the emergence of the bias. Accordingly, we replicated Höhle et al. (2009) with 24 bilingual 6-month-olds learning French and German simultaneously. All infants had been exposed to both languages for 30 to 70% of the time from birth. Using the Head Preference Procedure, infants were presented with two lists of stimuli, one made up of several occurrences of the pseudoword /GAba/ with word-initial stress (trochaic pattern), the second one made up of several occurrences of the pseudoword /gaBA/ with word-final stress (iambic pattern). The stimuli were recorded by a native German female speaker. Results revealed that these French/German bilingual 6-month-olds have a trochaic bias (as evidenced by a preference to listen to the trochaic pattern). Hence, their listening preference is comparable to that of monolingual German-learning 6-month-olds, but differs from that of monolingual French-learning 6-month-olds who did not show any preference (Höhle et al., 2009). Moreover, the size of the trochaic bias in the bilingual infants was not correlated with their amount of exposure to German. The present results thus establish that the development of a trochaic bias in simultaneous bilinguals is not delayed compared to monolingual German-learning infants (Höhle et al., 2009) and is rather independent of the amount of exposure to German relative to French.

Keywords: bilinguals, infants, language, prosody, lexical stress, dominance effects

INTRODUCTION

The majority of children around the world grow up in bilingual families or countries, acquiring more than one language at a time (Grosjean, 2010). Despite being exposed to a more complex language situation, bilingual children succeed in the task of simultaneously learning their two native languages and pass the language development milestones

at roughly the same ages as their monolingual peers (Byers-Heinlein et al., 2010). However, this does not mean that language acquisition proceeds in exactly the same way in mono- and bilingual children. When discriminating phonetic contrasts present in only one of their languages, bilinguals usually succeed as monolinguals (Burns et al., 2007; Sundara et al., 2008; Albareda-Castellot et al., 2011; Sundara and Scutellaro, 2011), although a U-shaped curve not observed in monolinguals has been found for bilingual infants in some studies (Bosch and Sebastián-Gallés, 2003; Sebastián-Gallés and Bosch, 2009). In some language-related tasks, bilingual infants have shown an advantage over monolinguals at both 10 months (García-Sierra et al., 2011; Bijeljac-Babic et al., 2012) and even 7 months of age (Kovács and Mehler, 2009). These data suggest that hearing two different languages provides bilingual infants with greater experience in processing a more variable input and develop cognitive flexibility in both linguistic and non-linguistic tasks.

While previous studies focused on early development of segmental phonology in this population, very little is known regarding how prosodic properties are processed and acquired by very young bilingual infants. The present study therefore explored how simultaneous bilingual infants acquire fundamental prosodic properties of their two native languages when these languages differ in the realization of lexical stress. Prior to presenting this work, we review what we know about early prosodic acquisition in bilingual infants compared to monolingual infants. At present, the bulk of the available data relates to early language discrimination or recognition, and processing of stress patterns.

Monolingual infants have been found to recognize their native language at birth (Mehler et al., 1988; Moon et al., 1993). Moreover, both newborn and 2-month-old monolinguals can discriminate languages only if they differ by their rhythmic properties (Christophe and Morton, 1998; Nazzi et al., 1998; Byers-Heinlein et al., 2010), while 3.5- to 5-month-old monolinguals can discriminate their native language from rhythmically similar ones (Nazzi et al., 2000; Butler et al., 2011; Molnar et al., 2013). Similarly, bilingual newborns prefer to listen to both of their languages over a rhythmically different one, and can discriminate them if they have different rhythms (Byers-Heinlein et al., 2010). By 3–5 months of age, they can discriminate their two native languages if they are rhythmically similar (Bosch and Sebastián-Gallés, 1997; Molnar et al., 2013). While some fine-grained differences in shift latencies and overall listening times (Bosch and Sebastián-Gallés, 1997) suggest that bilinguals might attend to their native languages differently than monolinguals, bilinguals appear to have highly similar language discrimination and recognition abilities as monolinguals, probably based on the processing of prosodic properties at the utterance level.

Regarding the processing of more local prosodic properties, several studies have compared stress pattern discrimination in French- vs. Spanish- (Skoruppa et al., 2009; Abboub et al., 2015) or German-learning (Friederici et al., 2007; Höhle et al., 2009; Bijeljac-Babic et al., 2012) monolingual infants. While lexical stress is found in most languages of the world, including Spanish, German, English, and Dutch, French does not use stress contrasts at the lexical level. However, French has fixed phrasal-final stress

which is mostly marked by a lengthening of the phrase-final syllable (Hayes, 1995; Di Cristo, 1999; Jun and Fougeron, 2000, 2002). Moreover, stress is not realized acoustically in the same way in all languages: for example, duration appears to have a more important role in prosodic phrasing in French than in German (Féry et al., 2011); and both F0 and intensity are reliably higher for a syllable that has phrasal stress in German, while they can be dissociated in French (Vaissière and Michaud, 2006; Nespor et al., 2008). The fact that French only has phrase-final stress appears to have an impact on stress pattern discrimination by French-speaking adults: compared to both Spanish and German adults, French speakers show a reduced sensitivity to stress (that has sometimes been called “stress deafness”) when processing either words presented in isolation (Dupoux et al., 1997, 2001) or continuous sequences made up of nonsense syllables or nonlinguistic sounds (Bhatara et al., 2013, 2015; Boll-Avetisyan et al., 2015). This reduced sensitivity in French adults (which does not prevent them to use phrase boundaries as cues for segmentation, Michelas and D’Imperio, 2015) is particularly marked when the stimuli presented are characterized by speaker or segmental variability, suggesting that crosslinguistic differences emerge in experimental contexts in which the stimuli need to be processed at a phonological level, rather than at an acoustic or phonetic level.

When are these crosslinguistic differences set into place? Are infants growing up with different linguistic backgrounds differentially sensitive to stress patterns depending on whether they can process that information at the phonetic level (when presented with stimuli lacking segmental variability), or whether they have to process it at the phonological level (when presented with stimuli containing segmental variability)? Accordingly, previous studies tested infants in two different conditions. In the no segmental variability condition, infants were presented with different tokens of a single item (e.g., /gaba/) recorded either with a stress-initial (trochaic) or stress-final (iambic) pattern, allowing for discrimination based on lower-level acoustic properties. In the more challenging (segmental variability) condition, infants were presented with lists of segmentally different items (e.g., /datu/, /sapi/, /kiba/, /nuki/.) recorded with either a trochaic or an iambic pattern, such that discrimination was only possible if infants could abstract and generalize the stress patterns over segmental variability.

For monolingual infants, and in the absence of segmental variability, discrimination was found in Italian newborns (Sansavini et al., 1997), English-learning 1- to 4-month-olds (Spring and Dale, 1977), German-learning 4-month-olds (electrophysiological data: Friederici et al., 2007; behavioral data: Herold et al., 2008), and Spanish-learning 6-month-olds (Skoruppa et al., 2013). Importantly, it was also found in French-learning infants from 4 to 10 months of age (Friederici et al., 2007; Höhle et al., 2009; Skoruppa et al., 2009, 2013). This establishes early stress discrimination abilities in the absence of segmental variability that appears to be independent of native language experience. However, French-learning infants’ sensitivity to lexical stress declines between 6 and 10 months of age. While at 6 months, they could discriminate stress patterns following a 1-min familiarization with one pattern, at 10 months

they needed 2 rather than 1 min of familiarization (Höhle et al., 2009; Skoruppa et al., 2009; Bijeljac-Babic et al., 2012). Thus, at 10 months, French-learning infants require more time to identify stress patterns, a developmental path reflecting early language-specific reorganization probably leading to the “stress deafness” found in French adults (Dupoux et al., 1997, 2001; Bhatara et al., 2013, 2015; Boll-Avetisyan et al., 2015).

Still for monolingual infants, what do we know about stress pattern discrimination in the presence of segmental variability? Such ability seems limited in young infants. Indeed, newborns were found to discriminate stress patterns when presented with lists of words varying on consonants (Sansavini et al., 1997) but not when the words varied on both consonants and vowels (Sansavini et al., 1994). Early limitations were further attested by Spanish- and French-learning 6-month-olds’ difficulty at discriminating stress patterns when presented with lists of segmentally different words (Skoruppa et al., 2013). Discrimination of stress patterns across segmentally varying words were found later in development, at 9 months in Spanish-learning infants, and at 8 and 12 months in English-learning infants (Skoruppa et al., 2009, 2011). Importantly, such discrimination abilities appear to be modulated by the native language: French-learning 9 to 10-month-olds failed to discriminate, thereby showing that they cannot process stress patterns across multiple segmentally varied items, whereas they can do so in tasks using only one item (Skoruppa et al., 2009; Abboub et al., 2015).

As for bilingual infants, only two studies explored stress pattern discrimination in either the absence (Bijeljac-Babic et al., 2012) or the presence (Abboub et al., 2015) of segmental variation. Given the earlier studies on monolingual infants showing that sensitivity to lexical stress changes during development as a function of the prosodic characteristics of the native language, these studies explored prosodic acquisition in bilingual infants learning two languages with different lexical stress pattern systems. In both studies, 10-month-old bilingual infants learning French (a language lacking lexical stress contrasts) and a language that has variable lexical stress (from a set of about 15 different languages) were found to discriminate stress contrasts, and to perform better than French-learning monolinguals of the same age. These findings thus establish that the presence of a second language with variable lexical stress maintains sensitivity to lexical stress in these bilingual infants learning French. This stress pattern discrimination in bilinguals was found in the absence as well as in the presence of segmental variability in the stimuli. Moreover, in these two studies, none of the bilingual infants were learning the language in which the stimuli had been produced (German for Bijeljac-Babic et al., 2012; Spanish for Abboub et al., 2015), demonstrating that these discrimination effects are not just based on the recognition of the exact acoustic cues used in their second language to mark lexical stress, but that they possibly reflect the sensitivity to abstract stress patterns.

These two studies also explored the effect of language dominance, in order to determine whether infants who receive less French input and more of the languages with lexical stress have better stress discrimination abilities. In Bijeljac-Babic

et al. (2012), stress pattern discrimination was significant in the subgroup of infants dominant in the languages with lexical stress (receiving 70–80% of their input in these languages) but was only marginal in the balanced bilinguals (receiving 40–60% of their input in both languages), suggesting an effect of language dominance. However, in Abboub et al. (2015), no effect of language dominance was found: first, discrimination performance did not differ for French-dominant (hearing French 60 to 70% of the time in their input) versus not French-dominant infants (hearing French only 30 to 50% of the time), and second, there was no correlation between performance and percentage of German heard. Taken together, the two studies reveal only a weak impact of language dominance on prosodic processing, at least for discrimination abilities, suggesting that already a limited amount of exposure to a language with lexical stress allows the maintenance of discrimination abilities. Would the same hold for the acquisition of a prosodic property (namely the predominant stress pattern of words in the native language) that requires not only to discriminate stress patterns but also to conduct distributional analyses on the relative frequency of each pattern within the input?

In monolinguals, early language-specific developmental changes have been found. This was revealed by the emergence of a preference for trochaic over iambic items lacking segmental variability in German-learning infants between 4 and 6 months of age, the trochaic pattern being predominant in German (and English), while such bias was not found in French-learning 6-month-olds (Höhle et al., 2009). Still in monolinguals, but using lists of segmentally varied words, a trochaic bias was found to emerge in English-learning infants between 6 and 9 months of age (Jusczyk et al., 1993). However, a different pattern was found for Spanish (Pons and Bosch, 2010), a language with a relatively balance of trochaic (60%) and iambic (40%) words, but in which stress assignment is related to syllabic structure (95% of CVC.CV words are trochaic; 93% of CV.CVC words are iambic). Presented with lists of segmentally varying words, Spanish-learning 9-month-olds showed no stress pattern preference for CV.CV words, a trochaic preference for CVC.CV words and an iambic preference for CV.CVC words. Taken together, these results show that after 6 months of age, monolingual infants learning a language with variable lexical stress have learned the predominant stress pattern of their native language (and its link to syllabic structure). These findings further suggest that recognizing this pattern becomes more efficient in the following months, allowing infants to abstract the stress pattern from segmentally varying strings. On the other hand, monolingual infants learning French appear not to develop a trochaic bias, as no such bias is present in their linguistic input.

The above acquisition pattern thus raises the question of whether and when bilingual infants learning French and a language with a predominant lexical stress pattern develop a preference for that predominant pattern. The present study explored this issue, extending Höhle et al. (2009) to French/German bilingual infants, in order to determine whether by 6 months of age, these infants have a trochaic bias when listening to German stimuli in the absence of segmental

variability, as has been found for their German- but not their French-learning age mates. The present study also asked the question of whether language dominance modulates this effect, by exploring whether the size of the trochaic preference is related to the relative amount of German heard, as estimated through parental language reports.

EXPERIMENT

Methods

Ethical Statement

This study was authorized by the ethics committee “Comité de Protection des Personnes Ile de France II” (decision 2011 06).

Participants

Twenty-four French/German 6-month-olds ($M = 6;6$; range: 6;00–7;4; 16 girls and 8 boys), were tested in Paris. All participants were born full term, without apparent health problems. They were recruited from birth-lists obtained through the Paris city hall archives and from the “CAFE bilingue,” an association for the promotion of bilingual education. Informed written consent was obtained from all parents. The infants’ relative exposure to their two languages, both within and outside (e.g., extended family, caregivers and friends) the nuclear family, was measured using the Language Exposure Questionnaire (Bosch and Sebastián-Gallés, 1997). Only infants exposed to both French and German between 30 to 70% of the time, and to no other languages, were included in the study. Mean exposure to German was 53.7%. Two additional infants were tested but did not complete the experiment due to fussiness.

Stimuli

The stimuli were those used in Höhle et al. (2009). They consisted of CVCV /gaba/ sequences, stressed either on the first syllable (trochaic pattern) or on the second syllable (iambic pattern). Several tokens of each stress pattern were recorded by a German female native speaker. The first syllables of the trochaic sequences had a mean duration of 283 ms ($SD = 20.8$), the second syllables of the trochaic sequences one of 308 ms ($SD = 25.0$). The analysis of pitch revealed an average of 195 Hz ($SD = 3.9$) on the first and 163 Hz ($SD = 15.9$) on the second syllables. The first syllables of the iambic sequences had an average duration of 173 ms ($SD = 11.0$) whereas the second syllables had a mean duration of 430 ms ($SD = 21.2$). The average pitch of the first syllables was 186 Hz ($SD = 5.2$), that of the second syllables 183 Hz ($SD = 5.9$).

Again following Höhle et al. (2009), the tokens were used to create six files for each stress pattern that differed in the order of presentation of the different tokens, the tokens in a file being separated by pauses of about 600 ms. The trochaic speech files contained 16 tokens and had an average duration of 18.39 s (range: 18.28–18.51 s) and the iambic files contained 15 tokens and had an average duration of 18.01 s (range: 18.00–18.07 s). The difference in number of tokens per file is due to the fact that the iambic bisyllables had a slightly longer average duration (603 ms) than the trochaic bisyllables (591 ms).

Procedure, Apparatus, and Design

We used the Headturn Preference Paradigm (HPP) as introduced by Hirsh-Pasek et al. (1987). The procedure, apparatus and design were the same as for the monolingual French infants in Experiment 3 of Höhle et al. (2009).

The experiment was run by a Dell Optiplex computer. During the experimental session, the infant was seated on the lap of a caregiver in the center of a test booth. The caregiver listened to music over headphones to prevent influences on the infant’s behavior. Furthermore, he or she was instructed not to interfere with the infant’s behavior during the experiment. Inside the booth, three lamps were fixated: a green one at the center wall, and red ones at each of the side walls. Directly above the green lamp on the center wall was a hole for the lens of a video-camera. On the inside of the test booth, two loudspeakers (SONY xs-F1722) were mounted just below the red side lamps. Each experimental trial started by the blinking of the green center lamp. When the infant oriented to the green lamp, this lamp went out and one of the red lamps on a side wall started to blink. When the infant turned her head toward the red lamp, the speech stimulus was presented from the loudspeaker on the same side as the blinking red lamp. The trial ended when the infant turned her head away for more than 2 s, or when the end of the speech file was reached. If the infant turned away for less than 2 s, the presentation of the speech file continued but the time spent looking away was not included in the total listening time.

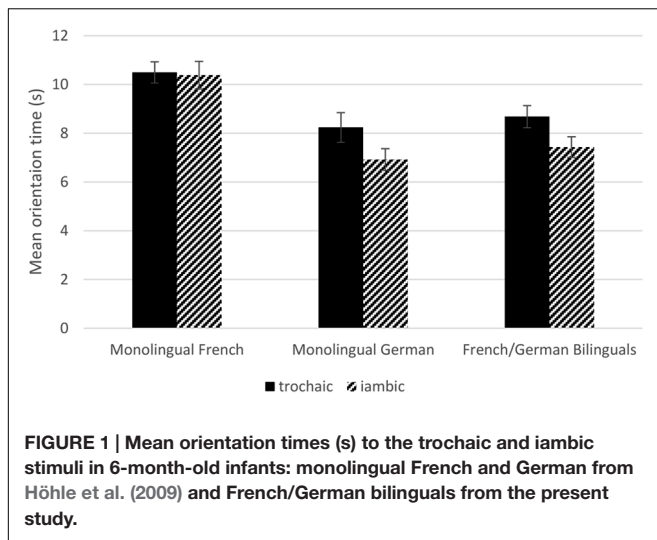
The first two speech files – one of the trochaic and one of the iambic pattern – served as warm-up trials and were not included in the data analysis. The remaining 12 experimental speech files were presented in random order. The duration of each experimental session lasted approximately 3–5 min.

RESULTS

As for the experiments in Höhle et al. (2009), all individual orientation times exceeding 18 s were reduced to 18 s; two trials were cut off, accounting for 0.7% of all trials. Mean orientation times for each of the two rhythmic patterns were calculated for each infant. On average, infants oriented for 8.68 s ($SE = 0.45$) to the trochaic sequences and for 7.43 s ($SE = 0.42$) to the iambic sequences (see **Figure 1**). This difference was significant, $t(23) = 2.43$, $p = 0.02$, two-tailed, Cohen’s $d = 1.01$, large effect. Sixteen of the 24 infants had longer orientation times to the trochaic than to the iambic sequences. In order to evaluate whether the size of the trochaic bias was influenced by infants’ amount of exposure to German, the difference in orientation times for trochaic minus iambic stimuli was correlated with their percentage of German input. No significant correlation was found, $r = -0.12$, $p = 0.58$.

DISCUSSION

The aim of the present study was to evaluate the emergence of a trochaic bias in simultaneous French/other language bilinguals, given prior evidence that such a bias emerges in several stress-based languages (German, English), but not in French (Juszyk



et al., 1993; Höhle et al., 2009). Our study used the exact same procedure and German stimuli as Höhle et al. (2009), which had found the emergence of a trochaic bias in monolingual German-learning infants between 4 and 6 months, but no preference in monolingual French-learning infants at 6 months. In this context, the finding of a trochaic preference in French/German bilingual 6-month-olds establishes for the first time that there is no delay in this area of prosodic acquisition for this bilingual population compared to German-learning monolinguals. Moreover, the fact that performance was not affected by the relative amount of exposure to the two languages suggests that even 30% of exposure to German is enough for bilingual infants to develop a trochaic bias that can be used when processing German stimuli.

In our previous studies, we had found that at 10 months of age, bilingual infants learning French and a language with lexical stress do not show the same decline in their sensitivity to stress contrasts than monolingual French-learners, and that they remain sensitive to such contrasts in contexts either lacking (Bijeljic-Babic et al., 2012) or incorporating (Abboub et al., 2015) segmental variability. This sensitivity to stress information constitutes a prerequisite to be able to process lexical stress information in their speech input, and thus be able to discover the predominant lexical stress pattern of the languages spoken in their environment. Taken together with the current findings, this suggests that bilingual infants' exposure to one language with lexical stress not only maintains sensitivity to this dimension but also provides them with a basis for learning its predominant prosodic pattern without any delay compared to monolinguals.

These findings, however, do not necessarily mean that French-"lexical stress language" bilinguals will process prosodic information as well as "variable stress language" monolinguals later in life. Indeed Dupoux et al. (2010) found that simultaneous French-Spanish bilingual adults had stress processing performance that fell in between that found for Spanish- and French-learning monolinguals. Moreover, specific experience with an L2 with variable stress was found to increase performance in a task in which French speakers learning an

L2 with variable lexical stress were judging whether syllable sequences were heard as made up of trochaic or iambic syllable pairs (Boll-Avetisyan et al., 2015) but not when they had to discriminate pairs of syllables with different stress patterns (Dupoux et al., 2008). Whether these differential effects are due to linguistic differences (e.g., in the L2 learned -German versus Spanish-, or the level of processing tapped by the experimental task used utterance vs. word) or the way experience with L2 was evaluated (see Boll-Avetisyan et al., 2015, for more detailed discussion) will need to be further explored. Such future studies will help specify the circumstances in which these early prosodic acquisitions in bilingual infants will (or will not) translate into efficient prosodic processing in adulthood.

Relatedly, one intriguing aspect of the present findings is the lack of an effect of the relative language exposure. Indeed, in our study, infants were hearing between 30 and 70% of German according to our estimation using a detailed language questionnaire. This means that only hearing 30% of German was apparently enough for these bilingual infants to acquire a trochaic bias at around the same age at which monolingual infants with a 100% of exposure to German show this bias. How can we reconcile these findings? First, it should be noted that little is known about the impact of language dominance on early language processing and acquisition. To the best of our knowledge, only one study explored the role of dominance for the processing of segmental information, and more precisely the acquisition of the phonotactic properties of the native language (Sebastián-Gallés and Bosch, 2002). They found that both Catalan-learning monolingual and Catalan-dominant Catalan/Spanish bilingual 11-month-olds had learned phonotactic properties of Catalan, while a similar but non-significant effect was found in Spanish-dominant Catalan/Spanish bilinguals of the same age, suggesting a weak dominance effect. For prosody, our two previous studies on stress pattern discrimination by 10-month-old bilinguals also revealed little to no effect of language dominance. Indeed, no impact of language dominance was found when using stimuli with segmental variability (Abboub et al., 2015). Moreover, the only marginal discrimination effect for the more balanced French/German bilinguals tested with materials lacking segmental variability compared to the significant effect of the German dominant bilinguals (Bijeljic-Babic et al., 2012) may reach significance by increasing statistical power when testing a higher number of infants. Overall, the existing evidence so far does not provide strong evidence that language dominance has large effects on infants' speech processing and their early phonological development.

So how could bilingual infants learn phonological properties of their native language at the same age as their monolingual peers, even though they very likely receive less input in each of their languages? One possibility is that the acquisition of linguistic properties requires not only a certain amount of exposure (and in the present case, 30% or more of German would be enough for bilinguals to get enough input), but also a certain amount of exposure time over development in order to detect and learn properties of the native language. This duration of exposure factor might be related to the need for some flexibility

in acquisition (in order not to learn too quickly erroneous properties, or to be able to unlearn an acquired property if it happens to be erroneous). It might also be linked to the need for certain neural networks, linguistic or cognitive abilities, to be set into place or reach a certain maturation level before phonological acquisition can take place. While this hypothesis would need to be tested empirically, note that it might be indirectly supported by data showing that while 6- and 8-month-old infants can learn a new consonant contrast in 2 min in the lab (Maye et al., 2002), it takes them around 8–10 months to learn native consonant categories in the real world (Werker and Tees, 1984; Kuhl et al., 1992), time during which they probably accumulate much more input than in the Maye et al. (2002) experiment. In this perspective, it would mean that the 30–70% of German input that our bilingual infants are receiving in the 6 first months of their lives is enough for them to learn that German words are predominantly trochaic. Note that this possible importance of duration of exposure might be more relevant for the acquisition of phonological properties than the acquisition of the lexicon, a domain of acquisition in which clear effects of quantity of input have been found (Hurtado et al., 2008), although certainly much more work on this issue will be needed, in both monolingual and bilingual populations.

A second reason for why our French/German bilingual infants were able to specify the predominant stress pattern of German within the same timeframe as German-learning monolinguals, related to bilingual acquisition *per se*, could be that bilingual infants have enhanced cognitive abilities, possibly as a result of hearing two languages at the same time. These enhanced abilities would allow them to learn properties of their native language more rapidly and with less input than needed by monolinguals. Evidence for such advantage has been found in early development, at 18 months of age in memory generalization studies (Brito and Barr, 2012), at 7 months in tasks requiring the acquisition of new linguistic rules (Kovács and Mehler, 2009), or at 6 months of age in studies measuring visual habituation as an index of efficiency in stimulus encoding (Singh et al., 2015). However, some authors have recently argued that the strength of this cognitive advantage in bilinguals remains to be confirmed, and its neural/cognitive bases specified (Costa and Sebastian-Galles, 2014). Future studies will then have to continue exploring these early language dominance effects, keeping in mind the difficulty of evaluating language input, and thus possibly using recording tools such as the LENA system (Oller et al., 2010) to quantify language dominance more objectively than by relying on parental reports.

The present findings also raise several questions regarding the specificity and generality of the prosodic acquisition trajectory found in our bilingual population. First, future studies should explore how the link between the prosodic properties of the two languages in acquisition impacts the developmental pattern that we uncovered. In the present case, French/German bilingual infants have to learn that one of their languages does not have lexical stress, while the other one has variable but predominantly trochaic lexical stress. The learning situation might be different, and lead to a different developmental trajectory, for infants learning two languages that both have lexical stress, but in

different positions within the words (that is for example, word-initial vs. word-final). This situation could constitute a more difficult acquisition context that might lead to delayed acquisition since infants would have to learn two different stress pattern assignment rules, rather than just one. Indeed, it remains to be determined whether some variation in the developmental trajectory related to early bilingual exposure can be found for the acquisition of prosodic properties, as has been found for the acquisition of segmental properties (e.g., Bosch and Sebastián-Gallés, 2003; Sebastián-Gallés and Bosch, 2009). Second, if the bias observed in the present study results from the acquisition of a language with trochaic lexical stress (German), then it should be observed in French/other language bilinguals if and only if their other language gives rise to a trochaic bias. In order to explore this prediction, we are in the process of testing bilingual infants, all learning French and a language other than German, separating these bilinguals depending on whether their second language would result in the acquisition of a trochaic bias or not.

Third, it will be of interest to determine whether the trochaic bias found in French/German bilingual 6-month-olds in the present experiment only applies to German stimuli, or whether it would also be found if such bilinguals were presented with stimuli that have segmental properties typical for French but not for German. Since infants theoretically can discriminate French and German from birth as these two languages have different rhythms (Mehler et al., 1988; Nazzi et al., 1998), they should be able to learn separate properties of these two languages and use them in language-appropriate ways already at 6 months of age.

CONCLUSION

The present study establishes the acquisition of a trochaic bias in French/German bilingual infants at 6 months of age, the same age at which this prosodic development has been found in monolingual infants (Höhle et al., 2009). This first study exploring the acquisition of a prosodic property by bilingual infants thus establishes that this acquisition is not delayed by bilingualism. Following up on this, it will be of interest to further explore the specificity of this developmental trajectory (as discussed above), and also its scope, in particular whether the acquisition of the trochaic bias in these bilingual infants will have implications at higher levels of linguistic processing. More specifically, it will be of interest to explore word form segmentation abilities in bilingual infants given evidence that the emergence of word form segmentation abilities is language-specific in monolingual infants, being based on the stress pattern in trochaic dominant languages (for English and Dutch: Jusczyk et al., 1999; Kooijman et al., 2005, 2009) but on the syllable in French (Nazzi et al., 2006; Goyet et al., 2010, 2013; Nishibayashi et al., 2015).

AUTHOR CONTRIBUTIONS

All authors designed the experiment, analyzed and discussed the results and contributed to the writing of the paper. RB-B organized the testing of the infants.

ACKNOWLEDGMENTS

This study was conducted with the support of an ANR-DFG grant (ANR-13-FRAL-0010 to RB-B and TN; DFG Ho 1960/15-1

to BH) and a LABEX EFL grant (ANR-10-LABX-0083 to RB-B and TN). Special thanks to the infants and their parents for their kindness and cooperation, and “CAFE bilingue” for help in recruiting the participants.

REFERENCES

- Abboub, N., Bijeljac-Babic, R., Serres, J., and Nazzi, T. (2015). On the importance of being bilingual: word stress processing in a context of segmental variability. *J. Exp. Child Psychol.* 132, 111–120. doi: 10.1016/j.jecp.2014.12.004
- Albareda-Castellot, B., Pons, F., and Sebastian-Galles, N. (2011). The acquisition of phonetic categories in bilingual infants: new data from an anticipatory eye movement paradigm. *Dev. Sci.* 14, 395–401. doi: 10.1111/j.1467-7687.2010.00989.x
- Bhatara, A., Boll-Avetisyan, N., Agus, T., Höhle, B., and Nazzi, T. (2015). Language experience affects grouping of musical instrument sounds. *Cogn. Sci.* doi: 10.1111/cogs.12300 [Epub ahead of print].
- Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., and Höhle, B. (2013). Native language and stimulus complexity affect rhythmic grouping of speech. *J. Acoust. Soc. Am.* 134, 3828–3843. doi: 10.1121/1.4823848
- Bijeljac-Babic, R., Serres, J., Höhle, B., and Nazzi, T. (2012). Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PLoS ONE* 7:e30843. doi: 10.1371/journal.pone.0030843
- Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T., and Höhle, B. (2015). Effects of experience with L2 and music on rhythmic grouping by French listeners. *Biling. Lang. Cogn.* (in press). doi: 10.1017/S1366728915000425
- Bosch, L., and Sebastián-Gallés, N. (1997). Native-language recognition abilities in four month-old infants from monolingual and bilingual environments. *Cognition* 65, 33–69. doi: 10.1016/S0010-0277(97)00040-1
- Bosch, L., and Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Lang. Speech* 46, 217–243. doi: 10.1177/00238309030460020801
- Brito, N., and Barr, R. (2012). Influence of bilingualism on memory generalization during infancy. *Dev. Sci.* 15, 812–816. doi: 10.1111/j.1467-7687.2012.1184.x
- Burns, T. C., Yoshida, K. A., Hill, K., and Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Appl. Psycholinguist.* 28, 455–474. doi: 10.1017/S0142716407070257
- Butler, J., Floccia, C., Goslin, J., and Panneton, R. (2011). Infants' discrimination of familiar and unfamiliar accents in speech. *Infancy* 16, 392–417. doi: 10.1111/j.1532-7078.2010.00050.x
- Byers-Heinlein, K., Burns, T. C., and Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychol. Sci.* 21, 343–348. doi: 10.1177/0956797609360758
- Christophe, A., and Morton, J. (1998). Is Dutch native English? Linguistic analysis by 2-month-olds. *Dev. Sci.* 1, 215–219. doi: 10.1111/1467-7687.00033
- Costa, A., and Sebastian-Galles, N. (2014). How does the bilingual experience sculpt the brain? *Nat. Rev. Neurosci.* 15, 336–345. doi: 10.1038/nrn3709
- Di Cristo, A. (1999). Vers une modélisation de l'accentuation du français: première partie. *J. French Lang. Stud.* 9, 143–179. doi: 10.1017/S095926950004671
- Dupoux, E., Pallier, C., Sebastian-Galles, N., and Mehler, J. (1997). A distressing “deafness” in French? *J. Mem. Lang.* 36, 406–421. doi: 10.1006/jmla.1996.2500
- Dupoux, E., Peperkamp, S., and Sebastian-Galles, N. (2001). A robust method to study stress “deafness”. *J. Acoust. Soc. Am.* 110, 1606–1618. doi: 10.1121/1.1380437
- Dupoux, E., Peperkamp, S., and Sebastian-Gallés, N. (2010). Limits on bilingualism revisited: stress “deafness” in simultaneous French-Spanish bilinguals. *Cognition* 114, 266–275. doi: 10.1016/j.cognition.2009.10.001
- Dupoux, E., Sebastian-Gallés, N., Navarrete, E., and Peperkamp, S. (2008). Persistent stress “deafness”: the case of French learners of Spanish. *Cognition* 106, 682–706. doi: 10.1016/j.cognition.2007.04.001
- Féry, C., Hörnig, R., and Pahaut, S. (2011). “Correlates of phrasing in French and German from an experiment with semi-spontaneous speech,” in *Intonational Phrasing in Romance and Germanic: Cross-Linguistic and Bilingual Studies*, Vol. 10, eds C. Gabriel and C. Lleo (Amsterdam: John Benjamins Publishing), 11–42. doi: 10.1075/hsm.10.03fer
- Friederici, A. D., Friedrich, M., and Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Curr. Biol.* 17, 1208–1211. doi: 10.1016/j.cub.2007.06.011
- Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., et al. (2011). Bilingual language learning: an ERP study relating early brain responses to speech, language input, and later word production. *J. Phonetics* 39, 546–557. doi: 10.1016/j.wocn.2011.07.002
- Goyet, L., de Schonen, S., and Nazzi, T. (2010). Syllables in word segmentation by French-learning infants: an ERP study. *Brain Res.* 1332, 75–89. doi: 10.1016/j.brainres.2010.03.047
- Goyet, L., Nishibayashi, L.-L., and Nazzi, T. (2013). Early syllabic segmentation of fluent speech by infants acquiring French. *PLoS ONE* 8:e79646. doi: 10.1371/journal.pone.0079646
- Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press.
- Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.
- Herold, B., Höhle, B., Walch, E., Weber, T., and Obladen, M. (2008). Impaired stress pattern discrimination in very low birth weight infants during the first six months of life. *Dev. Med. Child Neurol.* 50, 678–683. doi: 10.1111/j.1469-8749.2008.03055.x
- Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B., and Kennedy, L. (1987). Clauses are perceptual units for young Infants. *Cognition* 26, 269–286. doi: 10.1016/S0010-0277(87)80002-1
- Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., and Nazzi, T. (2009). The development of language specific prosodic preferences during the first half year of life: evidence from German and French. *Infant Behav. Dev.* 2, 262–274. doi: 10.1016/j.infbeh.2009.03.004
- Höhle, B., Pauen, S., Hesse, V., and Weissenborn, J. (2014). Discrimination of rhythmic pattern at 4 months and language performance at 5 years: a longitudinal analysis of data from German-learning infants. *Lang. Learn.* 64, 141–164. doi: 10.1111/lang.12075
- Hurtado, N., Marchman, V. A., and Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Dev. Sci.* 11, F31–F39. doi: 10.1111/j.1467-7687.2008.00768.x
- Jun, S. A., and Fougeron, C. (2000). “A phonological model of French intonation,” in *Intonation: Analysis, Modeling and Technology*, ed. A. Botinis (Dordrecht: Kluwer Academic Press), 209–242.
- Jun, S. A., and Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus* 14, 147–172. doi: 10.1515/prbs.2002.002
- Jusczyk, P. W., Cutler, A., and Redanz, N. J. (1993). Preference for the predominant stress patterns of English words. *Child Dev.* 64, 675–687. doi: 10.2307/1131210
- Jusczyk, P. W., Houston, D. M., and Newsome, M. (1999). The beginning of word segmentation in English-learning infants. *Cogn. Psychol.* 39, 159–207. doi: 10.1006/cogp.1999.0716
- Kooijman, V., Hagoort, P., and Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cogn. Brain Res.* 24, 109–116. doi: 10.1016/j.cogbrainres.2004.12.009
- Kooijman, V., Hagoort, P., and Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds. *Infancy* 6, 591–612. doi: 10.1080/15250000903263957
- Kovács, A., and Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6556–6560. doi: 10.1073/pnas.0811323106
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–608. doi: 10.1126/science.1736364
- Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111. doi: 10.1016/S0010-0277(01)00157-3

- Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition* 29, 143–178. doi: 10.1016/0010-0277(88)90035-2
- Michelas, A., and D'Imperio, M. (2015). Prosodic boundary strength guides syntactic parsing of French utterances. *Lab. Phonol.* 6, 119–146. doi: 10.1515/lp-2015-0003
- Molnar, M., Gervain, J., and Carreiras, M. (2013). Within-rhythm class native language discrimination abilities of Basque-Spanish monolingual and bilingual infants at 3.5 months of age. *Infancy* 19, 326–337. doi: 10.1111/inf.12041
- Moon, C., Panneton-Cooper, R., and Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behav. Dev.* 16, 495–500. doi: 10.1016/0163-6383(93)80007-U
- Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *J. Exp. Psychol.* 24, 756–776.
- Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., and Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: emerging evidence for crosslinguistic differences. *J. Mem. Lang.* 54, 283–299. doi: 10.1371/journal.pone.0079646
- Nazzi, T., Jusczyk, P. W., and Johnson, E. K. (2000). Language discrimination by English learning 5-month-olds: effects of rhythm and familiarity. *J. Mem. Lang.* 43, 1–19. doi: 10.1006/jmla.2000.2698
- Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., and Donati, C. (2008). Different phrasal prominence realization in VO and OV languages. *Lingue Linguaggio* 7, 1–28.
- Nishibayashi, L.-L., Goyet, L., and Nazzi, T. (2015). Early speech segmentation in French-learning infants: monosyllabic words versus embedded syllables. *Lang. Speech* 58, 334–350. doi: 10.1177/0023830914551375
- Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., et al. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13354–13359. doi: 10.1073/pnas.1003882107
- Pons, F., and Bosch, L. (2010). Stress pattern preference in Spanish-learning infants: the role of syllable weight. *Infancy* 15, 223–245. doi: 10.1111/j.1532-7078.2009.00016.x
- Sansavini, A., Bertoncini, J., and Giovanelli, G. (1994). Newborns discriminate stress patterns in phonetically complex Italian words. *Infant Behav. Dev.* 17:924.
- Sansavini, A., Bertoncini, J., and Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Dev. Psychol.* 33, 3–11. doi: 10.1037/0012-1649.33.1.3
- Sebastián-Gallés, N., and Bosch, L. (2002). Building phonotactic knowledge in bilinguals: role of early exposure. *J. Exp. Psychol.* 28, 974–989.
- Sebastián-Gallés, N., and Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: is the distributional account all there is to it? *Dev. Sci.* 12, 874–887. doi: 10.1111/j.1467-7687.2009.00829.x
- Singh, L., Fu, C. S. L., Rahman, A. A., Hameed, W. B., Sanmugam, S., Agarwal, P., et al. (2015). Back to basics: a bilingual advantage in infant visual habituation. *Child Dev.* 86, 294–302. doi: 10.1111/cdev.12271
- Skoruppa, K., Cristià, A., Peperkamp, S., and Seidl, A. (2011). English-learning infants' perception of word stress patterns. *J. Acoust. Soc. Am.* 130, EL50–EL55. doi: 10.1121/1.3590169
- Skoruppa, K., Pons, F., Bosch, L., Christophe, A., Cabrol, D., and Peperkamp, S. (2013). The development of word stress processing in French and Spanish infants. *Lang. Learn. Dev.* 9, 88–104. doi: 10.1080/15475441.2012.693881
- Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., et al. (2009). Language-specific stress perception by nine-month-old French and Spanish infants. *Dev. Sci.* 12, 914–919. doi: 10.1111/j.1467-7687.2009.00835.x
- Spring, D. R., and Dale, P. S. (1977). Discrimination of linguistic stress in early infancy. *J. Speech Hear. Res.* 20, 224–232. doi: 10.1044/jshr.2002.224
- Sundara, M., Polka, L., and Molnar, M. (2008). Development of coronal stop perception: bilingual infants keep pace with their monolingual peers. *Cognition* 108, 232–242. doi: 10.1016/j.cognition.2007.12.013
- Sundara, M., and Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *J. Phonet.* 39, 505–513. doi: 10.1016/j.wocn.2010.08.006
- Vaissière, J., and Michaud, A. (2006). “Prosodic constituents in French: a data-driven approach,” in *Prosody and Syntax*, eds I. Fonagy, Y. Kawaguchi, and T. Moriguchi (Amsterdam: John Benjamins Publishing), 47–64.
- Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63. doi: 10.1016/S0163-6383(84)80022-3

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Bijeljac-Babic, Höhle and Nazzi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Limits on Monolingualism? A Comparison of Monolingual and Bilingual Infants' Abilities to Integrate Lexical Tone in Novel Word Learning

Leher Singh*, Felicia L. S. Poh and Charlene S. L. Fu

Department of Psychology, National University of Singapore, Singapore, Singapore

OPEN ACCESS

Edited by:

Miquel Simonet,
University of Arizona, USA

Reviewed by:

Katharine Graf-Estes,
University of California, Davis, USA
Jessica Hay,
University of Tennessee, USA

*Correspondence:

Leher Singh
psyls@nus.edu.sg

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 31 January 2016

Accepted: 21 April 2016

Published: 10 May 2016

Citation:

Singh L, Poh FLS and Fu CSL (2016)
Limits on Monolingualism?
A Comparison of Monolingual
and Bilingual Infants' Abilities
to Integrate Lexical Tone in Novel
Word Learning. *Front. Psychol.* 7:667.
doi: 10.3389/fpsyg.2016.00667

To construct their first lexicon, infants must determine the relationship between native phonological variation and the meanings of words. This process is arguably more complex for bilingual learners who are often confronted with phonological conflict: phonological variation that is lexically relevant in one language may be lexically irrelevant in the other. In a series of four experiments, the present study investigated English–Mandarin bilingual infants' abilities to negotiate phonological conflict introduced by learning both a tone and a non-tone language. In a novel word learning task, bilingual children were tested on their sensitivity to tone variation in English and Mandarin contexts. Their abilities to interpret tone variation in a language-dependent manner were compared to those of monolingual Mandarin learning infants. Results demonstrated that at 12–13 months, bilingual infants demonstrated the ability to bind tone to word meanings in Mandarin, but to disregard tone variation when learning new words in English. In contrast, monolingual learners of Mandarin did not show evidence of integrating tones into word meanings in Mandarin at the same age even though they were learning a tone language. However, a tone discrimination paradigm confirmed that monolingual Mandarin learning infants were able to tell these tones apart at 12–13 months under a different set of conditions. Later, at 17–18 months, monolingual Mandarin learners were able to bind tone variation to word meanings when learning new words. Our findings are discussed in terms of cognitive adaptations associated with bilingualism that may ease the negotiation of phonological conflict and facilitate precocious uptake of certain properties of each language.

Keywords: lexical tone, phoneme discrimination, infant speech perception, Mandarin Chinese, word learning

INTRODUCTION

Languages of the world make use of sound in different ways to create words. A classic example is the use of vocal pitch in human languages. When learning a tone language like Mandarin Chinese, listeners must register particular changes in vocal pitch that distinguish the meanings of words. However, pitch variation is also a ubiquitous feature of non-tone languages such as English and is used to distinguish questions/statements, emotional states, and placement of stress and focus. In contrast to Mandarin learners, English learners must disregard pitch variation when determining the lexical identity of a word. It is therefore incumbent upon the young language

learner to determine how sound changes effect changes in word meaning in their native language to construct a vocabulary. By necessity, children learning two languages have to learn how words are defined in both of their native languages. This process is potentially complicated by the fact that the phonological rules of two languages can diverge as in the case of Mandarin and English where pitch varies lexically and non-lexically, respectively, causing a potential conflict. The purpose of the current study is to determine how bilingual infants resolve this conflict and negotiate cross-language phonological conflict when learning new words. Specifically, the present study focuses on English–Mandarin bilingual infants' abilities to define words according to lexical tone when listening to Mandarin and to disregard the same source of variation in pitch when defining new words in English. Bilingual infants' abilities to integrate pitch in a language-dependent fashion are interpreted in relation to those of monolingual tone language learners.

In prior research, children's abilities to integrate native phonological variation when learning new words have been widely studied in monolingual children (Stager and Werker, 1997; Pater et al., 2004; Dietrich et al., 2007; Rost and McMurray, 2009, 2010; Yoshida et al., 2009), but to a much lesser extent in bilingual children (but see Fennell et al., 2007; Mattock et al., 2010; Byers-Heinlein et al., 2013; Fennell and Byers-Heinlein, 2014). A substantial proportion of this research has used the Switch task, which has been productively used to investigate infants' abilities to map similar sounding words onto different meanings. In a common instantiation of this task, infants are familiarized with an on-screen display of two objects and their labels. Labels consist of novel words that are subtle phonemic variants – or minimal pairs (e.g., 'bih' and 'dih'). During a habituation phase, infants are presented with repetitions of each pairing until their attention to the objects wanes to a pre-set criterion. Following the habituation phase, infants are presented with two test trials. In one test trial (Same trial), infants are presented with the pairing with which they were familiarized. In the other test trial (Switch trial), infants are presented with the visual object with which they were familiarized but it is labeled with the name for the other object (e.g., what was learned as a 'bih' is now labeled as a 'dih'). Infants' fixation times to each trial type are compared: a relative elevation in fixation to the Switch trial versus the Same trial is interpreted as evidence of infants' sensitivity to the source of phonological variation incorporated into the task (i.e., to variation in place of articulation in the current example).

In a seminal study that pioneered the Switch task to investigate early word learning, Stager and Werker (1997) demonstrated that 14-month-old monolingual infants failed to incorporate phonological variation (i.e., the difference between 'b-' versus 'd-') when learning new words, although they could incorporate the same variation when recognizing familiar words (Fennell and Werker, 2003). Comparative studies with bilingual infants reveal a similar set of abilities provided that bilingual infants are provided with input that is consistent with the phonetic properties of their input (i.e., input that sounds native to them). In one such study by Mattock et al. (2010), the authors presented 17-month-old bilingual infants with tokens drawn from both

of their languages. Mattock et al. (2010) demonstrated that under these conditions, bilingual infants linked similar sounding words to their meanings at 17 months. More recently, Fennell and Byers-Heinlein (2014) demonstrated that both 17-month-old monolingual and bilingual infants succeeded in learning similar sounding words when the speaker matched their language background (i.e., when the speaker was monolingual or bilingual, respectively), although bilingual infants were not able to learn similar sounding words when presented with monolingual input (see Fennell et al., 2007). In sum, this set of studies suggests that both 17- to 18-month-old monolingual and bilingual infants maintain keen perceptual sensitivities to subtle phonetic detail that are optimally engaged when they listen to language input reminiscent of their environment.

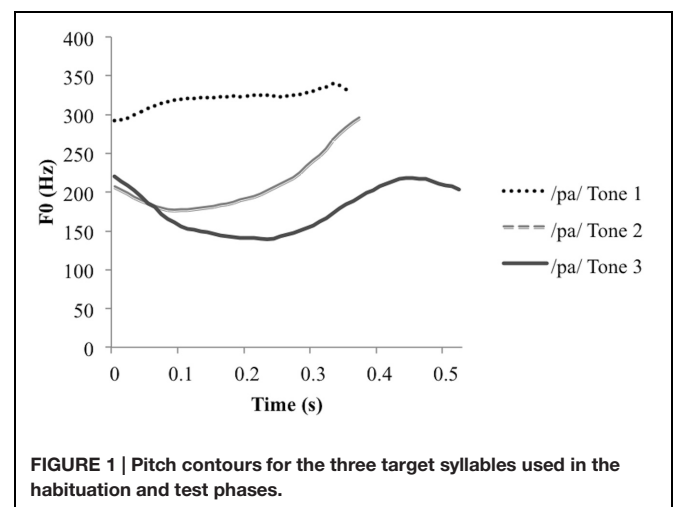
Previous research has focused on bilingual infants' sensitivity to phonological variation that draws lexical distinctions in both of their native languages (although the sub-phonetic realization of these sounds varies across languages; e.g., Fennell et al., 2007; Mattock et al., 2010; Fennell and Byers-Heinlein, 2014). Nevertheless, in each of the aforementioned studies, the phonemes used to distinguish word meanings belonged to separate phonetic categories in *both* languages. However, bilinguals often have to negotiate phonological conflict where the same source of variation draws lexical distinctions only in one language and not in the other. In this situation, learners of two languages have to alternate between activating and de-activating sensitivity to this source of variation depending on the language in use. For example, learners of Mandarin–Chinese and English have to inhibit integration of pitch variation when defining new words in English but have to incorporate certain forms of pitch variation (i.e., tone contrasts) when learning new words in Mandarin. One prior study has investigated bilingual English–Mandarin infants' abilities to integrate tone in English and Mandarin in a language-selective manner. In a word segmentation task investigating how effectively infants segment words from passages, Singh and Foong (2012) familiarized infants with isolated words and then tested infants' recognition of the familiarized words in fluent speech. Each infant was tested in English and in Mandarin in succession. The critical manipulation was that in the test phase, the target word was either matched or mis-matched in tone (Mandarin session) or matched or mis-matched in pitch (English session). Infants were tested at 7.5-, 9-, and 11-months. While infants did not demonstrate language-selective integration of pitch at 7.5- and 9-months (either integrating pitch/tone variation or disregarding pitch/tone variation in both languages), at 11 months, infants selectively defined words by tone in Mandarin and not by pitch in English. However, this study did not involve forming word-object associations, as it was an auditory-only word segmentation task, rendering it unclear as to whether infants linked the familiarized words to meaning. Additionally, the pitch transformations qualitatively differed between English and Mandarin sessions: Mandarin pitch variants encompassed Mandarin lexical tone contrasts, while English pitch variants were digitized, uniform transformations across the entire syllable. However, most crucially, word segmentation is thought to measure an infants' ability to track repetitions

of the same word and prior to 12 months, and is thought to precede an infants' determination of meaning (Jusczyk and Aslin, 1995).

Subsequent studies investigating integration of pitch and tone when forming word-object associations reveal more fragile abilities in young children when they are required to link words to meaning. Influences of tone variation in newly learned words have been investigated in English monolingual infants, non-tone language learning bilingual infants and English-Mandarin bilingual infants (Singh et al., 2014; Graf Estes and Hay, 2015; Hay et al., 2015). Collectively, these findings suggest that the language-specific functions of pitch are not consolidated as early as 11 months. Using a preferential looking paradigm, a study by Singh et al. (2014) involved teaching infants novel tone-marked words in a referential context. Infants were then tested on their recognition of tone-matched and tone-varying labels of familiarized words (as well as vowel matches/variants). The authors reported that non-tone learning infants (monolingual and bilingual) were similar to their Mandarin learning peers in that they were sensitive to tone as a source of lexical contrast, rejecting tone variants as labels for words at 18 months. It was not until 24 months that non-tone learning infants (monolingual and bilingual) demonstrated selective inhibition to tone in English when learning new words, whereas Mandarin learning infants continued to associate and integrate lexical tone into newly learned words at 24 months. Tone integration was reflected by participant's construal of tone changes as mispronunciations of newly learned words. In an investigation of tone sensitivity in English monolingual infants using the Switch paradigm, Hay et al. (2015) investigated English learning infants' sensitivity to rising and falling tones when learning new words at 14, 17, and 19 months. Infants exhibited developmental change in tone sensitivity between 14 and 17–19 months: while 14-month-old infants were sensitive to tone variation, at 17 and 19 months, infants were no longer sensitive to the same source of tone variation in the Switch paradigm. Posing this question with bilingual infants learning two non-tone languages, Graf Estes and Hay (2015) reported a protracted period of tone sensitivity in bilingual learners, demonstrating that these infants were sensitive to lexical tones at 14 and 19 months, but not at 22 months. In the aggregate, it appears that when infants are confronted with the added burden of forming word-object associations, their sensitivity to phonological variation appears much more fragile than when they are simply tracking repetitions of words across time as in Singh and Foong's study. However, in Singh et al. (2014), although tone learners were English-Mandarin bilinguals, the language context of newly learned words was not manipulated within bilingual participants. As such, it was not possible to examine whether bilingual participants could actually shift their interpretation of tone as befitted the language context. The ability on the part of bilingual learners to re-interpret the same phonetic information in a language-selective manner – termed perceptual switching – has been well researched in adult bilinguals (Flege and Eefting, 1987; Hazan and Boulakia, 1993; García-Sierra et al., 2009, 2012; Gonzales and Lotto, 2013) and to a limited extent, in children (Singh and Quam, 2016), but not yet in infants.

However, this process of rapid alternation is a fundamental component of bilingual proficiency. The current study focuses on monolingual and bilingual infants' abilities to alternate between the phonological systems of each of their languages when these systems conflict.

The primary goal of this study is to compare monolingual and bilingual phonological representations of lexical tone by assessing infants' responsiveness to tone mispronunciations in their native language(s). In light of the multi-functionality of pitch in English-Mandarin bilingual infants' environments, infants were provided with naming phrases ending with target words to cue a particular language (i.e., English or Mandarin). Prior research has demonstrated that bilingual infants make productive use of naming phrases to identify the relevant phonological rules (see Fennell and Byers-Heinlein, 2011). A secondary goal of the present study was to determine whether sensitivity to a change in lexical tone depended not only on the language in use, but furthermore, on the acoustic salience of the tone change. Mandarin Chinese has four lexical tones [high (Tone 1), rising (Tone 2), dipping (Tone 3), falling (Tone 4)], three of which (Tones 1, 2, and 3) were used in our study (please see **Figure 1** for an illustration of Tones 1, 2, and 3). Some tones are highly distinctive from one another (such as Tones 1 and 3) such that Mandarin speakers readily discriminate them (Chen, 2013). Other tones are highly similar, notably Tones 2 and 3, such that these tones are often poorly discriminated (Zue, 1976; Shen and Lin, 1991). Prior studies investigating infants' sensitivity to lexical tones have revealed that sensitivity to lexical tone pairs progresses asynchronously for different tone pairs (see Mattock and Burnham, 2006; Tsao, 2008; Yeung et al., 2013; Liu and Kager, 2014). An important determinant of lexical tone perception appears to be the salience of the tone contrast (see Liu and Kager, 2014 and Tsao, 2008 for investigations of sensitivity to high and low salience tone contrasts), a pattern also evidenced in production (e.g., Wong et al., 2005). Prior studies have demonstrated that emergent sensitivity to lexical tone contrasts do not necessarily generalize across the entire tone inventory (see Singh and Fu, 2016, for a review of this evidence in perception



and production). Conclusions drawn about tone sensitivity are therefore necessarily qualified by the relative similarity of a given tone pair. Tone similarity is commonly defined by properties of the pitch contour (Gandour, 1983), primarily by pitch direction and secondarily by pitch height (Chandrasekaran et al., 2010). In light of discrepant effects of similar and distinct tone pairs on tone sensitivity, in the current study, infants' sensitivity to lexical tone as a source of contrast was compared across similar and distinct tone pairs.

A series of four experiments are reported. In Experiment 1, 12–13-month-old bilingual English–Mandarin infants were tested on a similar task, but were tested in both Mandarin and English in direct succession. In Experiment 2, 12–13-month-old monolingual Mandarin learning infants were tested on their sensitivity to lexical tone contrasts when learning novel words in Mandarin. Experiments 3 and 4 were designed to further investigate the apparent insensitivity to lexical tone observed in Mandarin learning monolingual infants at 12–13 months. Experiment 3 investigated whether Mandarin learning monolingual infants could discriminate the tones used in Experiment 2, even though they did not appear sensitive to variation in these tones when learning novel words. Experiment 4 investigated whether Mandarin learning monolingual infants could integrate lexical tone contrasts at a later age, testing 18-month-old infants on the same word learning task administered to 12- to 13-month-old Mandarin infants in Experiment 2.

EXPERIMENT 1

In Experiment 1, we investigated whether bilingual infants, learning English and Mandarin, integrated tone in a language-selective manner within each of their native languages. The purpose of this experiment was to determine whether habitual exposure to two native languages that conflicted in their use of tone would facilitate a language-selective interpretation of tone. We hypothesized that the contrastive use of tone in each of the participants' native languages would contribute to a more mature understanding of the linguistic functions served by tone in each language.

Infants were familiarized with a word object pairing via the Switch paradigm. The label used to introduce the object was spoken in Tone 3. After successfully habituating to the pairing, infants were tested on their sensitivity to a similar (Tone 2 versus Tone 3) mispronunciation and to a distinct (Tone 1 versus Tone 3) mispronunciation. Infants were tested in each of their native languages: English and Chinese. Responses to each type of tone mispronunciation were compared across languages.

Methods

Participants

Our sample comprised eighteen 12- to 13-month-old Mandarin–English bilingual infants (age range: 12 months 10 days to 13 months 21 days, average age = 13 months 1 day). All infants were born healthy and full term. Another seven infants were

tested but excluded from the final sample due to fussiness during test ($n = 6$) or on account of data that deviated from the group mean by more than 3 standard deviations ($n = 1$). All infants received between at least 35% exposure to English or Mandarin with no third language exposure (range of English exposure: 38 to 63%, mean = 51%; range of Mandarin exposure: 37 to 62%; mean = 48%). Language exposure was determined by the Language Exposure Questionnaire developed by Bosch and Sebastián-Gallés (1997). Language exposure was derived from parental estimates of the relative proportion that each caregiver used when communicating directly to the child, and the amount of time each caregiver spent with the child in a typical week.

The age of testing was motivated by prior research investigating sensitivity to suprasegmental lexical variation (see Curtin, 2009). When tasked with learning minimally contrastive words differing in lexical stress, Curtin (2009) demonstrated that infants were sensitive to contrasts in stress at 1213 months. This finding stands in contrast to the broad swath of studies defining similar sounding words by consonant variation demonstrating that infants are challenged by this task prior to 14 months (e.g., Stager and Werker, 1997; Werker et al., 2002; Pater et al., 2004). As concluded by Curtin (2009), it appears that suprasegmental lexical variation is integrated into word meaning earlier than segmental variation. As our study manipulated suprasegmental lexical variation (i.e., tones), we tested infants at 12–13 months. This study was carried out with the approval of the National University of Singapore Institutional Review Board. Participants' parents or legal guardians gave written informed consent in accordance with the National University of Singapore Institutional Review Board requirements.

Stimuli

Auditory stimuli for the study consisted of seven Mandarin and seven English naming phrases adapted from Fennell and Waxman (2010) (see Table 1). The target word was the label “pa” produced in Tones 1, 2, and 3 by a female native speaker of Mandarin in the context of each naming phrase. All stimuli were produced in infant-directed speech. The mean duration of the Mandarin phrases was 1.28 s (SD : 0.4) and the mean duration of the English phrases was 1.14 s (SD : 0.3). English and Mandarin phrase durations did not differ significantly. The

TABLE 1 | Naming phrases used in Experiments 1, 2, and 4.

Mandarin carrier phrases	English carrier phrases
看, 那是个 ____.	Look, it's the ____.
这是个 ____.	Here's the ____.
看一看, 有个 ____.	Do you see the ____?
那是个 ____.	There's the ____.
哪里有个 ____?	Where's the ____?
看那个 ____.	Look at the ____.
我喜欢这个 ____.	I like the ____.

mean pitch range of the carrier phrases was 288.14 Hz (*SD*: 45.81; Mandarin) and 277.95 Hz (*SD*: 48.51; English), which again, did not differ significantly across languages. The mean duration of the target words was 0.47 s (*SD*: 0.07). The same tokens were spliced into English and Mandarin introductory phrases to mitigate possible effects of language-specific differences in tone productions. Each instantiation of the target syllable was separated by 800 ms.

The target word was labeled by the syllable /pa/, which begins with an unaspirated voiceless onset consonant. This segment was chosen for the entire series of experiments because it assimilates to the native phonological inventories of English and Mandarin. In English, the unaspirated /p/ typically follows a word-initial /s/, such as in “spa,” but it does not appear in the word-initial position. However, unaspirated voiceless stops in word-initial position sound native to English speakers and are classified as voiced stops (in this case, ‘ba’; Pegg and Werker, 1997). They are judged to be as good an instance of ‘ba’ as the voiced stop ‘ba’ when produced in word-initial position (Lisker and Abramson, 1964).

Acoustic analyses of the target syllable, /pa/, were conducted to ensure that the tokens matched monolingual productions within each language. The voice-onset-time (VOT) values of the three tokens ranged from 11 ms (Tone 3 production) to 18 ms (Tone 1 production). These values overlap with published VOT values of monolingual Mandarin productions that range from 11 to 18 ms (Liao, 2005; Chao et al., 2006; Chen et al., 2007; Deterding and Nolan, 2007) as well as with English monolingual values for /ba/ (Lisker and Abramson, 1967). Formant values also fell within the range of values reported for Mandarin and English monolingual productions (Mandarin monolingual F1: 1104 Hz, English monolingual F1: 850, bilingual F1: 802.7–1213.6 Hz, Mandarin monolingual F2: 1593.6 Hz, English monolingual F2: 1220, bilingual F2: 1046.3–1633.2 Hz; Peterson and Barney, 1952; Zee and Lee, 2001). F3 was not examined as it relates to lip rounding, which is not used contrastively for the target vowel in English or Mandarin. Auditory stimuli were accompanied by a visually presented novel object (see **Figure 2**) that moved in a circular path. Objects were counterbalanced to each language across participants.

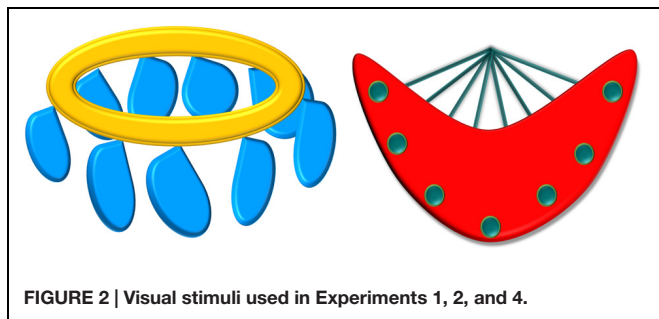
English and Mandarin versions of the task were created. The target word was paired with a different object in each language. However, the target word remained the same so as to determine

whether infants were capable of switching to a new set of phonological rules based on contextual cues alone.

Procedure

Before testing, all caregivers provided informed consent for their child’s participation, in accordance with guidelines set out by the National University of Singapore Institutional Review Board. Infants sat on their parents’ lap in a dimly lit testing suite facing a computer screen. Parents were asked not to interact with their child during the session. The experimenter observed the infant’s behavior from an adjoining room. During the experiment, both the parent and the experimenter listened to instrumental masking music.

During the task, novel objects were presented in the context of naming phrases to infants in the Switch task (Stager and Werker, 1997; Fennell and Waxman, 2010). The experiment consisted of a habituation and test phase. Before each trial, an attention getter was presented. Trials were initiated when infants oriented to the visual display. When the infant fixated to the visual display, the habituation phase commenced. Habituation consisted of repeated presentations of the target word /pa/ in Tone 3, embedded within the naming phrases and presented with the novel object. The habituation phase terminated when infant’s looking times to two trials decreased to less than 65% of two longest consecutive trials or until the infant completed a maximum of 24 trials. This habituation criterion was informed by a prior study that used the Switch task with carrier sentences (Fennell and Waxman, 2010). Once either of these criteria was met, the test phase commenced. The test phase included a Same trial and two Switch trials as adopted in previous studies (e.g., Curtin, 2010; Escudero et al., 2014). Trial order was counterbalanced across infants. For the Same trial, infants were presented with the word-object pairing to which they had habituated (i.e., /pa/ in Tone 3). The Switch trials violated this pairing, presenting infants with the same visual stimulus but paired with the target word /pa/ produced in Tones 1 and 2, respectively. Across all phases, trials lasted for a maximum of 20 s, or until the child looked away from the screen for more than 2 s. Trials were repeated if infants fixated to the screen for less than 1 s. Following the test trials, a post-test was presented. This consisted of a novel object produced by a different female speaker and labeled as a /pɪ/ produced in a novel tone (Tone 4). The object was animated to enlarge and shrink on the screen. A post-test trial is commonly included in the Switch paradigm to provide an indication of attention to the task during the terminal phase of the experiment. In prior studies (e.g., Fennell et al., 2007; Byers-Heinlein et al., 2013), fixation to the last habituation block has been compared to fixation to the post-test trial. Elevated attention (recovery) between these is recruited as an interpretative safeguard against a Type II error: in the event of a null result whereby fixation to Same and Switch trials do not differ, the presence of recovery between the last habituation block and the post-test trial indicates that this is unlikely to be accounted for by fatigue or disengagement from the experiment during the test trials. An example of the stimuli is provided in **Figure 2**. Infants were presented with a Mandarin and an English version of the same task. The order of presentation of the English and Mandarin task was counterbalanced across infants. Between



the two tasks, infants were presented with a 1-min non-verbal cartoon.

Both the Switch task and preferential looking approaches are well-established measures of infants' sensitivity to phonological variation when learning new words. However, in pilot studies, a preferential looking approach to the present task (including relevant parameters such as two languages, three test trial types) proved excessively demanding for participants. Each session was substantially longer than the auditory word segmentation task used within subjects by Singh and Foong (2012), and in recent research, a preferential looking approach to the question of perceptual switching was only successfully used in older children at 3–5 years of age (see Singh and Quam, 2016). As a consequence, the Switch task was selected for the current study. It should be noted that it is possible to use the Switch task to measure sensitivity to phonological variation using two objects (e.g., Werker et al., 2002; Fennell and Werker, 2003). However, familiarization with two objects could not be integrated into a design with a three trial [Same; Switch (distinct); Switch (similar)] test phase. An alternative design would have been to incorporate a two-trial (Same and Switch) test phase and manipulate contrast salience across participants. We prioritized the manipulation of salience as a within-subjects contrast in light of the fact that our sample comprised bilingual infants; a between-subjects comparison between two groups of bilinguals can introduce differences in performance due to background variables (specifically, the nature and extent of bilingual input, which are hard to match across bilingual groups with precision). Uncontrolled effects of error variance due to individual variation are somewhat mitigated by within-subjects comparisons, which motivated our decision to incorporate a single object and to manipulate salience within participants for each experiment. Although less common than a two-object paradigm, a single-object Switch paradigm has been used in several prior studies (see Stager and Werker, 1997; Werker et al., 1998; Pater et al., 2004; Thiessen, 2007; Fennell and Waxman, 2010; Fennell, 2012).

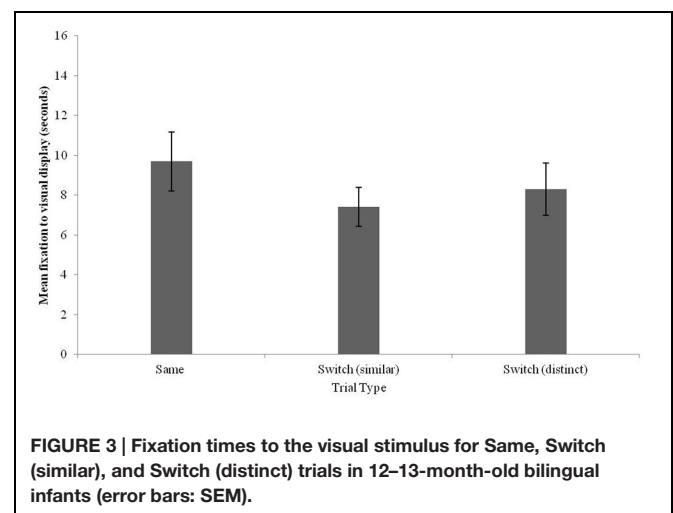
Results

All infants habituated within the 24 trial maximum habituation window. A preliminary analysis was conducted to determine whether participants recovered to the post-test by comparing the last habituation block to the post-test stimulus. A 2×2 (phase: last habituation block/post-test \times language: English/Mandarin) repeated-measures ANOVA revealed a main effect of phase [$F(1,34) = 13.91$, $p = 0.001$, $\eta_p^2 = 0.29$], accounted for by an elevation in fixation times between the last habituation block and the post-test. There were no effects of language on fixation times nor was there an interaction of phase and language on fixation times ($p > 0.8$).

An initial set of analyses was conducted to determine if there was an effect of test order on fixation times to test trials. A $3 \times 2 \times 2$ (Trial type: Same; Switch-similar; Switch-distinct \times Language: English; Mandarin \times Order: Mandarin first; English first) repeated-measures ANOVA was conducted with fixation times during test trials as the dependent variable. Results revealed no effects of interactions with order ($p > 0.3$). Fixation

times were therefore collapsed across test orders for subsequent analyses.

As the other of test trials was rotated across participants, a preliminary analysis was conducted to investigate effects and interactions test trial order, trial type, and language, revealing no effects or interactions with test trial order ($p > 0.6$). Test trial order was excluded from subsequent analyses. A 3×2 Trial type \times Language repeated-measures ANOVA was then conducted. Results revealed a main effect of trial type [$F(2,34) = 11.18$, $p = 0.0001$, $\eta_p^2 = 0.39$], no main effect of language ($p = 0.23$) and no interaction of trial type and language [$F(2,34) = 2.46$, $p = 0.1$]. Planned comparisons were conducted within each language to determine whether participants differed in how they responded to each tone change based on the language of testing. For each language, a repeated measures ANOVA was conducted to determine the effect of trial type (Same; Switch-similar; Switch-distinct) on fixation times to test trials. When participants were tested in Mandarin, results revealed a main effect of trial type [$F(2,34) = 10.56$, $p = 0.0001$, $\eta_p^2 = 0.39$]. Simple contrasts revealed higher fixation times to Switch-distinct trials than to Same trials [$F(1,17) = 20.35$, $p > 0.0001$, $\eta_p^2 = 0.54$] as well as higher fixation times to Switch-similar trials than to Same trials [$F(1,17) = 5.93$, $p = 0.03$, $\eta_p^2 = 0.26$]. A *post hoc* analysis comparing fixation times to Same and Switch trials for the two Switch trials (similar and distinct) demonstrated that differences in Same–Switch trials were greater for when the Switch involved a distinct contrast (i.e. change from Tone 3 to Tone 1) than when it involved a similar contrast [i.e., change from Tone 3 to Tone 2; $t(17) = 2.3$, $p = 0.04$ (Cohen's d : 0.57)]. This analysis revealed effects of perceptual salience on tone integration in Mandarin, although both similar and distinct substitutions were recognized as lexically contrastive. When participants were tested in English, results revealed a main effect of trial type [$F(2,34) = 3.27$, $p = 0.05$]. Simple contrasts revealed no significant difference in fixation to Switch-distinct trials than to same trials [$F(1,17) = 3.15$, $p = 0.1$] nor to Switch-similar trials than to Same trials [$F(1,17) = 0.54$, $p = 0.47$]. Fixation



times to each trial type for English and Mandarin are plotted in **Figure 3**.

Findings suggest that bilingual English–Mandarin infants recognized the lexical relevance of tone in English and Mandarin, responding differentially to tone variants based on the language in which words were introduced. In a second experiment, Mandarin monolingual infants were tested recognition of tone-matched and tone-varying words in the same task as employed in Experiment 1 (Mandarin version). The goal of this experiment was to provide a monolingual point of comparison for findings obtained in Experiment 1. Given that bilingual infants were sensitive to tone variation when words were introduced in Mandarin, it was expected that Mandarin monolingual infants would be comparably sensitive to tone variation.

EXPERIMENT 2

We investigated Mandarin monolingual infants' sensitivity to tone changes in a similar paradigm as that used in Experiment 1. The primary methodological difference with Experiment 1 was that all participants were tested in Mandarin only. As in Experiment 1, tone changes consisted of similar and distinct contrasts.

Method

Participants

Our sample comprised 18 12- to 13-month-old Mandarin monolingual infants (age range: 12 months 11 days to 13 months 13 days, average = 12 months 24 days). All infants were full-term births with no known developmental delays or disabilities. Data from two additional infants were excluded due to failure to complete the testing session. All infants had more than 90% exposure to Mandarin as measured by the Language Exposure Questionnaire (Bosch and Sebastián-Gallés, 1997).

Stimuli

Auditory and visual stimuli for the Mandarin testing session were identical to Experiment 1 (Mandarin version).

Procedure

The experimental procedure and all other experimental parameters were identical to the Mandarin version of Experiment 1.

Results

All infants habituated within the 24 trial maximum habituation window. The number of trials to habituation and the total habituation time for each experiment is reported in **Table 2**. As in previous studies (see Fennell et al., 2007; Byers-Heinlein et al., 2013), a preliminary analysis was conducted to determine whether participants recovered to the post-test by comparing the last habituation block to the post-test stimulus. A paired samples *t*-test revealed a significant elevation in fixation times between the last habituation block and the post-test [$t(17) = 2.57, p = 0.02$].

Fixation times were logged for Same trials, Switch (similar) and Switch (distinct) trials. These values are plotted in **Figure 4**.

TABLE 2 | Summary of habituation measures.

	Trials to habituation: mean (SD)		Total habituation time (seconds): mean (SD)	
	Mandarin session	English session	Mandarin session	English session
Experiment 1	6.44 (2.66)	6.56 (2.43)	86.36 (51.53)	70.16 (47.39)
Experiment 2	6.5 (3)		79.01(53.54)	
Experiment 3	7.27 (2.86)		67.61 (30.61)	
Experiment 4	6.05 (2.34)		76.71 (46.24)	

A preliminary analysis conducted with test trial order as a between-subjects factor and trial type as a within-subjects factor revealed no effects or interactions with test trial order ($p > 0.6$). A one-way repeated-measures ANOVA with test trial as the within-subjects factor revealed no effect of trial type [$F(2,34) = 1.31, p = 0.28$]. A comparison of fixation times to Same trials as compared to each Switch trial revealed no difference in fixation times to Same versus Switch (similar) trials [$t(17) = 0.89, p = 0.39$] or between Same and Switch (distinct) trials [$t(17) = 1.69, p = 0.11$].

In comparing the results of Experiments 1 and 2, it is striking that infants with monolingual exposure to Mandarin did not differentiate tones when learning a novel word whereas those learning English and Mandarin did demonstrate sensitivity to tone when listening to Mandarin. It is possible that bilingual infants' integration of tone in Mandarin was related to having had prior exposure to the object label during the English session. If this were the case, one would predict effects of the order of testing on performance in the Mandarin session. As half of the infants underwent an English testing session first and half underwent a Mandarin testing session first, a 3×2 [test trial (same; switch (similar); switch (distinct)) \times order (English first; Mandarin first)] mixed ANOVA was conducted with fixation times to the Mandarin test trials as a dependent variable. Results revealed a main effect of trial type [$F(2,32) = 9.97, p = 0.0001, \eta_p^2 = 0.38$], no effect of order of testing [$F(1,16) = 1.2, p = 0.28$] and no interaction of test order and trial type [$F(2,32) = 0.7, p = 0.94$].

A secondary set of analyses was performed on habituation data in order to determine whether monolingual and bilingual infants were distinguished by their habituation profiles. A one-way ANOVA was conducted to compare the total time accrued during habituation and on the number of trials to habituation between Mandarin monolinguals, English–Mandarin bilinguals (Mandarin session) and English–Mandarin bilinguals (English session). There was no effect of group on total time accrued during habituation [$F(2,53) = 0.18, p = 0.84$]. Likewise, there was no effect of group on the number of trials to habituation [$F(2,53) = 0.69, p = 0.51$]. These analyses suggest that the profile of stimulus encoding did not differ across groups.

The present results suggest that Mandarin monolingual infants were not sensitive to labels for a familiarized object that had undergone a tone substitution, whether the substitution was due to a shift to a similar or distinct tone. This was

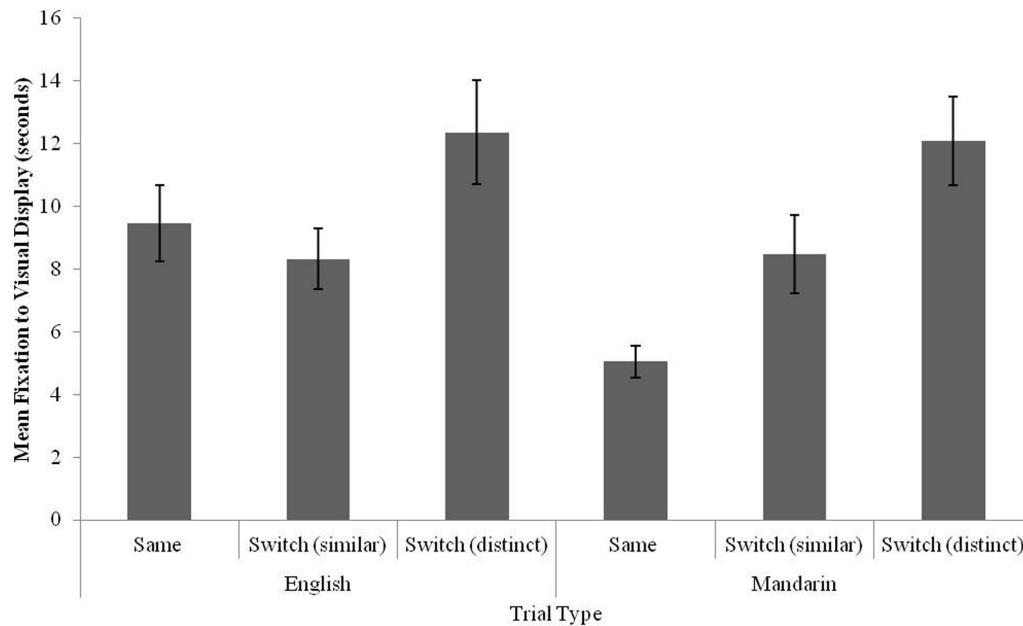


FIGURE 4 | Fixation times to the visual stimulus for Same, Switch (similar), and Switch (distinct) trials in 12–13-month-old Mandarin monolingual infants (error bars: SEM).

unexpected given findings from non-tone language learning infants demonstrating that infants at 14 and 17–18 months of age were sensitive to lexical tone distinctions when learning new words (Hay et al., 2015; Singh et al., 2015). Differences between experiments will be revisited in the Section “Discussion.” Using a different paradigm, Experiment 3 sought to determine whether Mandarin learning monolingual infants could discriminate the lexical tones that they were not able to integrate in Experiment 1. Given that tone learning infants have been shown to discriminate lexical tones at 4, 6, and 9 months of age (Mattock and Burnham, 2006; Yeung et al., 2013), it was hypothesized that Mandarin learning infants would discriminate Mandarin tones at 12–13 months.

EXPERIMENT 3

In this experiment, Mandarin monolinguals were tested on their ability to discriminate Tone 3 from Tones 1 and 2 in a phoneme discrimination paradigm. Participants were habituated to Tone 3 and then presented with an alternating string of Tone 3 and a contrastive tone (Tone 1 or 2). They were then re-exposed to Tone 3 and presented a second alternating string of Tone 3 and the other contrastive tone (Tone 1 or 2).

Method

Participants

Our sample comprised eighteen 12- to 13-month-old infants who had been monolingually exposed to Mandarin (age range: 12 months 11 days–13 months 22 days, average = 12 months 24 days). Data from two additional infants was excluded as testing

was incomplete due to fussiness. The language criteria used for this study was identical to that of Experiment 1.

Stimuli

Auditory stimuli consisted of the syllable /pa/, recorded in Mandarin Tones 1, 2, and 3. Multiple tokens were recorded, and four tokens of each tone were selected for the final stimuli. The VOT values and pitch contours for these syllables are equivalent to those described in Experiment 1. Stimuli were concatenated to form three trial types: (1) a Control trial, which featured only Tone 3 tokens, (2) an Alternating distinct tone pair trial, which had alternating tokens of Tones 1 and 3, and (3) an Alternating similar tone pair trial, which consisted of alternating tokens of Tones 2 and 3. All strings were 30 s long, and were created by repeating the stimuli systematically, with an interstimulus interval of 1 s. All strings were also paired with the visual stimulus of a stationary red-and-black checkerboard pattern presented against a white background.

Procedure

As with the previous experiments, testing was conducted in a quiet, dimly lit room, where the infant sat in their caregiver's lap, facing a computer screen. The experimenter observed the infants' responses via a CCTV system from an adjoining room. Both the experimenter and parent listened to instrumental music at a volume that masked the stimuli.

The procedure used was an adapted version of the stimulus alternating paradigm developed and previously used to assess discrimination of two contrasts within the same infant (Tyler et al., 2014; see Best and Jones, 1998; Maye et al., 2002; Mattock et al., 2008 for additional demonstrations of the

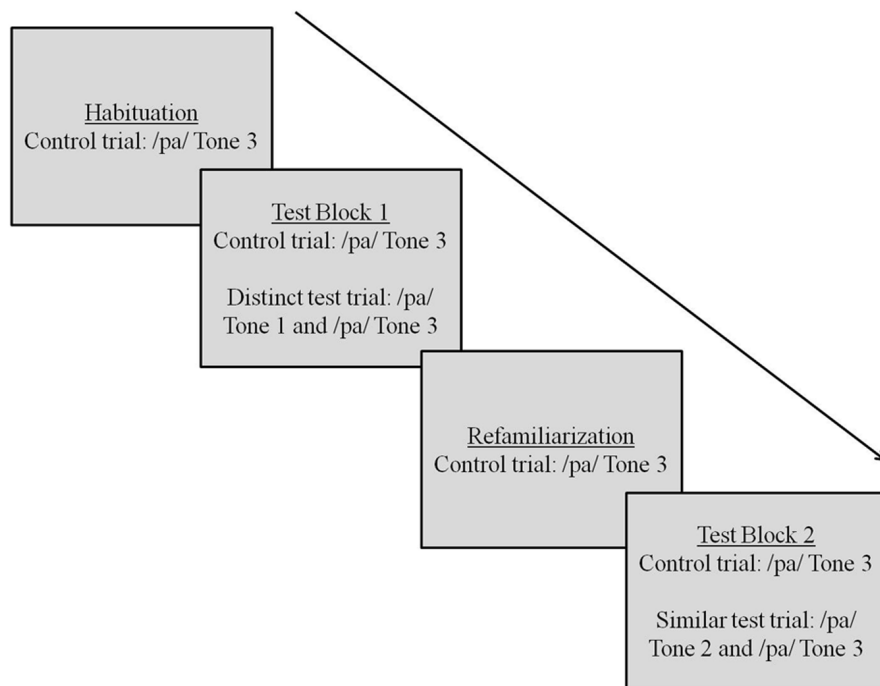


FIGURE 5 | Trial sequence for Experiment 2.

paradigm). Infants were first presented with the attention getter. At the first fixation to the visual display, the habituation phase commenced. In the habituation phase, infants were presented continuous tokens of Tone 3. Trials lasted for a maximum of 30 s, or until the infant looked away from the screen for more than 2 s. At the end of each trial, the attention getter was presented again. The habituation phase continued until the infant's looking time to the final three consecutive trials decreased to less than 50% of the total look time to the first three consecutive trials, or until the infant completed a maximum of 20 trials. This habituation criterion was informed by previous investigations of tone discrimination in Mandarin monolingual infants (see Gao et al., 2011). Once either of these criteria was met, the test phase was initiated.

The test phase consisted of three blocks. In the first test block, infants were first presented a Control trial (repetitions of Tone 3). This was followed by a Test trial, consisting of alternations of Tones 2 and 3 (similar) or of Tones 1 and 3 (distinct). Infants were then presented with three trials, each containing repetitions of Tone 3. The purpose of this phase was to reinstate Tone 3 as the basis for further comparisons with a contrastive tone. Following this, infants were presented with a second test block, comprising a Control trial (repetitions of Tone 3) and a second Alternating trial consisting of tonal alternations that had not been previously presented (either Tones 1 and 3 or Tones 2 and 3). The trial sequence for this experiment is depicted in **Figure 5**. The order of presentation of test blocks was counterbalanced across all infants, such that half the infants were presented with the distinct tone pair in the first alternating test trial and half were

presented with the similar tone pair in the first alternating test trial.

Results

All infants habituated within the 20 trial maximum habituation window. Difference scores were calculated for each infant, by subtracting the fixation times for each Control trial from the Alternating trial that followed it. Thus, infants each had two difference scores: one reflected dishabituation to the alternating trial consisting of a similar tone contrast (Tones 2 and 3) and one reflected dishabituation to the alternating trial consisting of the distinct tone contrast (Tones 1 and 3). A difference in fixation to the checkerboard display between Control and alternating blocks that deviates significantly from zero indicates that infants discriminated the Control tone from the tone presented in the alternating trial.

A 2×2 repeated measures ANOVA (Contrast: similar vs. distinct) \times 2 (Order: similar first vs. distinct first) was computed with difference scores as the dependent variables. No effects of order were found ($p > 0.3$) and thus order of presentation was excluded from subsequent analyses. To examine whether infants successfully discriminated each contrast, one-sample t -tests were used to analyze infants' difference scores in relation to baseline. This analysis revealed that infants' difference scores for the distinct contrast were significantly greater than zero, $t(17) = 3.31$, $p = 0.003$, Cohen's $d = 1.17$. Similarly, difference scores for the similar contrast ($M = 3.79$, $SD = 6.45$) were also significantly greater than zero, $t(16) = 2.44$, $p = 0.03$, Cohen's $d = 0.81$. Difference scores for these contrasts are depicted in **Figure 6**.

These results suggest that at 12–13 months, Mandarin monolinguals were sensitive to the same tone contrasts introduced in Experiment 2. Thus, while the Mandarin monolinguals successfully discriminated these contrasts, they appeared unable to integrate this information when learning new words. This conclusion should be qualified by the fact that a different paradigm was used to track auditory sensitivity to tone versus integration of tone when learning novel words. Hence, we do not conclude from this study that when presented with equivalent task demands in referential versus non-referential context, infants are sensitive to lexical tone only in the latter case. Rather, our claim is that in particular tasks known to elicit auditory sensitivity to tone contrasts, such as the Stimulus Alternating Paradigm, infants are indeed sensitive to the distinction between Tones 1 and 3 and Tones 2 and 3. Prior studies (e.g., Stager and Werker, 1997) that have tracked sensitivity to a single contrast have measured sensitivity in discrimination and word learning by using highly similar paradigms, replacing the object to be learned with a checkerboard. In our study, on account of simultaneously tracking sensitivity to two contrasts within the same participant and within a single experimental session, we opted for an equally well-established paradigm to measure phoneme discrimination. This paradigm allowed us to maintain some of the elemental components of the word-learning paradigm used in Experiment 2, specifically measurement of sensitivity to two contrasts within a single session and infant. It should also be noted that our findings from Experiment 3 are consistent with prior research using alternative discrimination paradigms that also demonstrate lexical tone discrimination in tone learning infants between 9 and 12 months of age (see Mattock and Burnham, 2006; Tsao, 2008; Yeung et al., 2013).

In light of the finding that Mandarin learning infants appeared to discriminate words based on tone, yet did not integrate these tones into newly learned words (albeit in a paradigm with different experimental parameters), Experiment 4 was designed to investigate whether older Mandarin monolingual infants

could integrate tones into newly learned words. Infants undergo significant change in their abilities to learn similar sounding words by 17–18 months (e.g., Stager and Werker, 1997) and specifically, non-tone language learning infants mature in their language-specific integration of tone between 14 and 17 months (Hay et al., 2015). It is possible that tone language learners also mature in this capacity as they approach 18 months and construe tones as a source of contrast when learning new words. It was therefore hypothesized that by 18 months, tone-learning infants would differentiate newly learned words based on tones.

EXPERIMENT 4

To determine whether older Mandarin monolinguals would be successful at detecting tone changes in a word-object association task, we tested 17- to 18-month-old Mandarin monolingual infants on the same procedure as Experiment 2.

Method

Participants

The sample comprised eighteen 17- to 18-month-old monolingual Mandarin infants (age range: 17 months 3 days to 18 months 29 days, average age = 17 months 21 days). Four additional infants were tested but excluded due to experimental error ($n = 2$), fussiness ($n = 1$) or on account of data that deviated from the group mean by more than 3 standard deviations ($n = 1$).

Stimuli

Stimuli were identical to those used in Experiment 2.

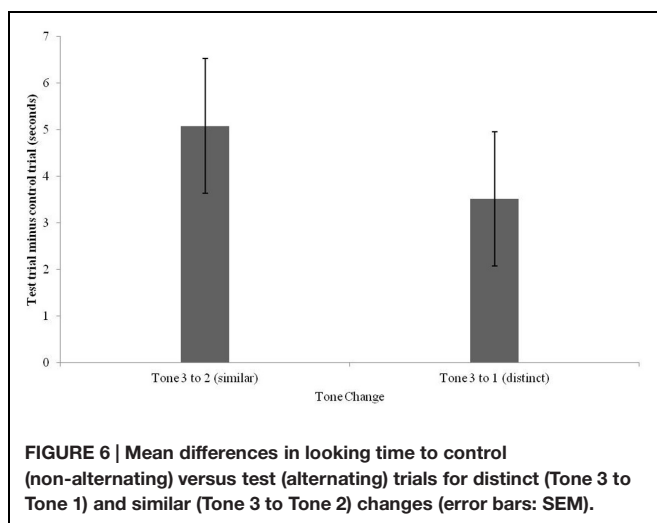
Procedure

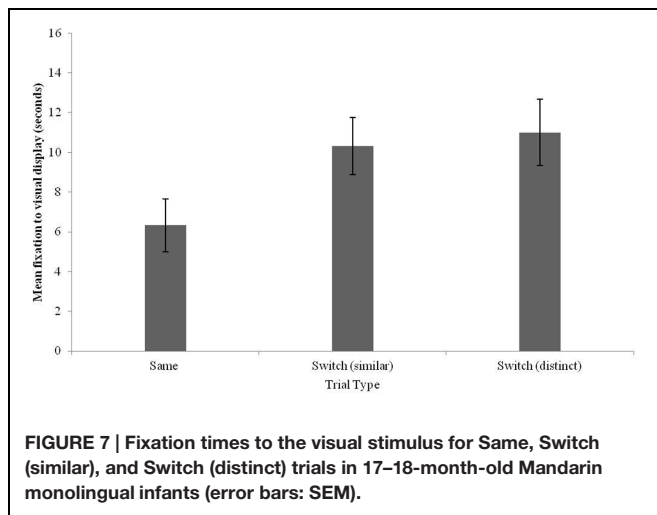
The procedure was identical to that of Experiment 2.

Results

All infants habituated within the 24 trial maximum habituation window. A comparison of the last habituation block and post-test trials revealed a significant increase in fixation to the post-test trial, $t(17) = 4.1$, $p = 0.001$. A preliminary analysis conducted with test trial order as a between-subjects factor and trial type as a within-subjects factor revealed no effects or interactions with test trial order ($p > 0.6$).

Further analyses focused on Same-Switch differences for each type of Switch trial (similar and distinct). A repeated-measures ANOVA was conducted to compare fixation times to each trial type [Same, Switch (similar) and Switch (distinct) tones], revealing a main effect of trial type [$F(2,34) = 5.63$, $p = 0.008$, $\eta_p^2 = 0.25$]. Planned contrasts revealed an increase in fixation to tone shifts that were both similar [$F(1,17) = 6.53$, $p = 0.02$, $\eta_p^2 = 0.28$] and distinct [$F(1,17) = 15.36$, $p = 0.001$, $\eta_p^2 = 0.48$]. These results are graphed in Figure 7. The results from Experiment 4 demonstrate that by 17–18 months, Mandarin learning infants are sensitive to similar and distinct tone variation when learning new words.





DISCUSSION

The present set of studies was designed to investigate the extent to which lexical tone is phonologically articulated within the bilingual and monolingual infant lexicon. Infants' sensitivity to lexical tone was examined across four experiments. Experiment 1 investigated bilingual English–Mandarin infants' sensitivity to lexical tone variation in each of their native languages. Infants exhibited language-selective integration of lexical tone at this stage, contrasting newly learned words by tone variation in a Mandarin context and disregarding tone variation in an English context even though they were tested in each language in immediate succession. In this experiment, there were effects of perceptual salience of the tone contrast on infants' sensitivity to tone variation in Mandarin. However, these effects were secondary in that they did not eclipse infants' overall recognition of the lexical functions fulfilled by pitch in Mandarin. In Experiment 2, we investigated 12- to 13-month-old monolingual Mandarin learning infants' abilities to integrate lexical tone into memories of newly learned words. Infants demonstrated a relative insensitivity to tone variation, irrespective of whether the variation was introduced by a similar or distinct mispronunciation. In Experiment 3, Mandarin monolingual infants were tested on their ability to discriminate the lexical tones presented in Experiment 1, revealing that both similar and distinct tone pairings were robustly discriminated in a habituation paradigm between 12 and 13 months. Finally, in Experiment 4, Mandarin learning monolingual infants were tested on the same paradigm as Experiment 2 at an older age (17- to 18-months), demonstrating an ability to integrate lexical tone variation into newly learned words and to detect similar and distinct mispronunciations in equal measure.

Previous investigations of infants' abilities to learn similar sounding words have focused on their sensitivity to segmental detail, most notably, to the onset consonant of a word (e.g., "bih" versus "dih") (e.g., Stager and Werker, 1997; Pater et al., 2004; Fennell et al., 2007; Fennell and Byers-Heinlein, 2014,

but see Curtin et al., 2009). In the aggregate, these findings suggest that monolingual infants are not able to learn similar sounding words differing by onset consonant until 17 months (Stager and Werker, 1997, but see MacKenzie et al., 2011), although this ability has been shown to emerge at 14 months when infants received contextual support (Fennell and Waxman, 2010). Our findings with Mandarin monolingual infants suggest that even with contextual support (i.e., naming phrases) infants were not able to map tonal variants onto different objects at 12–13 months and were only able to do so at 17–18 months. Given that the majority of prior studies investigating mastery of similar sounding words has been conducted with infants 14 months of age and older, it is difficult to compare the course of acquisition of segmental contrasts versus tone contrasts based on the present study. In contrast to Mandarin monolingual infants, the most surprising finding to emerge from the current set of studies is that bilingual infants demonstrated precocity in their ability to integrate tone variation in a language-selective manner as early as 12–13 months. Unlike monolingual infants, they were able to integrate variation in lexical tone in a Mandarin context. Within the same laboratory session, when presented with a new word-object pairing in English naming phrases, they were able to disregard the same sources of variation when tested in English. This finding is somewhat unexpected given the task demands faced by bilinguals in this study whereby they would have had to inhibit the phonological rules of one of their native languages in each task. The experiment was designed such that the phonetic properties of the target words remained the same across languages, suggesting that context alone may have enabled a language-specific integration of tone. In prior research, bilingual and monolingual infants have been shown to be similar to one another – assuming they receive input commensurate with their language environment – in learning similar sounding words with no clear evidence of a bilingual advantage (Mattock et al., 2010; Byers-Heinlein and Fennell, 2014). However, our study deviates from prior studies in this area in that previous research has focused exclusively on how bilinguals negotiate sound contrasts that distinguish meaning on both of their languages. In contrast, the current study investigates sensitivity to a source of phonological variation that categorically conflicts across languages (i.e., it is phonemic in one and non-phonemic in the other). Three possible reasons for a bilingual advantage in this task are discussed in turn.

First, it should be noted that tone does not only introduce phonological conflict for bilingual learners. Monolingual Mandarin learners also confront potential conflict within their native language on account of tone. Pitch movements drive lexical changes in tone but they also drive changes in intonation that are non-lexical in Mandarin. A learner of Mandarin therefore has to selectively integrate pitch variation that corresponds to lexical tone categories when learning new words and to disregard that which distinguishes intonational contrast when defining words. The challenge inherent in this duality is evidenced by findings that even adult speakers of Mandarin Chinese are sensitive to tone-intonation conflict in native sentence processing (e.g., Yuan and Shih, 2004).

Therefore, tone introduces intrinsic conflict for monolingual Mandarin learners as well as for bilingual learners. It is possible that bilingual infants are better able to negotiate this conflict on account of collateral cognitive changes that are thought to arise from bilingual experience. This possibility derives from a broad swath of research demonstrating a bilingual advantage in negotiating conflicting information both in linguistic and non-linguistic tasks (e.g., Bialystok et al., 2004, 2008; Costa et al., 2008; Kovács, 2009). The presence of conflict in the task may have harnessed bilinguals' extant advantages for cognitive control in the face of conflict, an advantage apparent in infancy (Kovács and Mehler, 2009a). As such, it is possible that cognitive control advantages conferred upon the bilingual infant permeate early language processing, aiding in the de-activation of the phonological structure of the one language when processing the other.

An explanation predicated on a bilingual advantage in conflict resolution presupposes that the advantage demonstrated in word learning is secondary to a general cognitive advantage to emerge from bilingual exposure. However, a second possibility is that the bilingual advantage observed herein is specific to language. Prior studies with bilingual children and adults have revealed a bilingual advantage in mastering the rules of the native languages, often characterized as a metalinguistic benefit of bilingualism (e.g., Bialystok, 1988; Bialystok et al., 2003). Although these studies have focused largely on mastery of the grammatical systems of each language, metalinguistic advantages appear to transcend grammatical knowledge and extend to mastery of the sound system (Campbell and Sais, 1995). A mechanism commonly advanced for why bilingualism may promote metalinguistic awareness may provide a second potential explanation for our findings. The mere presence of conflict – or structural differences – across languages may highlight relevant properties of each language to bilingual learners (Bialystok and Hakuta, 1994; Friesen and Bialystok, 2012). Although rhetorically, researchers have appealed to cross-language conflict as a basis for metalinguistic advantages (see Bialystok and Hakuta, 1994), tests of metalinguistic awareness in bilinguals have not generally measured sensitivities to linguistic cues that functionally conflict across the two languages of a bilingual. The normative approach has been to measure sensitivity to the rules of one language (see Bialystok, 2001, for a review). The current study suggests that mastering properties of languages that conflict, which intuitively should be more complex to negotiate, may be consolidated earlier in bilinguals. It is therefore possible that the precocity observed among bilingual infants in the present study derives directly from experience with conflicting linguistic rules. In other words, noticing that pitch cues effect referential change in one language but not in the other may facilitate an awareness of pitch as a relevant – and contrastive – feature of language to young learners.

Finally, the advantage observed in bilingual infants may derive from a specific sensitivity to pitch. Prior research demonstrates that bilinguals are more sensitive to prosody and more generally, to the encoding of pitch in comparison to monolinguals in both infancy and adulthood (Krizman et al., 2012; Gervain and Werker, 2013). In comparison with monolingual infants,

bilingual infants more readily incorporate pitch movements as a cue to linguistic structure even if they are not learning a tone language (Gervain and Werker, 2013). It is possible that the bilingual advantage observed in the present study may be limited to the specific source of variation contained within this study – vocal pitch. Further research could test this hypothesis by investigating sensitivity to segmental phonological conflict across languages in monolingual and bilingual learners.

In addition to demonstrating bilingual infants' facility with negotiating phonological conflict, a second contribution of the present study is to chart tone sensitivity in native learners of a tone language. From our findings, it appears that native tone language learners do not incorporate tone into newly learned words until 18 months. At 12–13 months, Mandarin learning infants appear insensitive to tone variation in newly learned words, an effect that does not reflect a limitation in discriminating the tone pairs used in this study but rather a specific limitation in integrating tones into novel word-object mappings. A disconnect between the capacity for auditory discrimination of native contrasts and integration of these contrasts into names for objects has been reported with regards to consonant variation (see Stager and Werker, 1997). However, this disconnect, often termed the word learning 'paradox,' is often alleviated when words are embedded in naming phrases that highlight the referential nature of the task at 14 months (Fennell and Waxman, 2010). In the present study, however, even when supported with naming phrases, 12- to 13-month-old monolingual Mandarin learning infants were not able to integrate tone variation into newly learned words. It is possible that the ability to profit from naming phrases develops closer to 14 months and was therefore not captured within the time frame under investigation in the present study. However, it is also possible that tone variation effected by pitch movements is more challenging to bind to the lexicon than segmental variation. Pitch serves a broad range of functions in all languages and tone languages are no exception. In Mandarin Chinese, pitch cues make important non-lexical distinctions, such as distinguishing questions versus statements (Yuan et al., 2002), contrastive prosodic stress (Xu, 1999), as well as contrasting vocal emotions (Li et al., 2011). The functional differentiation of pitch may be a complex process for tone language learner and this complexity may prolong the process of assigning distinct communicative functions to pitch variation. One source of support for this comes from prior developmental research demonstrating that pitch cues to tone and intonation are only robustly dissociated as late as 4–5 years of age in Mandarin learning children (Singh and Chee, 2016). Although bilingual infants contend with the same complexity with regards to pitch, or arguably even more, enhancements in cognitive control and/or metalinguistic awareness and/or enhanced pitch sensitivities may offset the effects of this complexity. Moreover, the mere presence of conflict across languages, often thought to underlie bilingual advantages in metalinguistic awareness, may facilitate phonological integration in bilingual infants.

The finding that Mandarin learning infants did not incorporate lexical tone into newly learned words at 12–13 months is somewhat surprising in light of prior studies demonstrating that other populations associate newly learned

words with tones. Integration of tones in non-tone language learning infants was evidenced at 14 and 18 months (Graf Estes and Hay, 2015; Hay et al., 2015; Singh et al., 2015) and in Mandarin–English bilinguals at 18 months (Singh et al., 2015), although it should be noted that none of these studies sampled Mandarin monolingual infants. Four possible explanations are offered for why Mandarin monolingual infants may have exhibited a different response to other language groups, such as English monolingual infants. First, as mentioned earlier, it is possible that the functional differentiation of pitch for a Mandarin learner is associated with a more complex learning pathway on account of the multiplexing of pitch in tone languages (e.g., pitch is used to contrast emotions, stress, communicative intent, and lexis). What appears to be a monolingual delay may be traced to monolingual learners gradually ‘distilling’ vocal pitch into its many communicative functions. The complexity of this process in tone languages may temporarily disfavor tone language learners. For non-tone language (e.g., English monolingual) learners, the division of labor carried by pitch is arguably more categorical: suprasegmental variation is more tightly bound to non-lexical functions and lexical contrast is marked by segmental variation. For Mandarin monolingual learners, the functions of suprasegmental variation are distributed over lexical and non-lexical functions, which may present a greater learning burden. So then why do bilingual learners of Mandarin and English not demonstrate effects of this burden? As discussed above, the presence of phonological conflict combined with a bilingual advantage for negotiating conflict may confer upon bilingual Mandarin–English learners early advantages less available to monolingual infants. This possibility is consistent with the bilingual advantage observed herein, but merits further empirical study. A second possibility derives from stimulus-specific effects. Each of the prior studies documenting tone integration in non-tone language learners (Graf Estes and Hay, 2015; Hay et al., 2015; Singh et al., 2015) used rising and falling tone contrasts (corresponding to Tones 2 and 4). These tones correspond closely to salient intonational categories in English and Mandarin, specifically, to the question/statement contrast (Singh and Chee, 2016). Young infants learning non-tone languages are astutely sensitive to the question/statement distinction (Geffen and Mintz, 2011; Frota et al., 2014), which serves an important pragmatic function in English as well as in Mandarin (Yuan, 2004, 2006). It is possible that these tone contrasts are integrated into lexical representations on account of their weighty pragmatic significance. One might expect tone contrasts that do not map directly onto intonational categories (such as those used in the present study) to be less salient to infants. It is possible that prior studies demonstrating tone integration in English learning infants engaged an extant sensitivity to intonational contrast, specifically, to the question/statement contrast. Sensitivity to this contrast in English learners may emerge earlier and may be more potent than sensitivity to native tones in Mandarin learners, although this account awaits empirical support. Third, it should be noted that Tone 3 is the most complex Mandarin tone on account of its bi-directionality (Gandour, 1983). It is acquired late relative to other Mandarin tones (Li and Thompson, 1977) and involves

relatively complex laryngeal coordination (Wong, 2012). Tone 3 is also invoked in a common phonological alternation (Tone 3 Sandhi) resulting in context-driven substitutions to Tone 2. On account of these factors, the representation of Tone 3 in young learners may indeed be more fragile than that of other tones. Our design was predicated on infants having a well-specified representation of Tone 3 in order to detect deviations to Tone 1 and 2. Although speculative, further studies could examine stimulus-specific effects by using a different tone as the point of comparison and by exploring whether effects observed herein are symmetrical (i.e., whether a change from Tone 2 to Tone 3 would be more accurately detected at 12–13 months based on the possibility that Tone 2 sensitivity may profit from greater representational strength). A fourth possibility that is worth noting is that tone sensitivity may actually change between 12 and 14 months of age, a transition documented by Liu and Kager (2014). Liu and Kager (2014) observed that 11–12 months represented a comparative ‘low point’ in terms of infant tone sensitivity, which then progressively increased by 14–15 months. While their study was conducted with Dutch monolingual infants, it is conceivable that this trajectory may generalize to tone language learners. Although beyond the scope of the current paper, a replication of the current study at 14 months may allow for more direct comparisons between the present and previous studies.

Our primary purpose in conducting this study was to investigate bilingual infants’ negotiation of tone as a source of phonological conflict. Currently, there is mounting public interest in the science of bilingualism, perhaps inspired by the ever increasing numbers of children raised in bilingual environments (Peña and Bedore, 2010). However, parents and educators often wonder about the developmental effects of early bilingual exposure and specifically whether early exposure to two languages has the potential to confuse a young baby and consequently, to delay language development. These questions have garnered considerable popular and scientific attention. A recent suite of studies has demonstrated that infants may benefit from early exposure to two languages in a range of cognitive domains: learning sequences of information, imitation, anticipating events, visual habituation, and visual recognition memory (Kovács and Mehler, 2009a,b; Brito and Barr, 2014; Singh et al., 2014). However, an open question exists as to whether early bilingual exposure influences the uptake of each language. Previous research comparing monolinguals and bilinguals on the uptake of the formal properties of each language has focused predominantly on vocabulary size. These studies have suggested that single language vocabulary size is sometimes reduced in bilingual versus monolingual children (Bialystok and Feng, 2011; Hoff et al., 2012), although when measured across both languages, vocabulary size estimates can match or even surpass that of monolingual peers (e.g., Pearson et al., 1993; De Houwer et al., 2013). The current study adds to an ongoing narrative on whether two languages facilitate or confound the language-learning journey and suggests that an elemental formal property of bilingual development, acquisition of the native phonological systems, may benefit from bilingual exposure. Moreover, such advantages may be evident prior to the onset of a substantial

productive vocabulary. Although prior studies have revealed bilingual advantages in learning the structure of languages, these studies have not typically assessed sensitivity to a property of language that causes cross-language conflict (e.g., Galambos and Goldin-Meadow, 1990; Campbell and Sais, 1995; Bialystok et al., 2014). Discursively, however, researchers have suggested that it may indeed be the presence of conflict that drives mastery of two systems, alluding to a direct relationship between incongruent language systems and gains in learning (Bialystok and Hakuta, 1994). This viewpoint is perhaps most famously exemplified by the now widely popularized statement by Bialystok and Hakuta that “it is precisely because the structures and concepts of different languages never coincide that the experience of learning a second language is so spectacular in its effects.” Providing one line of argument in support of this view, our findings invite the possibility that in some domains of bilingual development, cross-language conflict may not serve to confuse, but instead, to clarify.

CONCLUSION

The title of this paper alludes to prior research positing ‘limits on bilingualism’ (Cutler et al., 1989; Dupoux et al., 2010). The postulate that there are limits on bilingualism is predicated on the notion that bilingual learners may never attain the degree of single-language proficiency exhibited by native monolingual speakers of the same two languages. In contrast to this hypothesis, the present study proposes that the early establishment of the phonological lexicon may be fortified by bilingual exposure. In contrast to bilingual

infants, monolingually tone-exposed infants may follow a more protracted time course in determining the relationships between words and tones. Accordingly, mastery of two conflicting systems may potentially consolidate knowledge of the properties of each language, favoring phonological development in bilingual learners.

AUTHOR CONTRIBUTIONS

LS conceptualized the study, conducted data analyses, and drafted the manuscript. FP collected data for the study and drafted portions of the manuscript. CF collected data for the study, conducted data analyses, and drafted portions of the manuscript.

FUNDING

This research was supported by a grant from the National University of Singapore, Singapore to LS (HSS R-581-000-178-646), to a Ministry of Education Tier 1 Academic Research Fund (FY2013-FRC2-009) grant to LS, and to a grant from the Singapore Children’s Society to CF.

ACKNOWLEDGMENTS

We are grateful to Aloysia Tan, Desirene Poon, and Xian Hui Seet for assistance for assistance with stimulus recordings and to Cathi Best, Janet Werker, and Chris Fennell for methodological guidance.

REFERENCES

- Best, C., and Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behav. Dev.* 21, 295. doi: 10.1016/S0163-6383(98)91508-9
- Bialystok, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Dev. Psychol.* 24, 560–567. doi: 10.1037/0012-1649.24.4.560
- Bialystok, E. (2001). Metalinguistic aspects of bilingual processing. *Annu. Rev. Appl. Linguist.* 21, 169–181. doi: 10.1017/S0267190501000101
- Bialystok, E., Craik, F., Klein, R., and Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychol. Aging* 19, 290–303. doi: 10.1037/0882-7974.19.2.290
- Bialystok, E., Craik, F., and Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 859–873. doi: 10.1037/0278-7393.34.4.859
- Bialystok, E., and Feng, X. (2011). “Language proficiency and its implications for monolingual and bilingual children,” in *Language and Literacy Development in Bilingual Settings*, eds A. Y. Durgunoglu and C. Goldenberg (New York, NY: Guilford Press), 121–138.
- Bialystok, E., and Hakuta, K. (1994). *In Other Words: The Science and Psychology of Second Language Acquisition*. New York, NY: Basic Books.
- Bialystok, E., Majumder, S., and Martin, M. M. (2003). Developing phonological awareness: is there a bilingual advantage? *Appl. Psycholinguist.* 24, 27–44. doi: 10.1017/S014271640300002X
- Bialystok, E., Peets, K. F., and Moreno, S. (2014). Producing bilinguals through immersion education: development of metalinguistic awareness. *Appl. Psycholinguist.* 35, 177–191. doi: 10.1017/S0142716412000288
- Bosch, L., and Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition* 65, 33–69. doi: 10.1016/S0010-0277(97)00040-1
- Brito, N., and Barr, R. (2014). Flexible memory retrieval in bilingual 6-month-old infants. *Dev. Psychobiol.* 56, 1156–1163. doi: 10.1002/dev.21188
- Byers-Heinlein, K., and Fennell, C. T. (2014). Perceptual narrowing in the context of increased variation: insights from bilingual infants. *Dev. Psychobiol.* 56, 274–291. doi: 10.1002/dev.21167
- Byers-Heinlein, K., Fennell, C. T., and Werker, J. F. (2013). The development of associative word learning in monolingual and bilingual infants. *Biling. Lang. Cogn.* 16, 198–205. doi: 10.1017/S1366728912000417
- Campbell, R., and Sais, E. (1995). Accelerated metalinguistic (phonological) awareness in bilingual children. *Br. J. Dev. Psychol.* 13, 61–68. doi: 10.1111/j.2044-835X.1995.tb00664.x
- Chandrasekaran, B., Sampath, P. D., and Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *J. Acoust. Soc. Am.* 128, 456–465. doi: 10.1121/1.3445785
- Chao, K.-Y., Khattab, G., and Chen, L.-M. (2006). “Comparison of VOT patterns in Mandarin Chinese and in English,” in *Proceedings of the 4th Annual Hawaii International Conference on Arts and Humanities*, Vol. 840–859, Honolulu, HI.
- Chen, A. (2013). *Universal Biases in the Perception of Mandarin Tones: From Infancy to Adulthood*. Ph.D. thesis, Utrecht University, Utrecht.
- Chen, L., Chao, K., and Peng, J.-F. (2007). “VOT productions of word-initial stops in Mandarin and English: a cross-language study,” in *Proceedings of the 19th Conference on Computational Linguistics and Speech Processing*, Taipei, 303–317.
- Costa, A., Hernández, M., and Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: evidence from the ANT task. *Cognition* 106, 59–86. doi: 10.1016/j.cognition.2006.12.013

- Curtin, S. (2009). Twelve-month-olds learn novel word-object pairings differing only in stress pattern. *J. Child Lang.* 36, 1157–1165. doi: 10.1017/S0305000909009428
- Curtin, S. (2010). Young infants encode lexical stress in newly encountered words. *J. Exp. Child Psychol.* 105, 376–385. doi: 10.1016/j.jecp.2009.12.004
- Curtin, S., Fennell, C., and Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Dev. Sci.* 12, 725–731. doi: 10.1111/j.1467-7687.2009.00814.x
- Cutler, A., Mehler, J., Norris, D., and Segui, J. (1989). Limits on bilingualism. *Nature* 340, 229–230. doi: 10.1038/340229a0
- De Houwer, A., Bornstein, M. H., and Putnick, D. L. (2013). A bilingual-monolingual comparison of young children's vocabulary size: evidence from comprehension and production. *Appl. Psycholinguist.* 35, 1189–1211. doi: 10.1017/S0142716412000744
- Deterding, D., and Nolan, F. (2007). "Aspiration and voicing of Chinese and English plosives," in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 385–388.
- Dietrich, C., Swingle, D., and Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16027–16031. doi: 10.1073/pnas.0705270104
- Dupoux, E., Peperkamp, S., and Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: stress "deafness" in simultaneous French-Spanish bilinguals. *Cognition* 114, 266–275. doi: 10.1016/j.cognition.2009.10.001
- Escudero, P., Best, C. T., Kitamura, C., and Mulak, K. E. (2014). Magnitude of phonetic distinction predicts success at early word learning in native and non-native accents. *Front. Psychol.* 5:1059. doi: 10.3389/fpsyg.2014.01059
- Fennell, C. T. (2012). Object familiarity enhances infants' use of phonetic detail in novel words. *Infancy* 17, 339–353. doi: 10.1111/j.1532-7078.2011.00080.x
- Fennell, C. T., and Byers-Heinlein, K. (2011). Sentential context improves bilingual infants' use of phonetic detail in novel words. *Conf. Lang. Dev.* 178, 189.
- Fennell, C. T., and Byers-Heinlein, K. (2014). You sound like Mommy Bilingual and monolingual infants learn words best from speakers typical of their language environments. *Int. J. Behav. Dev.* 38, 309–316. doi: 10.1177/0165025414530631
- Fennell, C. T., Byers-Heinlein, K., and Werker, J. F. (2007). Using speech sounds to guide word learning: the case of bilingual infants. *Child Dev.* 78, 1510–1525. doi: 10.1111/j.1467-8624.2007.01080.x
- Fennell, C. T., and Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Dev.* 81, 1376–1383. doi: 10.1111/j.1467-8624.2010.01479.x
- Fennell, C. T., and Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Lang. Speech* 46, 245–264. doi: 10.1177/00238309030460020901
- Flege, J. E., and Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Commun.* 6, 185–202. doi: 10.1016/0167-6393(87)90025-2
- Friesen, D. C., and Bialystok, E. (2012). Metalinguistic Ability in Bilingual Children: the role of executive control. *Riv. Psicolinguist. Appl.* 12, 47–56.
- Frota, S., Butler, J., and Vigário, M. (2014). Infants' perception of intonation: is it a statement or a question? *Infancy* 19, 194–213. doi: 10.1111/inf.12037
- Galambos, S. J., and Goldin-Meadow, S. (1990). The effects of learning two languages on levels of metalinguistic awareness. *Cognition* 34, 1–56. doi: 10.1016/0010-0277(90)90030-N
- Gandour, J. T. (1983). Tone perception in far eastern-languages. *J. Phon.* 11, 147–175.
- Gao, J., Shi, R., and Li, A. (2011). "Lexical tone representations of mandarin-speaking toddlers," in *Poster Presented at the International Congress of Phonetic Sciences*, Hong Kong.
- García-Sierra, A., Diehl, R. L., and Champlin, C. (2009). Testing the double phonemic boundary in bilinguals. *Speech Commun.* 51, 369–378. doi: 10.1016/j.specom.2008.11.005
- García-Sierra, A., Ramírez-Esparza, N., Silva-Pereyra, J., Siard, J., and Champlin, C. A. (2012). Assessing the double phonemic representation in bilingual speakers of Spanish and English: an electrophysiological study. *Brain Lang.* 121, 194–205. doi: 10.1016/j.bandl.2012.03.008
- Geffen, S., and Mintz, T. (2011). "Seven-month-olds discrimination of statements and questions," in *Proceedings of the 36th Boston University Conference on Language Development* (Boston, MA: Boston University).
- Gervain, J., and Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nat. Commun.* 4, 1490. doi: 10.1038/ncomms2430
- Gonzales, K., and Lotto, A. J. (2013). A bafri, un pafri Bilinguals' pseudoword identifications support language-specific phonetic systems. *Psychol. Sci.* 24, 2135–2142. doi: 10.1177/0956797613486485
- Graf Estes, K., and Hay, J. F. (2015). Flexibility in bilingual infants' word learning. *Child Dev.* 86, 1371–1385. doi: 10.1111/cdev.12392
- Hay, J. F., Graf Estes, K., Wang, T., and Saffran, J. R. (2015). From flexibility to constraint: the contrastive use of lexical tone in early word learning. *Child Dev.* 86, 10–22. doi: 10.1111/cdev.12269
- Hazan, V. L., and Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Lang. Speech* 36, 17–38. doi: 10.1177/002383099303600102
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., and Parra, M. (2012). Dual language exposure and early bilingual development. *J. Child Lang.* 39, 1–27. doi: 10.1017/S0305000910000759
- Jusczyk, P. W., and Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cogn. Psychol.* 29, 1–23. doi: 10.1006/cogp.1995.1010
- Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Dev. Sci.* 12, 48–54. doi: 10.1111/j.1467-7687.2008.00742.x
- Kovács, Á. M., and Mehler, J. (2009a). Cognitive gains in 7-month-old bilingual infants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6556–6560. doi: 10.1073/pnas.0811323106
- Kovács, Á. M., and Mehler, J. (2009b). Flexible learning of multiple speech structures in bilingual infants. *Science* 325, 611–612. doi: 10.1126/science.1173947
- Krizman, J., Marian, V., Shook, A., Skoe, E., and Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7877–7881. doi: 10.1073/pnas.1201575109
- Li, A., Fang, Q., and Dang, J. (2011). "Emotional intonation in a tone language: experimental evidence from Chinese," in *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 1198–1201.
- Li, C. N., and Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *J. Child Lang.* 4, 185–199. doi: 10.1017/S0305000900001598
- Liao, S. J. (2005). *Interlanguage Production of English Stop Consonants: A VOT Analysis*. M.A. thesis, National Kaohsiung Normal University, Kaohsiung.
- Lisker, L., and Abramson, A. (1964). Cross-language study of voicing in initial stops: acoustical measurements. *Word* 20, 384–422. doi: 10.1080/00437956.1964.11659830
- Lisker, L., and Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Lang. Speech* 10, 1–28.
- Liu, L., and Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition* 133, 385–394. doi: 10.1016/j.cognition.2014.06.004
- MacKenzie, H., Graham, S. A., and Curtin, S. (2011). Twelve-month-olds privilege words over other linguistic sounds in an associative learning task. *Dev. Sci.* 14, 249–255. doi: 10.1111/j.1467-7687.2010.00975.x
- Mattock, K., and Burnham, D. (2006). Chinese and English infants' tone perception: evidence for perceptual reorganization. *Infancy* 10, 241–265. doi: 10.1207/s15327078in1003_3
- Mattock, K., Molnar, M., Polka, L., and Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition* 106, 1367–1381. doi: 10.1016/j.cognition.2007.07.002
- Mattock, K., Polka, L., Rvachew, S., and Krehm, M. (2010). The first steps in word learning are easier when the shoes fit: comparing monolingual and bilingual infants. *Dev. Sci.* 13, 229–243. doi: 10.1111/j.1467-7687.2009.00891.x
- Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111. doi: 10.1016/S0010-0277(01)00157-3
- Pater, J., Stager, C., and Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language* 80, 384–402. doi: 10.1353/lan.2004.0141
- Pearson, B. Z., Fernández, S. C., and Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Lang. Learn.* 43, 93–120. doi: 10.1111/j.1467-1770.1993.tb00174.x

- Pegg, J. E., and Werker, J. F. (1997). Adult and infant perception of two English phones. *J. Acoust. Soc. Am.* 102, 3742–3753. doi: 10.1121/1.420137
- Peña, E. D., and Bedore, L. M. (2010). “Bilingualism in child language disorders,” in *Handbook of Child Language Disorders*, ed. R. G. Schwartz (New York: Psychology Press), 280–307.
- Peterson, G. E., and Barney, H. L. (1952). Control methods used in the study of vowels. *J. Acoust. Soc. Am.* 24, 175–184. doi: 10.1121/1.1906875
- Rost, G. C., and McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Dev. Sci.* 12, 339–349. doi: 10.1111/j.1467-7687.2008.00786.x
- Rost, G. C., and McMurray, B. (2010). Finding the signal by adding noise: the role of noncontrastive phonetic variability in early word learning. *Infancy* 15, 608–635. doi: 10.1111/j.1532-7078.2010.00033.x
- Shen, X. S., and Lin, M. (1991). A perceptual study of mandarin tones 2 and 3. *Lang. Speech* 34, 145–156. doi: 10.1177/002383099103400202
- Singh, L., and Chee, M. (2016). Rise and fall: effects of tone and intonation on spoken word recognition in early childhood. *J. Phon.* 55, 109–118. doi: 10.1016/j.wocn.2015.12.005
- Singh, L., and Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition* 124, 128–142. doi: 10.1016/j.cognition.2012.05.008
- Singh, L., and Fu, C. S. (2016). A new view of language development: the acquisition of lexical tone. *Child Dev.* doi: 10.1111/cdev.12512 [Epub ahead of print].
- Singh, L., Fu, C. S. L., Rahman, A. A., Hameed, W. B., Sanmugam, S., Agarwal, P., et al. (2015). Back to basics: a bilingual advantage in infant visual habituation. *Child Dev.* 86, 294–302. doi: 10.1111/cdev.12271
- Singh, L., and Quam, C. M. (2016). Can bilingual children turn one language off? Evidence from perceptual switching. *J. Exp. Child Psychol.* 147, 111–125. doi: 10.1016/j.jecp.2016.03.006
- Singh, L., Tam, H. J., Chan, C., and Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: a cross-linguistic investigation. *Dev. Sci.* 17, 94–109. doi: 10.1111/desc.12097
- Stager, C. L., and Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388, 381–382. doi: 10.1038/41102
- Thiessen, E. D. (2007). The effect of distributional information on children’s use of phonemic contrasts. *J. Mem. Lang.* 56, 16–34. doi: 10.1016/j.jml.2006.07.002
- Tsao, F.-M. (2008). The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *Chin. J. Psychol.* 50, 111–124.
- Tyler, M. D., Best, C. T., Goldstein, L. M., and Antoniou, M. (2014). Investigating the role of articulatory organs and perceptual assimilation of native and non-native fricative place contrasts. *Dev. Psychobiol.* 56, 210–227. doi: 10.1002/dev.21195
- Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., and Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Dev. Psychol.* 34, 1289–1309. doi: 10.1037/0012-1649.34.6.1289
- Werker, J. F., Fennell, C. T., Corcoran, K. M., and Stager, C. L. (2002). Infants’ ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy* 3, 1–30. doi: 10.1207/S15327078IN0301_1
- Wong, P. (2012). Acoustic characteristics of three-year-olds’ correct and incorrect monosyllabic Mandarin lexical tone productions. *J. Phon.* 40, 141–151. doi: 10.1016/j.wocn.2011.10.005
- Wong, P., Schwartz, R. G., and Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *J. Speech Lang. Hear. Res.* 48, 1065–1079. doi: 10.1044/1092-4388(2005/074)
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *J. Phon.* 27, 55–105. doi: 10.1006/jpho.1999.0086
- Yeung, H. H., Chen, K. H., and Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *J. Mem. Lang.* 68, 123–139. doi: 10.1016/j.jml.2012.09.004
- Yoshida, K. A., Fennell, C. T., Swingle, D., and Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Dev. Sci.* 12, 412–418. doi: 10.1111/j.1467-7687.2008.00789.x
- Yuan, J. (2004). “Perception of Mandarin intonation,” in *Proceedings of the International Symposium on Chinese Spoken Language Processing 2004* (Rome: IEEE), 45–48. doi: 10.1109/CHINSL.2004.1409582
- Yuan, J. (2006). “Mechanisms of question intonation in Mandarin,” in *Chinese Spoken Language Processing*, eds Q. Huo, B. Ma, E.-S. Chng, and H. Li (Heidelberg: Springer Berlin), 19–30.
- Yuan, J., and Shih, C. (2004). “Confusability of Chinese intonation,” in *Proceedings of the Speech Prosody 2004*, Nara.
- Yuan, J., Shih, C., and Kochanski, G. P. (2002). “Comparison of declarative and interrogative intonation in Chinese,” in *Proceedings of the Speech Prosody 2002*, Aix-en-Provence, 711–714.
- Zee, E., and Lee, W. S. (2001). “An acoustical analysis of the vowels in Beijing Mandarin,” in *Proceedings of the 7th European Conference on Speech Communication and Technology, 2nd INTERSPEECH Event*, Aalborg.
- Zue, V. W. (1976). Some perceptual experiments on the Mandarin tones. *J. Acoust. Soc. Am.* 60, S45–S45. doi: 10.1121/1.2003354

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Singh, Poh and Fu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Differences in the Association between Segment and Language: Early Bilinguals Pattern with Monolinguals and Are Less Accurate than Late Bilinguals

Cynthia P. Blanco^{1*}, Colin Bannard² and Rajka Smiljanic¹

¹ Department of Linguistics, University of Texas at Austin, Austin, TX, USA, ² Department of Psychological Sciences, University of Liverpool, Liverpool, UK

OPEN ACCESS

Edited by:

Annie Tremblay,
University of Kansas, USA

Reviewed by:

Elizabeth A. McCullough,
University of Washington, USA
Alison Trude,
Johns Hopkins University, USA

*Correspondence:

Cynthia P. Blanco
cynthiapblanco@gmail.com

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 02 February 2016

Accepted: 16 June 2016

Published: 29 June 2016

Citation:

Blanco CP, Bannard C and
Smiljanic R (2016) Differences
in the Association between Segment
and Language: Early Bilinguals
Pattern with Monolinguals and Are
Less Accurate than Late Bilinguals.
Front. Psychol. 7:993.
doi: 10.3389/fpsyg.2016.00993

Early bilinguals often show as much sensitivity to L2-specific contrasts as monolingual speakers of the L2, but most work on cross-language speech perception has focused on isolated segments, and typically only on neighboring vowels or stop contrasts. In tasks that include sounds in context, listeners' success is more variable, so segment discrimination in isolation may not adequately represent the phonetic detail in stored representations. The current study explores the relationship between language experience and sensitivity to segmental cues in context by comparing the categorization patterns of monolingual English listeners and early and late Spanish-English bilinguals. Participants categorized nonce words containing different classes of English- and Spanish-specific sounds as being more English-like or more Spanish-like; target segments included phonemic cues, cues for which there is no analogous sound in the other language, or phonetic cues, cues for which English and Spanish share the category but for which each language varies in its phonetic implementation. Listeners' language categorization accuracy and reaction times were analyzed. Our results reveal a largely uniform categorization pattern across listener groups: Spanish cues were categorized more accurately than English cues, and phonemic cues were easier for listeners to categorize than phonetic cues. There were no differences in the sensitivity of monolinguals and early bilinguals to language-specific cues, suggesting that the early bilinguals' exposure to Spanish did not fundamentally change their representations of English phonology. However, neither did the early bilinguals show more sensitivity than the monolinguals to Spanish sounds. The late bilinguals however, were significantly more accurate than either of the other groups. These findings indicate that listeners with varying exposure to English and Spanish are able to use language-specific cues in a nonce-word language categorization task. Differences in how, and not only when, a language was acquired may influence listener sensitivity to more difficult cues, and the advantage for phonemic cues may reflect the greater salience of categories unique to each language. Implications for foreign-accent categorization and cross-language speech perception are discussed, and future directions are outlined to better understand how salience varies across language-specific phonemic and phonetic cues.

Keywords: speech perception, foreign-accented speech, bilingualism, language categorization, Spanish phonology, English phonology, metalinguistic awareness

INTRODUCTION

Listeners make judgments about talkers and their speech after only brief exposure. Considerable work has investigated the suprasegmental and segmental acoustic cues most important for listeners in their decisions about talker-specific characteristics like region of origin, age, and gender (Klatt and Klatt, 1990; Strand and Johnson, 1996; Harnsberger et al., 1997; Strand, 1999; Clopper and Pisoni, 2004, 2007; Tracy et al., 2015). Other cues may indicate that a talker grew up using a language other than the one being spoken, yielding a foreign accent (e.g., Flege, 1991; Flege and Munro, 1994; Flege et al., 1997a,b). At times it may even be necessary for listeners to identify which language a talker is using, for example, so that a bilingual can map a new word to the appropriate language or to facilitate a bilingual's access of a known word in one of their languages (Flege, 2007). However, unlike the work investigating associations of acoustic properties with indexical information like region of origin, cross-language speech perception tasks typically test only isolated vowels without a larger phonological context or consonants in a single CV syllable (although some work also presents stop bursts without context, e.g., Flege, 1984). These segments are often very limited in range (e.g., comparing neighboring vowels only). It is therefore unclear which segmental cues are most useful to listeners in making distinctions between their languages or whether listeners attend to all language-specific acoustic cues equally. The current project seeks to test listener sensitivity to a range of language-specific segments in nonce word contexts and considers how a listener's language background influences their use of these cues in a cross-language speech perception task.

Previous work has examined how listeners' language experience shapes their ability to categorize or discriminate isolated, or nearly-isolated, segments and subsegmental cues in cross-language speech perception. In these studies, bilingual listeners categorize or discriminate between pairs or triplets of sounds ranging along a continuum, most often the VOT continuum (e.g., between /t/ and /d/) or formant continua between neighboring vowels in the L2 (e.g., /i/ and /ɪ/). These studies have shown that monolingual English listeners and early bilinguals make similar distinctions among English categories (e.g., Mack, 1989; Flege et al., 1999a), and that this is especially true for bilinguals who have lower rates of continued use of or exposure to their L1 (Flege and MacKay, 2004). In some vowel discrimination tasks, even late bilinguals pattern like English monolinguals (Flege et al., 1994). However, listeners use a host of cues when perceiving speech beyond isolated segments or syllables, and in fact, differentiating native and non-native stop bursts may not require accessing linguistic representations at all, as is the case when listeners make parallel judgments between continua of non-speech sounds (Pisoni, 1977; Diehl and Walsh, 1989). It is possible that listeners use different, even non-linguistic and general auditory, strategies to make decisions about the isolated segments and syllables and acoustic cues used in these identification and discrimination tasks (Flege, 1987). Furthermore, these studies typically only evaluate listener sensitivity to cues in the L2, most often English, so very little is

known about how they process segments particular to their first language.¹

A few studies have attempted to extend the findings on the perception of segments in isolation or in syllables to the perception of language-specific speech and accented productions in longer stimuli. In a series of experiments, Flege (1984) found that listeners could distinguish native and non-native talkers of English after hearing CV syllables, single words, and three-word phrases. Even more remarkably, native English listeners could use input as brief as 30 ms of a stop burst to differentiate productions from native- and French-accented talkers. However, it is not clear that the strategies listeners used are the same across these varying materials despite the fact that listeners mostly accurately categorized stimuli from across this range of input. For the longer utterances, listeners may not have necessarily made use of stop burst differences at all, even though they can identify these differences in other tasks. Instead, listeners may pay more attention to other segmental and suprasegmental cues present in the longer stretches of speech. That is, the presence of a usable language-specific cue like a stop burst does not necessarily mean that this will be the most useful cue when other cues are present, and other cues may in fact be more salient to listeners than VOT. For example, evidence from a perceptual-similarity task using phrase-length stimuli from 17 languages suggests that marked back consonants and front vowel rounding might be particularly salient dimensions for non-native listeners (Bradlow et al., 2010). However, there remains some question about the interpretation of at least the vowel dimension in the perceptual-similarity study, so the number of cues present in even short phrases makes it difficult to identify the most influential acoustic factors.

Flege and Munro (1994) tested listener sensitivity to the multiple cues available in word-length stimuli by asking monolingual English listeners to categorize productions of *taco* as having been produced in English or in Spanish. The length of VOT associated with the initial /t/ explained more variance in listeners' responses than any other acoustic cue, but this language-specific difference is confounded with having occurred so early in the word – listeners may not have attended to the whole word if they could confidently make a decision based on the first segment or syllable. Since all four segments were Spanish-like or English-like in any production of *taco*, the results also do not reveal which cue(s) listeners would rely on, in the absence of the other cues. The VOT of /t/ was the strongest cue, but it is unclear if the other cues would have been sufficient for listeners to categorize productions accurately. The sensitivity of monolingual listeners to language-specific stops in Flege (1984) and Flege and Munro (1994) suggests that listeners can compare the VOT of the stimulus to their stored representations of what is an acceptable or atypical VOT for English stops. It remains to be seen whether bilinguals would show the same sensitivity to these cues in more naturalistic, word-length contexts. By manipulating a single cue in a stimulus word, and holding constant the remaining segments, we can begin to understand whether listeners from

¹See Carlson et al. (2015) for recent work on early bilinguals' use of L1 phonotactics in speech perception.

different language backgrounds can make use of a given cue when evaluating their lexical representations.

Work from mispronunciation studies indicates that bilingual listeners who can easily discriminate segments or syllables in isolation might be less able to identify those same differences in word-length stimuli, and this disparity across tasks is true even for early, highly proficient bilinguals. Listeners in these studies complete identification and discrimination tasks, and then identify whether a stimulus is the typical pronunciation of the word or if it is mispronounced. For the segment identification tasks contrasting neighboring vowels in Catalan (e.g., /*ɛ*/~/*e*/), there are conflicting results: highly proficient Spanish-dominant Spanish-Catalan bilinguals in Barcelona were unable to reliably distinguish the Catalan mid-vowels in isolation (Sebastián-Gallés and Soto-Faraco, 1999), while their peers in Majorca were successful (Amengual, 2015). However, Spanish-dominant bilinguals in both locales responded similarly poorly in the mispronunciation tasks, in which they heard a word's actual mid-vowel replaced with the neighboring vowel (e.g., /*ɛ*/ replaced with /*e*/, as in /*ə*rel/ 'root' pronounced as */*ə*rel/). Sebastián-Gallés and Soto-Faraco (1999) and Sebastián-Gallés et al. (2005) attribute the lack of detail in Spanish-dominant bilinguals' representations of Catalan to their exposure to Spanish in the first years of life, before acquiring Catalan. However, Amengual's results indicate that early Spanish exposure itself is not the cause of early bilinguals' decreased discrimination abilities in the mispronunciation task, since listeners in Majorca could reliably perceive differences when the segments were presented in isolation. This suggests that, in both regions, the Spanish-dominant bilinguals' lexical representations of Catalan contain less phonetic detail for Catalan-specific contrasts, despite the ability of some listeners to discriminate the segments in other tasks. This difference in the detail of bilinguals' lexical representations reflects the kinds of variation to which listeners are exposed, and the construction of representations is likely more complex than would be suggested by a listener's ability to discriminate isolated sounds or syllables. It is therefore important that investigations into the nature of bilinguals' representations of their languages use tasks that force listeners to respond to more complex input as language to better understand the level of detail encoded in lexical representations and to more closely approximate the challenge of processing naturalistic speech.

In fact, lexical representations incorporate not only phonological variation but social information associated with that variation as well. These indexical features, such as speaker and contextual characteristics, are encoded in the lexical representations, and they may be incorporated even after only brief exposure in the lab (e.g., Nygaard and Pisoni, 1998; Allen and Miller, 2004; Kraljic and Samuel, 2006, 2007). If the Spanish-Catalan bilinguals heard more variable input in the productions of real words, their representations of Catalan may have included both productions as possible, explaining their difficulty identifying mispronunciations, whereas the monolinguals in Flege (1984) and Flege and Munro (1994) may have been exposed to less variation in English and so were more sensitive to deviations from typical productions. There is also evidence demonstrating that listeners with exposure to specific

accents, even in absence of knowing the L2, show improved processing and categorization of those accents (Clopper and Pisoni, 2004, 2007; Vieru et al., 2011; Witteman et al., 2013), so language and a talker's language proficiency must also be linked to specific productions.

These associations of indexical information with productions, and the incorporation of acoustic variation in lexical representations, are in line with exemplar theories of speech perception (Johnson, 1997; Pierrehumbert, 2002). Listeners use stored exemplars – those from an exposure period in a lab or from hearing productions in normal life – to inform their expectations about unheard productions and word forms. Thus, listeners can generalize over a number of stored exemplars about what kinds of stops, for example, occur in English or in the productions of a particular talker of English. Listeners like bilinguals who have experience with a sound category in both languages must associate productions with each language in order to make the appropriate conclusions about the phonological categories in each language (as in the related BLINCS model in Shook and Marian, 2013). For example, a Spanish-English bilingual who hears a word produced with a /*t*/ will store with this exemplar whether the sound was produced in English or Spanish, and information about how it was produced (e.g., the VOT of the stop) will be added to the listener's representation for the production of /*t*/ in the language. Spanish-English bilinguals will therefore have developed detailed phonological representations for English and Spanish, and their sensitivity to the distribution of sounds particular to each language might be expected to be greater than that of English monolinguals, who have only English productions on which to base their language representations. While English monolinguals may have some, or even significant, exposure to Spanish-accented English, their knowledge of Spanish phonology will be less than that of bilinguals who have acquired Spanish since birth. In fact, due to existence of multiple (language-specific) categories in the same phonological space, Spanish-English bilinguals' representations might also be unlike English monolinguals' in other ways: bilinguals might use categories more extreme than monolinguals to maximize differences between languages (cf. Flege, 1995), or bilinguals' categories may show evidence of cross-linguistic transfer and be less like the monolinguals', especially for later-acquired sounds and for later learners (Flege, 2007).

The present study tests the effect of language experience on listener sensitivity to language-specific segments to better understand how language-specific sounds are represented and related in the bilingual lexicon. We use a novel task in which listeners are told they are hearing snippets of continuous speech (either in Spanish or English) and are asked to associate the nonce words containing a Spanish- or English-specific sound with the appropriate language. Accuracy and reaction times (RTs) are compared across listener groups for each of the classes of segment. The use of nonce words has two advantages. First, presenting word-length stimuli forces listeners to process the sounds linguistically and not just auditorily, and there is evidence that listeners in previous studies may have perceived segments without linguistic context differently than when the same sounds

were processed as words. Second, unlike real words, nonce-word stimuli avoid inducing lexical effects related to listeners' actual exposure to the phonological variations of real words. Finally, the use of word-length nonce stimuli, purportedly taken from naturally produced speech, forces listeners to generalize over the phonological properties of their languages and decide in which language a given stimulus must have been produced. The present study also extends previous work, which mostly tested contrasts from only one language (e.g., English in Flege's work and Catalan in the work of Sebastián-Gallés and Amengual), by including cues from both English and Spanish to more fully investigate how listeners' language backgrounds influence perception in both languages.

The nonce words tested here include segmental categories that are unique to English or Spanish ("phonemic" cues) and segments that vary in how they are implemented phonetically along a continuum between the Spanish variant and the English variant ("phonetic" cues). Similar distinctions among segments have been made for the perception of non-native sounds that vary in similarity to native categories (Best, 1991) and for the acquisition of second language sounds, in the Speech Learning Model (Flege, 1987, 1995). Evidence suggests that sound categories that are "new" to an L2 and have no counterpart in the L1, like the phonemic cues proposed here, are easier to perceive as a distinct category and to produce authentically than "similar" L2 phones that differ along some particular acoustic-articulatory dimension from the L1 variant, like the phonetic cues described here. One study (Flege and Munro, 1994) has specifically examined phonetic cues in context and found that listeners could use these cues to varying degrees depending on the language background of the talker, but no work has directly compared phonemic and phonetic cues. Following Flege and Munro (1994) and the predictions outlined in the Speech Learning Model for new and similar phones, both classes of cues are expected to be successfully associated with their respective languages but phonemic cues are expected to be stronger indicators of language than phonetic cues in a language categorization task.²

Finally, this study also systematically compares the sensitivity of monolingual English listeners and early and late Spanish-English bilinguals. Previous work in cross-language speech perception indicates similarities between English monolinguals and early Spanish-English bilinguals in the categorization of English sounds, but evidence regarding how late bilinguals compare to these groups is more limited. It is expected that the bilingual groups will show greater sensitivity to language-specific cues from both languages than the monolinguals, since the bilinguals' considerable exposure to both English and Spanish productions should foster more reliable associations between language and the phonetic detail in stored representations.

²While the Speech Learning Model distinguishes between new and similar phones in a second language, this binary may not be sufficient to include all relationships between the sounds of one's native language and the categories in a second language. For example, it is unclear how to classify a shared phone with different statuses in each language, e.g., both Spanish and English use the tap [ɾ], but this sound is phonemic in Spanish and allophonic in English.

MATERIALS AND METHODS

Materials

Language-Specific Target Segments

Three language-specific phonemic cues were chosen for the categorization task: the English-specific segments /θ/ and /ɹ/, and the Spanish-specific trill /r/. We limited the selection of phonemic cues to those sounds that form categories not present in the other language and that do not form a continuum. For example, the English voiced alveolar approximant /ɹ/ and the Spanish voiced alveolar trill /r/ are not different extremes of a continuum between /ɹ/ and /r/, in the way that English and Spanish voiced and voiceless stops vary along a single dimension (VOT). That is, there is not a single dimension or acoustic correlate that distinguishes /ɹ/ and /r/ that could be increased or decreased to derive one from another, since the two sounds are produced with fundamentally different manners of articulation (/ɹ/ as an approximant and /r/ as a trill). One additional English-specific cue was identified for inclusion as a phonemic cue, /θ/. Although /θ/ is a phoneme in Peninsular Spanish (it is produced as /s/ in Latin America), it was included as an English-specific phoneme since exposure to Peninsular Spanish among our listeners was expected to be very limited, and native speakers of Peninsular Spanish were excluded from the study. Early Spanish-English bilingual listeners living in Central Texas, where this study was conducted, may have some exposure to Peninsular Spanish, for example through movies, but are most familiar with Latin American dialects of Spanish. The late bilingual participants likely have more exposure to Peninsular Spanish than early bilinguals, but it is not expected that this exposure would be more influential on L1 representations than native dialect phonology. In fact, many monolingual English listeners probably have exposure to the trill /r/ in Scottish English, also through media, but it would be surprising if their language-segment associations reflected occasional exposure to the trill /r/ in English.³ Vowels were excluded as phonemic cues for this language pair for two reasons. First, all five Spanish vowel categories exist in English, minimally in English diphthongs, so there were no Spanish-specific vowels to consider for phonemic cues. Second, English-specific vowels (e.g., /ɪ/) can be differentiated from the nearest shared vowels (e.g., /i/) by both spectral cues and duration differences; while native listeners attend to the spectral differences in these English-specific vowels, non-native listeners may rely on vowel duration to distinguish these categories (Flege et al., 1997a; Escudero, 2006; Kondaurova and Francis, 2008). In this case, non-native listeners would be able to use the duration continuum between the short /ɪ/ and the long /i/. Instead, we wanted to ensure as much as possible that all listener groups included in this study were attending to the same acoustic property of the target segment.⁴

³In fact, our results suggest that late bilingual listeners were even more sensitive than the other listener groups to the association of /θ/ with English. See the discussion for additional analysis of how the different listener groups categorized stimuli with /θ/.

⁴While vowels can be described as differing from one another along (minimally) three continuous dimensions (F1, F2, and duration), there can in fact be phonemic or "new" categories across languages. This would be the case, for example, for

In addition to the phonemic cues, we also tested phonetic cues, which vary along a continuum. These sound categories exist in both languages but their articulation in each language is characterized by sub-phonemic differences in place of articulation. Two language-specific phonetic segments were chosen for the task, the lateral approximant /l/ and the high back vowel /u/. The lateral approximant is produced as a 'light' [l] at the alveolar ridge in Spanish, while in American English the segment is realized as the 'darker' [ɫ], with an additional closure near the velum, particularly in closed syllables (Recasens, 2004, 2012). The back vowel differs along F2 in English and Spanish: it is fronted to [ʊ] for many speakers of American English and is produced further back, as [u], in Spanish (Mendez, 1982; Bradlow, 1995; Clopper et al., 2005).

Nonce Words

Nonce words were created to test the contributions of specific sounds to listeners' conceptualizations of Spanish and English. All nonce words were disyllabic trochees with either two open syllables (i.e., CVCV) or /l/ in coda position of the first syllable (i.e., CVl/CV). The CVl/CV structure was included in the nonce words to provide two phonological contexts for /l/ stimuli that were both permissible in Spanish and in which /l/ was most likely to be velarized to [ɫ] in American English (Recasens, 2012). The inclusion of disyllabic words with stress on the first syllable meant that the second English vowel would be reduced to schwa, resulting in an additional vowel-quality cue beyond the language-specific target segment. However, this strategy was preferred to the development of monosyllabic words for several reasons. Spanish has relatively few monosyllabic words compared to English (cf. Costa and Caramazza, 1999) so monosyllables may be biased toward English responses. The set of possible word-final consonants in Spanish is very small: /ð, s, n, l, r/. Some of these are subject to lenition (/ð/) or aspiration (/s/), or are already included as a language-specific target segment (/l/). Words ending in /r/ are associated with infinitive morphemes, and /r/ is also in free variation with /r/ word-finally. The inclusion of a second syllable and vowel reduction was therefore preferred. Vowel reduction and its potential influence on listeners' language decisions are addressed in the discussion.

Each nonce word included one language-specific segment that served as a cue to language categorization. The remaining segments in the nonce words exist in both English and Spanish (at least phonemically, as in the case of the English unstressed schwa) and are not expected to differ between the two languages, so that listeners would be obligated to use the target segment for the language categorization decision. The segments identified as common to both English and Spanish were the fricatives /m, f, s, h/⁵ and the affricate /tʃ/, which do not differ between the

English listeners perceiving French /y/, which does not exist as a category in English, even though it may initially be confused with English /u/ or French /u/ (Flege, 1987); English listeners treat French /y/ as a language-specific category sooner than they recognize French /u/ as a category unique from English /u/. This, however, is not the case for any Spanish-specific vowel, which are in line with the French /u/-English /u/ relationship.

⁵The phoneme identified here as /h/ is alternately realized as /x/ in some dialects of Spanish (Hualde, 2005). The speaker chosen to record the stimuli uses /h/ in his dialect of Spanish; see "Stimuli Recordings and Speaker."

languages in point of articulation or in voicing, and the vowels /i, a/. While /i, a/ are realized somewhat differently in English and Spanish, with the English variants sometimes transcribed as /ij/ and /a/, respectively, these vowels were preferable over others. Mid-vowels are diphthongized in American English, and /u/ was included as a target segment due to the variation in its articulation in English and Spanish. The symbol /i/ is used here to indicate the vowel in Spanish *mi* 'my' /mi/ and English *me*, and /a/ is used to represent Spanish *la* /la/ 'the' and the vowel in English *cot*. Although /a/ is more variable than /i/ across the languages (Bradlow, 1995), it was included to increase the number of possible stimuli.

For each target segment, eight nonce CVCV and CVl/CV words were constructed from the set of segments overlapping in English and Spanish. Each nonce word was a possible, but non-existent, word in both English and Spanish, and all words ended with /a/, which was reduced to [ə] in the English stimuli. See **Table 1** for the set of stimuli containing language-specific phonemes and **Table 2** for the set of stimuli containing language-specific phonetic segments. One phonemic stimulus, *racha*, was identified as a real Spanish word meaning 'gust of wind' after the study had been completed, so it was excluded from the following analyses. The English nonce word /ɪatʃə/ was also removed due to its similarity to the Spanish *racha* /ratʃa/, since bilingual listeners may have interpreted this stimulus as the Spanish word *racha* produced with an English accent and not as a uniquely English word.

Stimuli Recordings and Speaker

A single speaker was chosen to record both English and Spanish stimuli, and this was crucial to the experimental task. A single speaker was preferred over recording two monolinguals to avoid voice being a cue to language, and using natural productions of the stimuli ensured there were no acoustic artifacts from splicing or otherwise manipulating segments within a word frame. Using natural productions from a single talker also permitted the selection of the desired segments as target segments, regardless of difficulties isolating them (e.g., with the English /ɪ/).

Since it was also important for the stimuli to lack any language-specific cues, or accent, beyond the controlled target

TABLE 1 | Nonce words with language-specific phonemes /θ, ɪ, r/.

English phoneme /θ/	English phoneme /ɪ/	Spanish phoneme /r/
/tʃiθə/	/tʃaɪə/	/tʃira/
/fiθə/	/fiɪə/	/fara/
/hiθə/	/hiɪə/	/fira/
/maθə/	/maɪə/	/mara/
/saθə/	/saiə/	/mira/ ¹
/siθə/	/siɪə/	/ratʃa/
/θiʃə/	/θimə/	/ritʃa/
/θisa/	/θiɪə/	/sira/

¹Note that the Spanish nonce-word /mira/, which would be written *mirra*, is distinct from the real Spanish word *mira* /mira/ 'look,' which is produced with the tap /ɾ/. Such minimal pairs contrasting /r/ and /ɾ/ exist elsewhere in Spanish; consider *carro* /karo/ 'car' vs. *caro* /karo/ 'expensive' and *perro* /pero/ 'dog' vs. *pero* /pero/ 'but.'

TABLE 2 | Nonce words with language-specific phonetic variants of /l,u/.

/l/		/u/	
English	Spanish	English	Spanish
[tʃaɪsə]	[tʃaltʃa]	[tʃutʃə]	[tʃuma]
[faɪmə]	[filfa]	[fʊtʃə]	[fufa]
[hiɪfə]	[laɸa]	[fuɸə]	[fusa]
[litʃə]	[litʃa]	[fusə]	[mufa]
[lifə]	[lifa]	[hutʃə]	[muma]
[maɪfə]	[maɪfa]	[husə]	[sutʃa]
[saɪfə]	[silma]	[mʊmə]	[hutʃa]
[siɪtʃə]	[halfa]	[sufə]	[husa]

segment, care was taken to recruit a balanced Spanish-English bilingual who produced both languages as natively as possible. The chosen talker was a 37-year-old Spanish-English bilingual who was born and raised in Colombia until the age of 7 at which point he moved to the state of New York with his family. He continued to speak Spanish at home in New York, and as an adult he moved to Texas for graduate school, during part of which he lived in Guatemala and Spain to conduct research. While most of his current daily interactions were in English, he also used Spanish on a daily basis with his family and frequently for translating and interpreting professionally at work. An accentedness rating study was conducted to ensure that the talker’s English and Spanish productions sounded native-like to native English and native Spanish speakers, respectively. In both languages, the talker was rated as native-like as other talkers who grew up as monolingual speakers of each language. See the **Appendix** for a complete description of the accentedness ratings.

The English and Spanish nonce words were recorded in separate sessions to further ensure minimal cross-linguistic transfer. The recordings took place in a sound-attenuated booth using a MOTU UltraLite-mk3 Hybrid recorder at a sampling frequency of 44.1 kHz (16 bit). The talker repeated each nonce word three times so that the clearest repetition could be chosen. The words were written in English and Spanish orthography (e.g., English *leefuh* for [lifə] and Spanish *chirra* for [tʃira]) and not in the International Phonetic Alphabet (IPA), so for some items the talker was coached to arrive at the intended pronunciation. The pitch contours were manipulated to match a naturally produced token with a falling contour using Praat (Boersma and Weenink, 2012). The beginning and end points of the F0 contours were set to 170 and 124 Hz to match the values of model token. The intervening pitch points were interpolated between the two end points.

Participants

Participants (*n* = 53) were recruited through the Department of Linguistics subject pool and received course credit for their participation. To supplement the subject pool participants with the listeners who had the needed language backgrounds, the remaining Spanish–English bilinguals, both early and late (*n* = 27) were recruited through the University of Texas Events Calendar. These participants were paid \$10/h for their time.

Listeners completed a language history questionnaire (Chan, 2014) that included questions about participants’ biographical information, the places they had lived and for how long, their language exposure and proficiency, and their language(s) of education. Based on their responses to the questionnaire, participants were divided into three groups: monolingual English speakers with minimal or no exposure to Spanish (Monolingual), Spanish-English bilinguals from the U.S. who acquired both languages in early childhood (Early Bilinguals), and Spanish–English bilinguals from Spanish-speaking countries who acquired English as adults (Late Bilinguals). Participants who did not fit into one of these groups were not included in the final sample (*n* = 24). See **Table 3** for a summary of participant characteristics.

Forty participants (21 females) were included in the Monolingual group. All members of this group were from the U.S., had heard English from birth, did not hear another language at home, and were not proficient in any other language. Participants ranged in age between 18 and 29, and the mean age of the group was 20. Of the 40 Monolingual listeners, 24 had studied Spanish in middle and/or high school. One additional participant had some Spanish classes in elementary school, and one further participant reported learning some Spanish as a toddler outside the home. All 26 listeners with some exposure to Spanish reported very low proficiency in the language.

The Early Bilinguals group included 18 participants (15 females) who ranged in age from 18 to 29, with a mean age of 20 years. Eleven of the listeners in the Early Bilinguals group were born and raised in the United States, and the remaining seven participants were born in Mexico (*n* = 6) or Colombia (*n* = 1) and moved to the U.S. before they began elementary school. All listeners in the Early Bilinguals group had learned Spanish at home since birth. Seven participants also learned English at home since birth (four of the U.S.-born participants, three of the foreign-born participants). The remaining 11 participants began learning English when they started elementary school.

Twenty-two listeners (11 females) were categorized as Late Bilinguals since they were born and raised in a Spanish-speaking country and moved to the U.S. after age 14. Listeners in this group ranged in age between 18 and 43, with a mean age of 28 years. Only Late Bilinguals from Latin America participated;

TABLE 3 | Demographic information and language background of participants.

	Monolinguals	Early bilinguals	Late bilinguals
N	40	18	22
Mean age	20	20	28
Age range	18–29	18–29	18–43
Females	21	15	11
Mean age (in years) when learned English	0	3.7	10
Mean age (in years) when learned Spanish	12.5	0	0
Mean age (in years) when moved to U.S.	NA	1.3	20.1

listeners from Spain were excluded since /θ/ is phonemic in Peninsular Spanish and the present study included /θ/ as an English-specific phoneme. Listeners were from Mexico ($n = 11$), Argentina ($n = 2$), Peru ($n = 2$), Ecuador ($n = 2$), Bolivia ($n = 1$), Venezuela ($n = 1$), Colombia ($n = 1$), the Dominican Republic ($n = 1$), or some combination of these countries ($n = 1$). Late Bilinguals ranged in the age at which they moved to the U.S. between 14 and 28, with mean age of arrival of 20. All listeners had learned only Spanish at home since birth. Although all had studied English at least informally in school before they moved to the U.S., Spanish was the only language of instruction in both primary and secondary school for all Late Bilingual participants.

Procedure

Participants completed the nonce-word categorization experiment in the UT Sound Lab in the Department of Linguistics at the University of Texas at Austin. The study was approved by the Institutional Review Board at UT Austin, and the experimenter obtained written informed consent from the participant before beginning the study, in accordance with the IRB's recommendations. Listeners answered an online language history questionnaire and were tested for normal hearing, followed by the categorization experiment.

Listeners performed the language categorization task in a sound-attenuated booth on a PC running E-Prime 2.0 (Psychology Software Tools, 2010). Listeners wore Sennheiser XX headphones and were oriented to the serial response button box (Psychology Software Tools, 2010). Participants were instructed to place the index and middle fingers of their dominant hand on the two leftmost buttons, which were labeled with "ENG" and "SPAN," the order of which was counterbalanced across participants. The language that corresponded to each button was also presented on the computer screen, e.g., "ENGLISH" appeared on the left side of the screen for the group of participants who used the left button to indicate English words. Listeners began with a practice block in which they read instructions presented on-screen and decided if each word sounded more like English or more like Spanish. The practice block included 20 real words (10 English, 10 Spanish).

After the practice block, the test portion began. At test, listeners were told they would hear "snippets of speech that were taken out of longer recordings while the speaker was talking in either English or Spanish," and they were asked to decide if what they heard sounded more like it came from the English recording or the Spanish recording. This wording and context was provided after piloting indicated that some listeners had the impression that they were hearing accented productions instead of words from two languages. To avoid this confusion between accent and language, the categorization task was rephrased to ask about the language being used to produce the word.⁶ Listeners

categorized the 56 nonce words (listed in **Tables 1** and **2**) eight times, and stimuli were randomized within each of the eight blocks, for a total of 448 trials. There was a one second pause between a listener's response and the onset of the audio for the next stimulus. RT was calculated from the onset of the audio file, and categorization decision and RT were recorded for each trial.

RESULTS

Categorization decision (Spanish or English) and RT were recorded for each trial. Decisions were coded as accurate if words containing the English-specific phoneme /ɪ/ or /θ/ or the English variants [ɪ] or [θ] were classified as English and if words with the Spanish-specific phoneme /r/ or the Spanish variants [l] or [u] were classified as Spanish. Trials with the Spanish stimulus *racha* /ratʃa/ and the English stimulus /ɪatʃə/ were excluded from the analyses (cf. Nonce Words). RTs were calculated by subtracting the length of the stimulus.wav file from the time calculated by E-Prime between trial onset and button press. This ensured that the RTs analyzed here reflected the length of time for the listener to make a categorization decision, after hearing the end of the stimulus word. Trials with RTs less than 200 ms ($n = 665$; 1.9%) were discarded as spurious responses. RTs were log-transformed from milliseconds to normalize the distribution of responses for the regression analyses. Less than 0.5% of responses exceeded 5000 ms and the distance of these from the mean was reduced in the log transformation. Trials more than three standard deviations above or below a participant's log-transformed mean were excluded as outliers ($n = 228$; 0.7%). The spurious responses and outliers accounted for 2.6% of all trials ($n = 893$), after *racha* and the English /ɪatʃə/ were removed. The following analyses include the remaining 33667 trials (Monolinguals: $n = 16800$; Early Bilinguals: $n = 7441$; Late Bilinguals: $n = 9426$). Accuracy (correct, incorrect) and log-transformed RT were submitted to separate regression analyses, which were analyzed using Bayesian inference with the *glmer2stan* package (v0.995) in R (v3.2.2) to interface with Stan via RStan (v2.8.2).

Acoustic Analyses

Segmental properties of each stimulus were measured using Praat to ensure that the Spanish and English productions differed in the expected dimensions. The duration and first three formants of both vowels of each stimulus were measured, and the same measures were taken for the /l/ variant in the stimuli containing an English or Spanish /l/. Formant measurements were taken at the vowel midpoint and at 25 and 75% through the vowel. Recall that the vowels /i,a/ were used in the first vowel position of the disyllabic nonce words to create a sufficient number of non-word stimuli, and the second vowel (V_2) of each nonce word was realized as the full-vowel [a] in Spanish words and as the reduced [ə] in English words. The Spanish [u] and English [ʊ] segments were target vowels representative of phonetic cues. The acoustic

words uniformly ended in /a/, it is not a feature that distinguishes some stimuli from others. See discussion for potential language-specific properties of the nonce words.

⁶This phrasing invites the possibility that listeners may have looked for other patterns in the stimuli to make their categorization decisions, such as the appearance of language-specific morphemes in the nonce words. All nonce words did end in /a/, which is the Spanish morpheme for feminine adjectives (e.g., *rojo* /roho/ 'red-MASC' vs. *roja* /roha/ 'red-FEM') and is also one of the morphemes for third-person singular (e.g., *habla* /abla/ 'speaks-3SG'). However, since all nonce

TABLE 4 | Acoustic properties of segments.

	Duration (ms)		F1 (Hz)		F2 (Hz)		F3 (Hz)	
	English	Spanish	English	Spanish	English	Spanish	English	Spanish
(A) Non-target vowels								
/i/	87.0 (22.6)	95.6 (20.3)	369.7 (47.4)	361.0 (31.9)	2245.3 (243.7)	2196.3 (107.9)		
/a/	116.9 (19.0)	99.1 (14.4)	878.8 (67.4)	835.7 (15.1)	1189.4 (74.6)	1524.6 (55.1)		
V ₂	174.4 (29.0)	141.5 (31.4)	693.7 (67.6)	769.8 (130.8)	1367.4 (143.3)	1484.5 (97.7)		
(B) Target segments								
/i/	63.8 (22.9)	77.7 (17.9)	581.6 (134.7)	383.4 (88.3)	1141.4 (260.3)	1917.4 (331.8)	2999.2 (253.4)	2937.6 (375.9)
/u/	81.7 (11.9)	82.7 (18.3)	415.8 (22.2)	484.5 (170.9)	1560.9 (178.5)	1174.0 (372.5)		

properties of the segments are reported in **Table 4**: in (A) are reported the mean duration and formant values for the English and Spanish productions of the non-target vowels, and in (B) are the measurements of the language-specific variants of the target segments /i,u/. Formant values are the mean of the measurements taken at the midpoint of each vowel. Standard deviations are included in parentheses.

In order to test whether the English and Spanish variants were distinct from each other, the concordance statistic (c-statistic) of a logistic regression model was analyzed. The c-statistic is the proportion of outcomes that are correctly predicted by the fitted model. For each vowel, a logistic regression model was constructed in R (RStudio 0.99.489; RStudio Team, 2015) using the *rms* package (v4.2-1) with language (English, Spanish) as the dependent variable and the duration and midpoint measures of F1 and F2 as fixed effects. Measurements were centered and scaled, and duration was removed from the model where singularity remained. The model for English and Spanish /i/ additionally included the midpoint measure of F3 as a fixed effect. Constructing such a model for the c-statistic was preferable to testing for differences between each fixed effect separately since listeners hear the multiple acoustic cues at once; that is, listeners may attend to differences in all three dimensions (F1, F2, and duration), so all three should be considered together when determining if the sounds were distinct in the two languages.

For the two target segments that were measured, /i/ and /u/, it was expected that the formants and the duration of the segment would be sufficient to distinguish the English and Spanish variants. The model with these three main effects as well as the midpoint of F3 made perfect discrimination between the English [i] and the Spanish [i] ($C = 1.000$). For English [u] and Spanish [u], the duration variable was removed to avoid singularity, and the model with the midpoints of F1 and F2 was also highly successful ($C = 0.969$).

The other three segments were the two vowels /i,a/, which were used in the first syllables of the nonce words, and the final vowel of the nonce words. The initial model for /i/, with duration and the midpoint measurements for F1 and F2, produced a c-statistic of 0.681, which represents a moderately good fit to the differences in /i/ in English and Spanish words, but which falls short of the clear distinction between the

phonetic variants described above. For /a/ in the position of nucleus of the first syllable, the model was highly successful for discrimination ($C = 1.000$). Finally, the model for the second (unstressed) vowel in the nonce words fit well ($C = 0.853$). The acoustic distance between English and Spanish /a/ in stressed and unstressed positions, as well as those between the /i/ variants, was expected (cf. Bradlow, 1995); see “the discussion” for how the accuracy and RT results should be understood in light of these differences.

Accuracy Analysis

The mean accuracy score of each group for each stimulus type is presented in **Table 5**. The accuracy results were analyzed using a Bayesian mixed effects logistic regression model with listener language group (Monolingual, Early Bilingual, Late Bilingual), stimulus language (English, Spanish), and stimulus type (phonemic, phonetic) as fixed effects and participant and stimulus word as random intercepts. The models were fitted via a Markov Chain Monte Carlo procedure using STAN (Gelman et al., 2015). Model comparison was performed using the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002). A model with a three-way interaction among the fixed effects provided an improved fit over models with two-way interactions or with only main effects (see **Table 6** for the model summary). The reference group, reflected in the model intercept, represents the accuracy of Monolinguals categorizing stimuli with an English phoneme. The fitted log odds of accuracy for each stimulus language and listener language group are plotted in **Figure 1**, with the phonemic cues in the left panel and the phonetic cues in the right panel. The error bars represent the 95% Bayesian credible intervals.

TABLE 5 | Mean accuracy of each listener group for each stimulus type.

		Monolinguals	Early bilinguals	Late bilinguals
English Cues	Phonemic	72.7 (44.5)	78.8 (40.9)	86.1 (34.6)
	Phonetic	70.5 (45.6)	73.2 (44.3)	76.0 (42.7)
Spanish Cues	Phonemic	95.6 (20.5)	96.9 (17.3)	97.6 (15.4)
	Phonetic	91.0 (28.7)	90.4 (29.5)	90.6 (29.1)

Standard deviations are presented in parentheses.

TABLE 6 | Summary of mixed effects logistic regression model fitting accuracy results.

Predictor	Mean	Posterior SD	95% CI	p-value
Intercept (Monolingual, English phonemes)	1.391	0.299	(0.763, 1.983)	<0.0001
Phonetic cues	-0.200	0.356	(-0.916, 0.489)	NS
Early bilinguals	0.299	0.273	(-0.244, 0.818)	NS
Late bilinguals	1.014	0.236	(0.546, 1.468)	<0.0001
Spanish cues	2.242	0.459	(1.402, 3.190)	<0.0001
Phonetic * Early	-0.247	0.097	(-0.442, -0.059)	NS
Phonetic * Late	-0.722	0.097	(-0.911, -0.533)	NS
Phonetic * Spanish	-0.562	0.559	(-1.629, 0.521)	<0.0001
Early * Spanish	-0.223	0.230	(-0.659, 0.233)	<0.0001
Late * Spanish	-0.686	0.231	(-1.135, -0.226)	<0.0001
Phonetic * Early * Spanish	-0.165	0.256	(-0.674, 0.322)	<0.0001
Phonetic * Late * Spanish	0.126	0.253	(-0.365, 0.619)	<0.0001
Random effects Variance				
Listener	0.892			
Stimulus word	0.970			

Comparing Spanish and English Phonemic and Phonetic Cues

Overall, listeners responded more accurately to Spanish cues than to English cues, and to phonemic cues than to phonetic cues. The difference between the languages was greater for phonemic cues than for phonetic cues. The Spanish phoneme was categorized more accurately than the English phonemes (Monolinguals: $\beta = 2.242$, posterior $SD = 0.459$, $p < 0.0001$; Early Bilinguals: $\beta = 2.019$, posterior $SD = 0.484$, $p < 0.0001$; Late Bilinguals: $\beta = 1.556$, posterior $SD = 0.491$, $p < 0.001$), and the

Spanish phonetic cues were also categorized more accurately than the English phonetic cues (Monolinguals: $\beta = 1.680$, posterior $SD = 0.367$, $p < 0.0001$; Early Bilinguals: $\beta = 1.292$, posterior $SD = 0.373$, $p < 0.001$; Late Bilinguals: $\beta = 1.120$, posterior $SD = 0.372$, $p < 0.001$). The Early Bilinguals trended toward categorizing the English phonemic cues more accurately than the English phonetic cues ($\beta = 0.448$, posterior $SD = 0.358$, $p = 0.09$). The Late Bilinguals categorized English phonemic cues significantly better than English phonetic cues ($\beta = 0.922$, posterior $SD = 0.358$, $p < 0.01$). All groups categorized the Spanish phonemic cue more accurately than the Spanish phonetic cue (Monolinguals: $\beta = 0.763$, posterior $SD = 0.451$, $p < 0.01$; Early Bilinguals: $\beta = 1.175$, posterior $SD = 0.477$, $p < 0.0001$; Late Bilinguals: $\beta = 1.359$, posterior $SD = 0.480$, $p < 0.0001$).

Comparing Listener Groups

The three listener groups responded very similarly within each segment type, with the exception of the categorization of nonce words with an English phoneme. For the English phonemes, Monolinguals and Early Bilinguals responded less accurately than the Late Bilinguals (vs. Monolinguals: $\beta = 1.014$, posterior $SD = 0.236$, $p < 0.0001$; vs. Early Bilinguals: $\beta = 0.715$, posterior $SD = 0.294$, $p < 0.05$). There were no group differences in the English phonetic cue conditions, and there were also no significant group differences in response to the Spanish phonemic or the Spanish phonetic cues.

Reaction Time Analysis

The mean RTs (in milliseconds) of each group for correct responses to each stimulus type are presented in Table 7. Log-transformed RTs were analyzed using a Bayesian mixed effects linear regression model with listener language group (Monolingual, Early Bilingual, Late Bilingual), stimulus language (English, Spanish), stimulus type (phonemic, phonetic), and accuracy (correct, incorrect) as fixed effects. Participant and stimulus word were included as random intercepts. These models

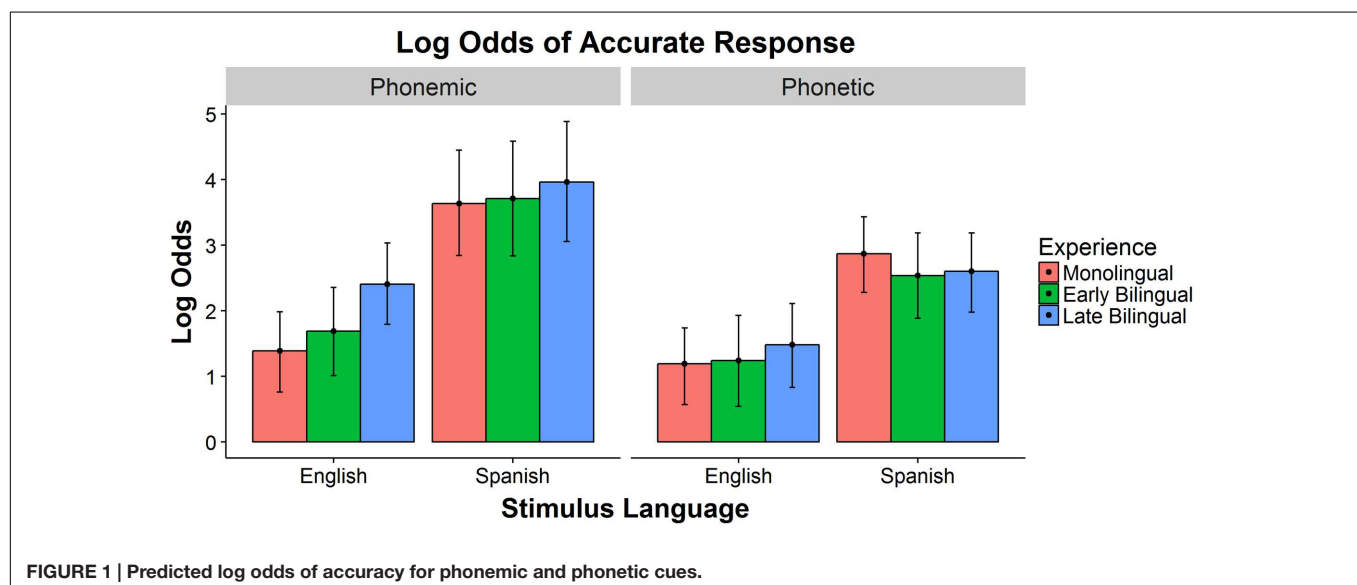


TABLE 7 | Mean RT (in milliseconds) for correct trials for each listener group and stimulus type.

		Monolinguals	Early bilinguals	Late bilinguals
English cues	Phonemic	542.0 (594.1)	629.8 (727.8)	662.7 (640.7)
	Phonetic	592.3 (742.9)	715.5 (833.4)	770.8 (791.5)
Spanish cues	Phonemic	538.0 (591.8)	530.4 (545.1)	639.8 (675.8)
	Phonetic	595.4 (636.6)	641.2 (711.6)	777.1 (792.7)

were also fitted via a Markov Chain Monte Carlo procedure using STAN, as described above. Testing for a significant effect of categorization accuracy evaluated the possibility that listeners' RTs were unaffected by the accuracy of the categorization decision. A model with the same three fixed effects as the accuracy model – listener group, stimulus language, and stimulus type – was significantly improved by adding accuracy as a fixed effect. RTs thus significantly differed between accurate and inaccurate trials, and subsequent models calculated separate betas for each type of trials. The model with a four-way interaction among the fixed effects provided a better fit than models with only main effects, with two-way interactions, or with three-way interactions. See **Table 8** for the model summary. The reference group, reflected in the model intercept, represents the log RT of inaccurate responses by Monolinguals categorizing stimuli with an English phoneme. The fitted log RT for correct responses to each target segment and listener language group are plotted in **Figure 2**. The error bars represent 95% Bayesian credible intervals. The following sections report the results of correct trials from the four-way interaction and the differences between correct and incorrect responses.

Comparing Spanish and English Phonemic and Phonetic Cues

For the four cue types, there were few significant differences in RTs. The only differences appeared for the Spanish cues: the Early Bilinguals trended toward faster RTs for the Spanish phonemic cue compared to the Spanish phonetic cues ($\beta = 0.144$, posterior $SD = 0.073$, $p = 0.08$), and the Late Bilinguals responded significantly faster to the Spanish phoneme than to the Spanish phonetic cues ($\beta = 0.164$, posterior $SD = 0.073$, $p < 0.05$). There was no difference between the Spanish categories for Monolingual listeners. The differences in RT between the English phonemic cues and the English phonetic cues did not reach significance for any listener group. There were also no differences in RTs between the English and Spanish phonemic cues or between the English and Spanish phonetic cues.

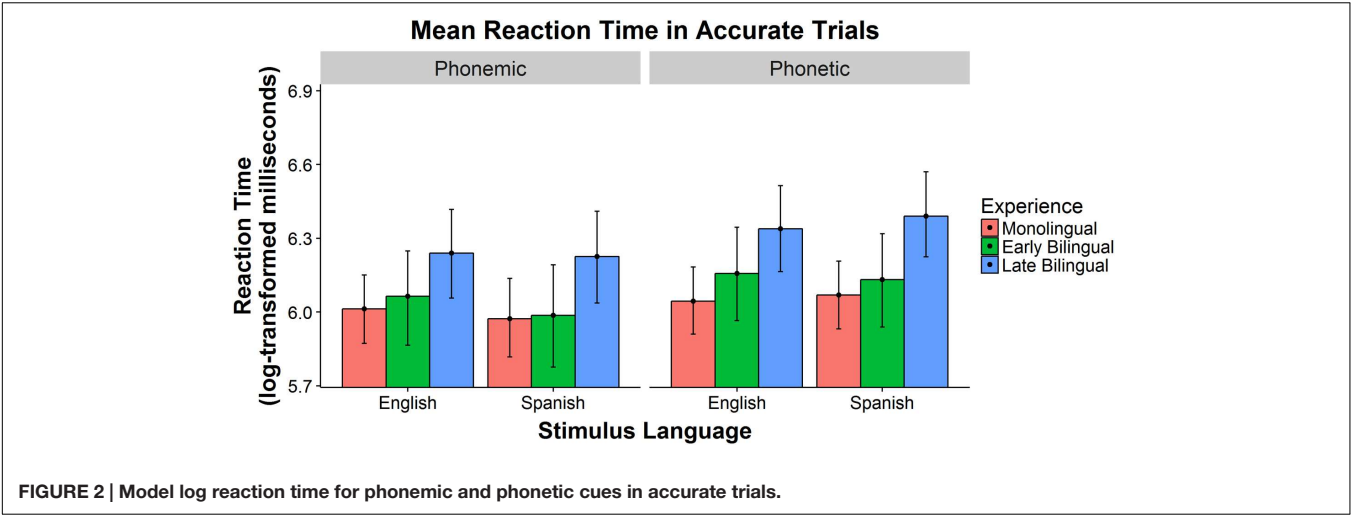
Comparing Listener Groups

The pattern of differences in RTs among the listener groups was mostly constant across segments: Monolinguals and Early Bilinguals responded with similar RTs, and both these groups were faster than Late Bilinguals. For the Spanish phonemic cue, there was no difference between Monolinguals and Early Bilinguals, and both groups were significantly faster than Late Bilinguals (vs. Monolinguals: $\beta = 0.252$, posterior $SD = 0.100$, $p < 0.01$; vs. Early Bilinguals: $\beta = 0.238$, posterior $SD = 0.124$, $p < 0.05$). For English phonemes, Monolinguals and Early

TABLE 8 | Summary of mixed effects linear regression model fitting log-transformed RT results.

Predictor	Mean	Posterior SD	95% CI	p-value
Intercept	6.191	0.074	(6.046, 6.333)	<0.0001
(Monolingual, English phonemes)				
Phonetic cues	0.040	0.059	(−0.074, 0.158)	NS
Early bilinguals	−0.011	0.114	(−0.236, 0.216)	NS
Late bilinguals	0.359	0.107	(0.146, 0.557)	<0.0001
Spanish cues	0.026	0.102	(−0.179, 0.226)	NS
Correct response	−0.178	0.025	(−0.224, −0.128)	<0.01
Phonetic * Early	−0.055	0.053	(−0.155, −0.051)	NS
Phonetic * Late	−0.192	0.052	(−0.296, −0.090)	<0.01
Phonetic * Spanish	0.037	0.121	(−0.208, 0.273)	<0.10
Early * Spanish	0.155	0.152	(−0.145, 0.449)	<0.05
Late * Spanish	−0.194	0.153	(−0.492, 0.106)	<0.01
Phonetic * Correct	−0.009	0.033	(−0.073, 0.054)	NS
Early * Correct	0.062	0.045	(−0.025, 0.150)	<0.10
Late * Correct	−0.132	0.046	(−0.223, −0.041)	NS
Spanish * Correct	−0.066	0.078	(−0.221, 0.091)	<0.01
Phonetic * Early * Spanish	0.045	0.164	(−0.265, 0.371)	<0.001
Phonetic * Late * Spanish	0.389	0.166	(0.058, 0.706)	<0.0001
Phonetic * Early * Correct	0.116	0.061	(−0.003, 0.237)	NS
Phonetic * Late * Correct	0.259	0.059	(0.144, 0.378)	<0.05
Phonetic * Spanish * Correct	0.029	0.088	(−0.150, 0.200)	<0.05
Early * Spanish * Correct	−0.192	0.156	(−0.499, 0.115)	<0.05
Late * Spanish * Correct	0.219	0.157	(−0.093, 0.520)	NS
Phonetic * Early * Spanish * Correct	−0.059	0.170	(−0.397, 0.267)	NS
Phonetic * Late * Spanish * Correct	−0.389	0.172	(−0.717, −0.046)	<0.01
Random effects		Variance		
Listener		0.366		
Stimulus Word		0.151		

Bilinguals also responded faster than Late Bilinguals (vs. Monolinguals: $\beta = 0.227$, posterior $SD = 0.100$, $p < 0.01$; vs. Early Bilinguals: $\beta = 0.176$, posterior $SD = 0.124$, $p < 0.05$), and there was again no difference between the Monolinguals and Early Bilinguals. For trials with Spanish phonetic cues, Monolinguals and Early Bilinguals responded faster than Late Bilinguals (vs. Monolinguals: $\beta = 0.320$, posterior $SD = 0.099$, $p < 0.0001$; vs. Early Bilinguals: $\beta = 0.258$, posterior $SD = 0.123$, $p < 0.01$), and there was no differences in RTs for the Monolinguals and Early Bilinguals. Finally, for nonce words with an English phonetic cue, Monolinguals and Early Bilinguals were also significantly faster than Late Bilinguals (vs. Monolinguals: $\beta = 0.294$, posterior $SD = 0.100$, $p < 0.0001$; vs. Early Bilinguals: $\beta = 0.182$, posterior $SD = 0.123$, $p < 0.05$), and Monolinguals trended faster than Early Bilinguals ($\beta = 0.112$, posterior $SD = 0.109$, $p = 0.06$).



Comparing Accurate and Inaccurate Trials

Overall, RTs for correct responses were faster than for incorrect responses. For Monolinguals, this difference reached significance for all four types of nonce words (English phonemic: $\beta = 0.178$, posterior $SD = 0.25$, $p < 0.01$; Spanish phonemic: $\beta = 0.244$, posterior $SD = 0.74$, $p < 0.01$; English phonetic: $\beta = 0.187$, posterior $SD = 0.023$, $p < 0.01$; Spanish phonetic: $\beta = 0.224$, posterior $SD = 0.035$, $p < 0.01$). For Early Bilinguals, correct trials were faster than incorrect trials for the Spanish cues (phonemic: $\beta = 0.374$, posterior $SD = 0.133$, $p < 0.0001$; phonetic: $\beta = 0.297$, posterior $SD = 0.052$, $p < 0.001$), but there was no difference for the English cues. For Late Bilinguals, the difference between correct and incorrect trials was significant for both kinds of Spanish cues (phonemic: $\beta = 0.157$, posterior $SD = 0.131$, $p < 0.05$; phonetic: $\beta = 0.267$, posterior $SD = 0.047$, $p < 0.01$) and for the English phonemes ($\beta = 0.310$, posterior $SD = 0.040$, $p < 0.001$), but not for the English phonetic cues.

The results of the accuracy and RT analyses are summarized in **Tables 9** and **10**. **Table 9** summarizes how Spanish and English stimuli were categorized by each listener group (A) and how the listeners categorized the different stimuli classes (B). **Table 10** summarizes how the listener groups compared within each stimulus type. The “=” is used to illustrate differences that were not significant, and the “>” and “<” indicate significant differences. The “ \gg ” and “ \ll ” represent differences that approached significance.

DISCUSSION

The current study tested the sensitivity of monolingual and early and late bilingual adults to language-specific sounds in a nonce-word categorization task to determine which segments listeners are most sensitive to and how language experience influences listeners’ sensitivity. Overall, listeners very accurately categorized phonemic cues and Spanish cues but struggled more with English cues and phonetic cues. There was also a significant interaction between stimulus language and cue type,

TABLE 9 | Summary of results from stimuli comparisons.

Accuracy		Reaction times
(A) Cross-language comparisons		
Monolinguals	Spanish phonemic > English phonemic	Spanish phonemic = English phonemic
Early Bilinguals	Spanish phonemic > English phonemic	Spanish phonemic = English phonemic
Late Bilinguals	Spanish phonemic > English phonemic	Spanish phonemic = English phonemic
(B) Cross-class comparisons		
Monolinguals	Spanish phonemic > Spanish phonetic English phonemic = English phonetic	Spanish phonemic = Spanish phonetic English phonemic = English phonetic
Early bilinguals	Spanish phonemic > Spanish phonetic English phonemic \gg English phonetic	Spanish phonemic \ll Spanish phonetic English phonemic = English phonetic
Late bilinguals	Spanish phonemic > Spanish phonetic English phonemic > English phonetic	Spanish phoneme < Spanish phonetic English phonemic = English phonetic

TABLE 10 | Summary of results from listener group comparisons.

	Accuracy	Reaction times
Spanish phonemes	Monolinguals = Early = Late	Monolinguals = Early < Late
English phonemes	Monolinguals = Early < Late	Monolinguals = Early < Late
Spanish phonetic	Monolinguals = Early = Late	Monolinguals = Early < Late
English phonetic	Monolinguals = Early = Late	Monolinguals \ll Early < Late

with the difference between phonemic and phonetic cues greater for Spanish than for English. This difference also significantly interacted with listener group, such that the difference between Spanish and English phonemic cues and Spanish and English phonetic cues was smaller for Late Bilinguals and greater for Early Bilinguals. The categorization accuracy of the Monolinguals, Early Bilinguals, and Late Bilinguals was very similar overall, with the only significant difference between groups occurring for the English phonemic cues, which Late Bilinguals categorized

more accurately than the other groups. The response times for Monolingual and Early Bilingual listeners were comparable, and both of these groups responded more quickly than Late Bilinguals for all cue types. Based on models of native and second-language speech perception (Flege, 1987, 1995; Best, 1991), we predicted a greater sensitivity to phonemic properties of lexical and language representations than to phonetic cues. The results here provide new evidence supporting these predictions in a language-decision task with word-length stimuli: early and late bilinguals can use both kinds of segments for categorization, but they were more sensitive to phonemic cues than phonetic cues. Unexpectedly, all listeners were more sensitive to Spanish-specific cues than English-specific cues. Finally, language background had only a limited effect on listeners' access to these representations.

Overall, there were no differences between the Monolingual and Early Bilingual listeners. The Late Bilinguals were as sensitive to some cues as the other two listener groups, and there was limited evidence that Late Bilinguals might even be more sensitive to some cues. The Late Bilinguals also responded significantly more slowly than the other groups, so it is possible that there was a speed-accuracy trade-off for these listeners; however, it only appeared for the Late Bilinguals' categorization of English phonemic cues, for which they were significantly more accurate than Monolinguals and Early Bilinguals but also significantly slower. The performance of the Monolinguals and Early Bilinguals reveals that the language representations of the Early Bilinguals, despite their having learned Spanish at home before English, do not differ in the phonemic categories or the phonetic detail encoded in their language representations. This is not to say that our Early Bilinguals would not have shown evidence of their Spanish exposure in other tests, such as production or phoneme identification tasks. The current results do suggest that the ability of Early Bilinguals to generalize about the properties of their native languages and associate phonological properties in particular with each language is not distinct from Monolinguals' awareness of these language-specific properties. This sets our early Spanish-English bilinguals apart from the early Spanish-Catalan bilinguals in Sebastián-Gallés et al. (2005), whose sensitivity to Catalan-specific contrasts was purportedly compromised by their early exposure to Spanish. Rather, the similarity between our responses from Monolinguals and Early Bilinguals supports the language assessment used by Amengual (2014, 2015), in which adults' current language exposure and use seem to override the effect of non-simultaneous early exposure and contribute to their equivalent performance (Gertken et al., 2014). The role of ongoing exposure in addition to and even superseding age of acquisition is also supported by Flege and colleagues who found that among listeners with similar ages of acquisition, greater exposure to, use of, and education in the L1 led to less native-like perception and production (Flege, 1991; Flege et al., 1997b; Flege and MacKay, 2004) and grammaticality judgments (Flege et al., 1999b) in the L2. It is important for future work on the association of language and segments to consider dominance and exposure to each language as factors influencing cross-linguistic speech perception in context.

While we only indirectly assessed the bilingual listeners' language dominance and exposure through the language

background questionnaire, the Monolingual and Early Bilingual groups did share some commonalities. Examining those further may assist in understanding the similarities in their categorization decisions and potentially why the Late Bilinguals outperformed these groups in the English phoneme trials. Our Early Bilinguals live and study immersed in their (chronological) L2, English, and as a result, they may have the same awareness of the generalizability of the phonological properties of each of their languages as the monolingual speakers who know only English. The difference between the two bilingual groups for the English phoneme category, on the other hand, may reflect variation in dominance, exposure, or the method of English acquisition. Most of the Early Bilinguals (11 of 18) learned English when they began kindergarten, and language instruction at this age is likely to be much less explicit than the middle and high school foreign-language classrooms in which the Late Bilinguals learned English. Even where there are parallels in L2 teaching at these ages, the experience of English language learning is much more recent for the Late Bilinguals than for the Early Bilinguals, and attending foreign language classes, practicing the language, and laboring to master the rules of and achieve proficiency in the L2 may lead the Late listeners to a greater metalinguistic awareness about properties of the language (Dąbrowska and Street, 2006), including increased sensitivity to language-segment associations. The study of phonological and metalinguistic awareness in adults has been limited to literacy and disorders (e.g., Pennington et al., 1990), although additional work with children has investigated bilingualism (Bruck and Genesee, 1995; Bialystok, 2001) and literacy development (e.g., Anthony and Francis, 2005). It is therefore unclear how metalinguistic awareness and cue sensitivity may affect cross-language speech perception in adults. The current findings suggest that the listeners who acquired an L2 in early childhood may lack the metalinguistic awareness evident in the Late Bilingual listeners, or that this sensitivity may decline into adulthood. Over time and as English proficiency increases, young bilingual listeners may lose their initial phonological sensitivity and may later categorize segments no differently than Monolingual adults who acquired their only language in infancy.

Given the potential differences in language teaching and language learning in kindergarten and high school, the Late Bilinguals may have increased sensitivity to some language-specific phonological properties due to the circumstances of their bilingualism and not necessarily due to the age of acquisition. In fact, this formal training may also explain why there were group differences for the English phonemic cues but not for the English phonetic ones. Phonemic differences across languages may get more attention in foreign-language classes than subsegmental differences between categories shared by the two languages. Just as the phonetic cues were more difficult for listeners in general, Late Bilinguals may not have had the same metalinguistic instruction about English phonetic differences and so may have been less able to associate those cues with English, even though this was possible for the phonemic cues. Future work on cue sensitivity should work to separate recency of language acquisition from method of language acquisition to disentangle how these factors influence phonological awareness

and especially awareness of subsegmental differences. For example, Early Bilinguals may be more sensitive to English phonemic cues during earlier stages of English acquisition, and we might also expect listeners who acquire a language without formal classes (e.g., from being immersed in a new community) to be less sensitive to language-specific cues, especially phonemes, than listeners who study the language in a formal setting.

The consistency of categorization accuracy across the three listener groups suggests that language experience was less important than cue salience in this task. Phonemic cues were more accurately categorized than phonetic cues, for both English and Spanish, supporting the parallel distinction made between new and similar phones in Flege (1987, 1995)'s Speech Learning Model (SLM). In this model, second language learners create independent categories for sounds judged to be "new" (unique to the L2 and not present in the L1), which facilitates the production and perception of such sounds. Phones that are recognized as similar to existing L1 segments are discriminated less well if no new category is established for them. The phonemes in the present task may be like the SLM's new phones, even for the Monolinguals who have not acquired Spanish, and as such they are immediately recognizable as language-specific sounds (Best, 1991), which leads to more accurate categorization. In contrast, the phonetic cues pattern like the SLM's similar phones, a category for which, according to Best (1991), the L2 or non-dominant language sounds would be mapped to the L1 or dominant-language categories. This would cause more competition in deciding between English or Spanish for the language identity of the word.

There may have also been an effect of the specific segments included in each category. Since there was only one Spanish-specific phonemic cue included, the Spanish phoneme category in fact represents listener responses to a single sound, the Spanish trill /r/, which was easily perceived and strongly associated with Spanish phonology for all three listener groups. The English phoneme category may have been very different in this sense, since it included the English rhotic /ɹ/ and the interdental fricative /θ/. Fricatives and interdentals in particular are acquired late by English-learning children (Clark, 2003; Dodd et al., 2003), and even native-English-speaking adults are susceptible to mishearing /θ/ more than they mishear other segments (Cutler et al., 2004). That is, there may be inherent differences in the perceptual salience of the two English phonemes, irrespective of the strengths of associations between English and each segment. Since only a single Spanish phonemic cue was available and given the asymmetry in salience of the English phonemic cues, future work should more systematically compare a wider range of phonemes in other language pairs to consider whether there may be variability within the phonemic category. However, despite the inherent difficulty of at least the English /θ/, it is even more striking that the Late Bilinguals outperformed the groups that had acquired the English phonemes in childhood. In fact, since the Late Bilinguals may be aware of /θ/ being a phonemic sound in Peninsular Spanish, we might have expected this awareness to cause confusion and thus fewer accurate responses in English phoneme trials for the Late Bilinguals, but just the opposite was the case. This suggests that the absence of this phoneme

in the native language and dialects of the Late Bilinguals may have heightened their sensitivity to /θ/. Instead, the difficulty all listeners had responding to the English phoneme category may be motivated by perceptual salience more generally, and future work should further probe variation with each of these cue types.

The difficulty listeners from all backgrounds experienced in accurately categorizing phonetic cues also requires further investigation. The English [ɫ] is more velarized, i.e., produced with the tongue further back in the oral cavity, than the Spanish [l], while the English [ʉ] is fronted, so the difference between English and Spanish phonetic cues is unlikely to be due to a single property that sets English apart from Spanish, since the English variants differ in opposite directions from the Spanish ones. It may be that listeners hear more variation in English input between lighter or darker /l/ and more or less fronted /u/ across dialects, speakers, and phonological contexts than exists for Spanish [l] and [u]. However, it would be surprising if our monolingual English listeners were also sensitive to the greater consistency of these segments in Spanish, given their lack of exposure to the language.⁷ Furthermore, if the variability present in the realization of these sounds in English motivated the difference in accuracy between English and Spanish segments, we should expect a different categorization pattern entirely. A light [l] or a backed [u] may be either from Spanish or English, since these variants exist in many dialects of English, so the Spanish phonetic cues should have received responses more mixed between the languages. It is the darker [ɫ] and fronted [ʉ] that should be unambiguously associated with English, but in fact we find the English cues receive more of a mix of Spanish and English categorization decisions while the Spanish cues are relatively consistently identified as Spanish.

While every effort was made to create nonce words that were equally plausible in both languages, except for the language-specific target segment, the naturally produced stimuli used here inevitably carried additional indicators of language. The phonotactic restrictions of Spanish may have meant that the CVCV stimuli were simply more Spanish-like than English-like, even though this word structure is permitted in English. The Spanish-ness of these stimuli is supported by the reactions of participants in two pilot studies; in the first pilot, theoretically congruous stimuli that overlapped English and Spanish in all segments, e.g., /tʃima/, were categorized as Spanish significantly more than English, and in the second pilot (cf. Procedure), listeners reported confusion about whether words were English or English-accented Spanish. In the present study, listeners from all three language backgrounds were able to overcome this potential bias toward Spanish for English: the log odds of responding correctly were significantly above 0 (chance performance) in all four cases, including for the English segments. Therefore, listeners showed sensitivity to the English-ness of the English cues even if the word structure is less common in English than it is in Spanish. Furthermore, Monolinguals might not be expected to suffer from such a potential bias as

⁷We would additionally have to assume that exposure to Spanish-accented English is sufficient for the development of phonological categories that accurately reflect the properties of these categories as they are realized in Spanish.

much as the bilingual groups, since the Monolinguals do not have representations of Spanish phonotactics against which to judge the nonce word forms. Instead, their categorization patterns were in line with the bilingual groups'. Why, then, might listeners have been less accurate in categorizing stimuli with English cues?

The difficulties that persisted for English cues are especially interesting given that the naturally produced nonce words used here likely contained multiple phonetic cues to language. As was mentioned in the discussion of nonce words, the disyllabic nature of the nonce words meant that the unstressed vowel /a/ in the second syllable was reduced to [ə] in the English words; therefore, all the English nonce words contained both a language-specific target segment (e.g., /ɪ/) and the reduced vowel. Furthermore, the acoustic analyses of the /i/ and /a/ vowels in the first syllable of the nonce words indicate that there were also language-specific differences in the productions of these non-target segment (cf. Acoustic Analyses). But again, despite these potential additional cues to language, listeners categorized the English-specific segments less accurately than Spanish cues. Given the more accurate performance of the Late Bilinguals than the other groups for English phonemes we might be tempted to conclude that the Late Bilinguals were better able to use these supplementary language-specific cues than their peers, but their accuracy did not significantly differ from the Monolinguals and Early Bilinguals in the English phonetic condition. If the Late Bilinguals were more sensitive to the English-ness of the nonce word filler vowels in the phonemic condition, where they outperformed their peers, it is unclear why they wouldn't have been able to make use of the additional cues in the English phonetic words.

Moving forward, it will continue to be important to consider the contributions of language-specific segments in the context of a word, as discussed earlier, since listeners may use different processing strategies and respond to the same sound categories differently when presented in isolation and in context. To this end, it will be necessary to also involve language pairs for which there are more language-specific contrasts and a wider variety of segments to be studied than those available for English and Spanish. All phonemic cues used here were consonants, with a necessary but confounding overreliance on the differences in rhotics across the languages. Similarly, the mispronunciation studies in Spanish and Catalan by Sebastián-Gallés et al. (2005) and Amengual (2014, 2015) were restricted in scope, and focused only on vowels. Contrasting a language pair that differs more significantly in both consonants and vowels at the phonemic and phonetic levels would provide the evidence needed to further test the conclusions drawn from the present results.

Finally, the current study speaks to other related speech perception phenomena, namely foreign-accent detection. To date, our knowledge of the perception of foreign-accented speech has been largely based on monolingual listeners, but the findings of the present study support the inclusion of listeners actually proficient in, and not just familiar with, the L1 of the accented speech. Based on our results, bilingual listeners might be expected to identify accented talkers as well as monolingual listeners, and if the foreign accent contains non-native phonemic cues like those tested here, late bilinguals might be more sensitive

to accented speech than other listeners. Benefits of exposure to accented speech have likewise been reported for categorizing sentences produced in regional (Clopper and Pisoni, 2004, 2007) and foreign (Vieru et al., 2011) accents. High-exposure listeners also processed foreign-accented words faster and more accurately than low-exposure listeners (Witteman et al., 2013), so listeners with experience can attend to the relatively few cues available in a single word. Even so, given the nature of the naturally produced words and sentences used in these studies, it is not clear what cues the listeners with greater exposure were using in their processing, or which cues the less-experienced listeners were not able to capitalize on. We might expect foreign-accented speech to contain more of the difficult phonetic cues that most challenged our Monolingual listeners, and this could explain the performance of the low-familiarity listeners in Vieru et al. (2011) and Witteman et al. (2013). The contribution of phonemic and phonetic cues to foreign-accented speech detection could be tested by controlling these cues in real words, as was done in the present study with nonce words, to determine if real foreign-accented words with deviant phonemic cues are in fact categorized more easily than words with phonetic cues. Furthermore, the processing of foreign-accented speech may also be influenced by the presence of phonemic and phonetic cues. Since phonetic cues are less clearly linked to a specific language and listeners of all backgrounds are less sensitive to deviations in phonetic cues, speech that contains only phonetic deviations (e.g., from more proficient L2 speakers) may be easier to process than speech that also contains phonemic deviations.

In summary, the results of the nonce-word categorization task indicate that listeners are better able to use Spanish-specific cues than English-specific cues and that listeners categorize phonemic cues, modeled on Flege's (1987, 1995) "new" sounds, better than phonetic cues. This distinction supports similar divisions made between native and non-native sounds in speech perception literature more generally and for second language acquisition in particular (Flege, 1987, 1995; Best, 1991). Our findings also show similarities in categorization patterns across listener groups, in parallel with the work of Mack (1989) and Flege et al. (1999a) on early bilinguals' phoneme discrimination, and even the late bilinguals categorized the nonce-word stimuli like early learners. The early bilinguals' sensitivity to English-specific cues was not degraded by their early exposure to and proficiency in Spanish, deviating from the conclusions of Sebastián-Gallés et al. (2005), but their knowledge of Spanish also did not improve the accuracy of their language classification decisions for Spanish nonce words, which might have been expected given the advantages for high-exposure listeners in accent categorization tasks (e.g., Witteman et al., 2013). Such facilitation was observed for the late bilinguals for words with English phonemic cues, although the late bilingual listeners responded significantly more slowly than the other groups for all cues. The study of additional language pairs will strengthen the conclusions we make here about differences in listener sensitivity to language-specific phonemic and phonetic cues by providing additional segments and contrasts and allowing for systematic comparisons, e.g., of consonantal and vowel contributions to each category. The finding that listeners use phonemic cues more

successfully than phonetic cues in word contexts should shape future directions of work on the perception of foreign-accented speech and cross-language speech perception.

AUTHOR CONTRIBUTIONS

CPB, CB, and RS jointly designed the project. CPB and RS designed the stimuli, and CPB and CB worked together to analyze the data. All three authors were responsible for interpreting the results. CPB wrote the manuscript draft, and CB and RS added considerable critical commentary and revisions. All three have approved this manuscript and are accountable for all aspects of the work regarding accuracy and integrity.

REFERENCES

- Allen, J. S., and Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *J. Acoust. Soc. Am.* 115, 3171–3183.
- Amengual, M. (2014). The perception and production of language-specific mid-vowel contrasts: shifting the focus to the bilingual individual in early language input conditions. *Int. J. Biling.* 20, 133–152. doi: 10.1177/1367006914544988
- Amengual, M. (2015). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Appl. Psycholinguist.* 1–31. doi: 10.1017/S0142716415000557
- Anthony, J. L., and Francis, D. J. (2005). Development of phonological awareness. *Curr. Dir. Psychol. Sci.* 14, 255–259. doi: 10.1111/j.0963-7214.2005.00376.x
- Best, C. T. (1991). The emergence of native-language phonological influences in infants: a perceptual assimilation model. *Haskins Lab. Status Rep. Speech Res.* SR-107/108, 1–30.
- Bialystok, E. (2001). Metalinguistic aspects of bilingual processing. *Annu. Rev. Appl. Linguist.* 21, 169–181.
- Boersma, P., and Weenink, D. (2012). *Praat: Doing Phonetics by Computer [Computer Program]*. Version 5.3.32. Available at: <http://www.praat.org/>
- Bradlow, A. (1995). A comparative acoustic study of English and Spanish vowels. *J. Acoust. Soc. Am.* 97, 1916–1924. doi: 10.1121/1.412064
- Bradlow, A., Clopper, C., Smiljanic, R., and Walter, M. A. (2010). A perceptual phonetic similarity space for languages: evidence from five native language listener groups. *Speech Commun.* 52, 930–942. doi: 10.1016/j.specom.2010.06.003
- Bruck, M., and Genesee, F. (1995). Phonological awareness in young second language learners. *J. Child Lang.* 22, 307–324. doi: 10.1017/S0305000900009806
- Carlson, M. T., Goldrick, M., Blasingame, M., and Fink, A. (2015). Navigating conflicting phonotactic constraints in bilingual speech perception. *Biling. Lang. Cogn.* 1–16. doi: 10.1017/S1366728915000334
- Chan, C. L. (2014). *NUSub-db: Northwestern University Subject Database [Web Application]*. Evanston, IL: Northwestern University.
- Clark, E. V. (2003). *First Language Acquisition*. Cambridge: Cambridge University Press.
- Clopper, C. G., and Pisoni, D. B. (2004). Homebodies and army brats: some effects of early linguistic experience and residential history on dialect categorization. *Lang. Var. Change* 16, 31–48. doi: 10.1017/S0954394504161036
- Clopper, C. G., and Pisoni, D. B. (2007). Free classification of regional dialects of American English. *J. Phon.* 35, 421–438. doi: 10.1016/j.wocn.2006.06.001
- Clopper, C. G., Pisoni, D. B., and de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *J. Acoust. Soc. Am.* 118, 1661–1676. doi: 10.1121/1.2000774
- Costa, A., and Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Biling. Lang. Cogn.* 2, 231–244. doi: 10.1017/S1366728999000334
- Cutler, A., Weber, A., Smits, R., and Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *J. Acoust. Soc. Am.* 116, 3668–3678. doi: 10.1121/1.1810292
- Dąbrowska, E., and Street, J. A. (2006). Individual differences in language attainment: comprehension of passive sentences by native and non-native English speakers. *Lang. Sci.* 28, 604–615. doi: 10.1016/j.langsci.2005.11.014
- Diehl, R. L., and Walsh, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *J. Acoust. Soc. Am.* 85, 2154–2164. doi: 10.1121/1.397864
- Dodd, B., Holm, A., Hua, Z., and Crosbie, S. (2003). Phonological development: a normative study of British English-speaking children. *Clin. Linguist. Phon.* 17, 617–643. doi: 10.1080/0269920031000111348
- Escudero, P. (2006). “The phonological and phonetic development of new vowel contrasts in Spanish learners of English,” in *English With a Latin Beat: Studies in Portuguese/Spanish-English Interphonology, Studies in Bilingualism*, Vol. 31, eds B. O. Baptista and M. A. Watkins (Amsterdam: John Benjamins), 149–161.
- Flege, J. E. (1984). The detection of French accent by American listeners. *J. Acoust. Soc. Am.* 76, 692–707. doi: 10.1121/1.391256
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *J. Phon.* 15, 47–55.
- Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *J. Acoust. Soc. Am.* 89, 395–411. doi: 10.1121/1.400473
- Flege, J. E. (1995). “Second language speech learning: theory, findings, and problems,” in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. W. Strange (Timonium, MD: York Press), 233–277.
- Flege, J. E. (2007). “Language contact in bilingualism: phonetic system interactions,” in *Laboratory Phonology 9*, eds J. Cole and J. Hualde (Berlin: Mouton de Gruyter), 353–382.
- Flege, J. E., Bohn, O.-S., and Jang, S. (1997a). Effects of experience on non-native speakers’ production and perception of English vowels. *J. Phon.* 25, 437–470. doi: 10.1121/1.3466879
- Flege, J. E., Frieda, E. M., and Nozawa, T. (1997b). Amount of native-language (L1) use affects the pronunciation of an L2. *J. Phon.* 25, 169–186. doi: 10.1006/jpho.1996.0040
- Flege, J. E., and MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Stud. Second Lang. Acquis.* 26, 1–34. doi: 10.1159/000235659
- Flege, J. E., MacKay, I. R. A., and Meador, D. (1999a). Native Italian speakers’ perception and production of English vowels. *J. Acoust. Soc. Am.* 106, 2973–2987. doi: 10.1121/1.428116
- Flege, J. E., and Munro, M. J. (1994). The word unit in second language speech production and perception. *Stud. Second Lang. Acquis.* 16, 381–411. doi: 10.1017/S0272263100013437
- Flege, J. E., Munro, M. J., and Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *J. Acoust. Soc. Am.* 95, 3623–3641. doi: 10.1121/1.409931
- Flege, J. E., Yeni-Komshian, G. H., and Liu, S. (1999b). Age constraints on second-language acquisition. *J. Mem. Lang.* 41, 78–104. doi: 10.1006/jmla.1999.2638

ACKNOWLEDGMENTS

Funding for this work was provided in part by the Department of Linguistics at the University of Texas at Austin, a Carlota Smith Fellowship, and a Harrington Dissertation Fellowship awarded to the first author. The authors would like to thank the UT Sound Lab research assistants for their assistance with data collection, Sally Amen for her help analyzing the results of the accentedness rating study, and participants in the Acoustical Society of America’s biannual meetings for their thoughtful feedback as this project developed. We are also grateful for the insights of two reviewers and the editor. All errors remain our own.

- Gelman, A., Lee, D., and Guo, J. (2015). A probabilistic programming language for bayesian inference and optimization. *J. Educ. Behav. Stat.* 40, 530–543. doi: 10.3102/1076998615606113
- Gertken, L. M., Amengual, M., and Birdsong, D. (2014). “Assessing language dominance with the Bilingual Language Profile,” in *Measuring L2 proficiency: Perspectives from SLA*, eds P. Leclercq, A. Edmonds, and H. Hilton (Bristol: Multilingual Matters), 208–225.
- Harnsberger, J. D., Shrivastav, R., Brown, W. S. Jr., Rothman, H., and Hollien, H. (1997). Speaking rate and fundamental frequency as speech cues to perceived age. *J. Voice* 22, 58–69. doi: 10.1016/j.jvoice.2006.07.004
- Hualde, J. I. (2005). *The Sounds of Spanish*. New York, NY: Cambridge University Press.
- Johnson, K. (1997). “Speech perception without speaker normalization,” in *Talker Variability in Speech Processing*, eds K. Johnson and J. W. Mullennix (New York, NY: Academic Press), 145–166.
- Klatt, D. H., and Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* 87, 820–857. doi: 10.1121/1.398894
- Kondaurova, M. V., and Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *J. Acoust. Soc. Am.* 124, 3959–3971. doi: 10.1121/1.2999341
- Kraljic, T., and Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychon. Bull. Rev.* 13, 262–268. doi: 10.3758/BF03193841
- Kraljic, T., and Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *J. Mem. Lang.* 56, 1–15. doi: 10.3758/s13414-015-0856-y
- Mack, M. (1989). Consonant and vowel perception and production: early English-French bilinguals and English monolinguals. *Percept. Psychophys.* 46, 187–200. doi: 10.3758/BF03204982
- Mendez, A. (1982). Production of American English and Spanish vowels. *Lang. Speech* 25, 191–197. doi: 10.1177/002383098202500206
- Munro, M. J., and Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 Speech: the role of speaking rate. *Stud. Second Lang. Acquis.* 23, 451–468.
- Nygaard, L. C., and Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Percept. Psychophys.* 60, 355–376. doi: 10.3758/BF03206860
- Pennington, B. F., van Orden, G. C., Smith, S. D., Green, P. A., and Haith, M. M. (1990). Phonological processing skills and deficits in adult dyslexics. *Child Dev.* 61, 1753–1778. doi: 10.2307/1130836
- Pierrehumbert, J. B. (2002). “Word-specific phonetics,” in *Laboratory Phonology*, Vol. 7, eds C. Gussenhoven and N. Warner (Berlin: Mouton de Gruyter), 101–139.
- Pisoni, D. B. (1977). Identification and discrimination of the relative voice onset time of two component tones: implication for voicing perception in stops. *J. Acoust. Soc. Am.* 61, 1352–1361. doi: 10.1121/1.381409
- Psychology Software Tools (2010). *Psychology Software Tools, Inc. E-Prime 2.0*. Available at: <http://www.pstnet.com>
- Recasens, D. (2004). Darkness in [l] as a scalar phonetic property: implications for phonology and articulatory control. *Clin. Linguist. Phon.* 18, 593–603. doi: 10.1080/02699200410001703556
- Recasens, D. (2012). A cross-language acoustic study of initial and final allophones of [l]. *Speech Commun.* 54, 368–383. doi: 10.1016/j.specom.2011.10.001
- RStudio Team (2015). *RStudio: Integrated Development for R*. RStudio, Inc. Boston, MA: RStudio Team. Available at: <http://www.rstudio.com/>
- Sebastián-Gallés, N., Echeverría, S., and Bosch, L. (2005). The influence of initial exposure on lexical representation: comparing early and simultaneous bilinguals. *J. Mem. Lang.* 52, 240–255. doi: 10.1016/j.jml.2004.11.001
- Sebastián-Gallés, N., and Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition* 72, 111–123. doi: 10.1016/S0010-0277(99)00024-4
- Shook, A., and Marian, V. (2013). The Bilingual language interaction network for comprehension of speech. *Biling. Lang. Cogn.* 16, 304–324. doi: 10.1017/S1366728912000466
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *J. R. Stat. Soc. Ser. B* 64, 583–639. doi: 10.1111/1467-9868.00353
- Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *J. Lang. Soc. Psychol.* 18, 86–99. doi: 10.1177/0261927X99018001006
- Strand, E. A., and Johnson, K. (1996). “Gradient and visual speaker normalization in the perception of fricatives,” in *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference*, Bielefeld, ed. D. Gibbon (Berlin: Mouton), 14–26.
- Tracy, E. C., Bainter, S. A., and Satariano, N. P. (2015). Judgments of self-identified gay and heterosexual male speakers: which phonemes are the most salient in determining sexual orientation? *J. Phon.* 52, 13–25. doi: 10.1016/j.wocn.2015.04.001
- Vieru, B., Boula de Mareüil, P., and Adda-Decker, M. (2011). Characterisation and identification of non-native French accents. *Speech Commun.* 53, 292–310. doi: 10.1016/j.specom.2010.10.002
- Witteman, M. J., Weber, A., and McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Atten. Percept. Psychophys.* 75, 537–556. doi: 10.3758/s13414-012-0404-y

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Blanco, Bannard and Smiljanic. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

To ensure that the stimuli talker's productions were native-like in both languages, an accentedness rating study was completed. Native English and native Spanish listeners rated the nativeness of the productions of eight talkers, including the stimuli talker. All talkers recorded *Æsop's The North Wind and the Sun* in Spanish and English, and the final set of talkers included one male and one female from each of the following four groups: monolingual English talkers, L1 English talkers who learned Spanish late and had completed college and graduate coursework in Spanish, L1 Spanish talkers from Latin America who learned English late and had moved to the U.S. to attend college, and early Spanish-English bilinguals (including the stimuli talker). The recordings from these eight talkers were divided into seven phrases, yielding 56 sound files of the talkers' English and 56 sound files of their Spanish.

The raters included ten monolingual English listeners and 10 L1 Spanish listeners from Latin America who learned English after age 14. None participated in the main study. Raters heard productions in their native language and decided how native- or foreign-sounding each production was by using the mouse to click on a horizontal line. The line appeared on the screen after the audio presentation of each sentence and represented a continuum between "Perfectly native sounding" (labeled as such at the left extreme) and "Very foreign sounding" (so labeled at the right extreme). The Spanish translations "Suena totalmente nativo" and "No suena nada nativo" were used in the Spanish version with the native Spanish listeners and the talkers' Spanish productions. The accentedness rating was recorded as the x-intercept of the mouse at the click. The 56 sentences were randomized for each listener.

Accentedness ratings were converted to z -scores to account for listeners using the continua differently, and the z -transformed accentedness ratings for English and Spanish productions were submitted to separate mixed-effects linear regression models using the *lme4* (v1.1-7) and *lmerTest* (v2.0-20) packages in R (RStudio 0.99.489; RStudio Team, 2015). Listener was included as a random intercept, and testing talker as a fixed effect significantly improved the fit of a model with the random intercept alone, for both the English model ($\chi^2 = 1317.3$, $df = 7$, $p < 0.001$) and the Spanish model ($\chi^2 = 948.25$, $df = 7$, $p < 0.001$). See **Table A1** for the model summaries. The stimuli talker (early bilingual male) was designated as the referent class for the talker variable. The intercept for the stimuli talker was significantly less than zero ($p < 0.001$) in both the English and Spanish models and was thus significantly closer to the "Perfectly native sounding" extreme than to the center for both languages. The

TABLE A1 | Model summaries for mixed-effects linear regression models predicting accentedness ratings.

Predictor	Estimate	Standard error	t-value	p-value
(A) English productions				
Intercept (Stimuli talker)	−0.632	0.037	−17.186	<0.001
Monolingual male	0.0554	0.052	1.065	0.29
Monolingual female	−0.131	0.052	−2.516	<0.05
L1 English male	−0.082	0.052	−1.575	0.12
L1 English female	0.163	0.052	3.127	<0.01
Early bilingual female	0.613	0.052	11.785	<0.001
L1 Spanish male	2.123	0.052	40.797	<0.001
L1 Spanish female	2.318	0.052	44.537	<0.01
Random effects		Variance		
Listener	<0.001			
Residual	0.095			
(B) Spanish productions				
Intercept (Stimuli talker)	−0.873	0.051	−17.062	<0.001
Monolingual male	2.272	0.072	31.404	<0.001
Monolingual female	2.241	0.072	30.970	<0.001
L1 English male	1.292	0.072	17.861	<0.001
L1 English female	0.661	0.072	9.144	<0.001
Early bilingual female	0.458	0.072	6.323	<0.001
L1 Spanish male	−0.018	0.072	−0.255	0.80
L1 Spanish female	0.077	0.072	1.070	0.29
Random effects		Variance		
Listener	<0.001			
Residual	0.183			

stimuli talker's English was not rated as significantly different from the monolingual English male ($p = 0.29$) or the L1 English male ($p = 0.12$), and he was rated as significantly more native sounding than all other talkers (at least $p < 0.01$) except the monolingual English female ($p < 0.05$).⁸ The stimuli talker's Spanish was also rated as significantly more native sounding than all the other talkers ($p < 0.001$), except for the L1 Spanish male and female, with whom there was no significant difference in rating (for L1 Spanish male, $p = 0.80$; for L1 Spanish female, $p = 0.29$).

⁸ The monolingual English female was also rated as significantly more native sounding than the monolingual English male ($p < 0.001$) and the L1 English female ($p < 0.001$), who were also raised as monolingual English speakers. The speed with which the monolingual English female read the story may have influenced how accented she was rated (cf. Munro and Derwing, 2001), but importantly, the stimuli talker's accent in English was not rated different from two male talkers who grew up as monolingual English speakers.



Establishing New Mappings between Familiar Phones: Neural and Behavioral Evidence for Early Automatic Processing of Nonnative Contrasts

Shannon L. Barrios^{1*}, Anna M. Namyst², Ellen F. Lau², Naomi H. Feldman^{2,3} and William J. Idsardi²

¹ Department of Linguistics, University of Utah, Salt Lake City, Utah, USA, ² Department of Linguistics, University of Maryland, College Park, College Park, MD, USA, ³ Institute for Advanced Computer Studies, University of Maryland, College Park, College Park, MD, USA

OPEN ACCESS

Edited by:

Annie Tremblay,
University of Kansas, USA

Reviewed by:

Wendy Herd,
Mississippi State University, USA
Christine E. Shea,
University of Iowa, USA
Adrian Garcia-Sierra,
University of Connecticut, USA

*Correspondence:

Shannon L. Barrios
s.barrios@utah.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 30 January 2016

Accepted: 16 June 2016

Published: 30 June 2016

Citation:

Barrios SL, Namyst AM, Lau EF,
Feldman NH and Idsardi WJ (2016)
Establishing New Mappings between
Familiar Phones: Neural and
Behavioral Evidence for Early
Automatic Processing of Nonnative
Contrasts. *Front. Psychol.* 7:995.
doi: 10.3389/fpsyg.2016.00995

To attain native-like competence, second language (L2) learners must establish mappings between familiar speech sounds and new phoneme categories. For example, Spanish learners of English must learn that [d] and [ð], which are allophones of the same phoneme in Spanish, can distinguish meaning in English (i.e., /deɪ/ “day” and /ðeɪ/ “they”). Because adult listeners are less sensitive to allophonic than phonemic contrasts in their native language (L1), novel target language contrasts between L1 allophones may pose special difficulty for L2 learners. We investigate whether advanced Spanish late-learners of English overcome native language mappings to establish new phonological relations between familiar phones. We report behavioral and magnetoencephalographic (MEG) evidence from two experiments that measured the sensitivity and pre-attentive processing of three listener groups (L1 English, L1 Spanish, and advanced Spanish late-learners of English) to differences between three nonword stimulus pairs ([idi]-[iði], [idi]-[iri], and [iði]-[iri]) which differ in phones that play a different functional role in Spanish and English. Spanish and English listeners demonstrated greater sensitivity (larger d’ scores) for nonword pairs distinguished by phonemic than by allophonic contrasts, mirroring previous findings. Spanish late-learners demonstrated sensitivity (large d’ scores and MMN responses) to all three contrasts, suggesting that these L2 learners may have established a novel [d]-[ð] contrast despite the phonological relatedness of these sounds in the L1. Our results suggest that phonological relatedness influences perceived similarity, as evidenced by the results of the native speaker groups, but may not cause persistent difficulty for advanced L2 learners. Instead, L2 learners are able to use cues that are present in their input to establish new mappings between familiar phones.

Keywords: L1 Spanish, L2 English, L1 allophones, novel contrasts, MMN, allophonic split, perceptual categorization, phonological status

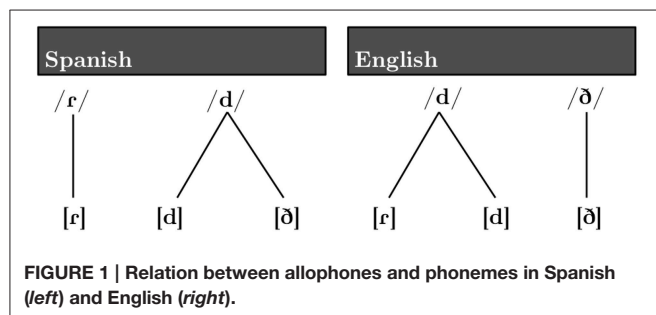
INTRODUCTION

Linguistic experience shapes listeners' sensitivities to phonetic distinctions. Specifically, extensive experience with one's native language (coupled with a lack of experience with nonnative sounds and contrasts) limits listeners' sensitivity to nonnative phonemic distinctions (Lisker and Abramson, 1970; Goto, 1971; Werker et al., 1981; Näätänen et al., 1997, to name just a few). This differential sensitivity to native vs. nonnative speech contrasts develops very early in life (Werker and Tees, 1984; Kuhl et al., 1992; Polka and Werker, 1994), and shapes the initial stages of second language (L2) speech perception (Escudero, 2005; Best and Tyler, 2007). These findings, and many others like them, have led to the development of models of cross-language and L2 speech perception and production (Best, 1995; Flege, 1995; Iverson et al., 2003; Escudero, 2005; Best and Tyler, 2007) which make predictions about how naive nonnative and L2 listeners will perceive and acquire target language sounds and contrasts. More recently, however, there has been growing interest in how allophones (i.e., phones which are present in the ambient language, but which are not used to distinguish word meanings) are represented and processed by adults (Kazanina et al., 2006; Boomershine et al., 2008; Johnson and Babel, 2010), and how this knowledge of phonological status develops in infants (Seidl and Cristia, 2012).

The present study contributes to this literature on sound category learning by investigating the role of language-specific phonological patterning in L2 phonological development. We use both behavioral methods and magnetoencephalographic (MEG) recordings to investigate how adult second language learners' knowledge of native language (L1) phonological patterns impacts the acquisition of their second language sound system. In particular, we ask whether advanced adult late-learners of a second language overcome native language mappings to establish new phonological relations between familiar phones.

Languages differ in their mappings between predictable surface variants (i.e., allophones) and more abstract phonological categories (i.e., phonemes) (Kenstowicz, 1994). Consider, for example, the relation between the phonological systems of Spanish and English (Figure 1), in which sets of sound categories with very similar acoustic distributions map onto different sets of phonemes in the two languages.

Although three very similar phonetic categories, [d], [ð], and [r], exist in both Spanish and English, the functional significance of these categories varies between the two languages.



The phones [d] and [ð] distinguish word meanings in English (i.e., [ðei] “they” and [dei] “day”). In contrast, a productive phonological pattern causes the voiced obstruents /b, d, g/ to surface as the approximants [β, ð, ɣ] intervocalically in Spanish¹. Thus, whereas [d] and [ð] are contrastive in English, the two distinct acoustic realizations are phonologically conditioned variants (allophones or positional variants) of the same phoneme category in Spanish. An important component of native speakers' knowledge of an allophonic alternation of this sort is that allophonic variants are tied to particular phonological contexts, whereas the phoneme is not. Thus, native Spanish speakers have internalized knowledge of the contexts in which the allophonic variants [d] and [ð] occur. While the exact pattern of allophony is known to vary by Spanish dialect (See Carrasco et al., 2012 for a review of this literature, as well as acoustic analyses characterizing the differences between Costa Rican and Madrid varieties of Spanish), the approximant (rather than the stop) is expected intervocalically in all dialects. On the other hand, the phones [d] and [r] are contrastive in Spanish, but not in English. In American English /d/ (and /t/) surface as [r] in post-tonic intervocalic position, and [d] elsewhere (i.e., [ˈaɪrːd] “ride” vs. [ˈaɪrːɹ̩] “rider”)².

A consequence of cross-linguistic variation in the mapping between speech sounds and phonemes is that L2 learners may need to establish new mappings between familiar phones. For example, to attain native-like competence in English, a Spanish learner must learn that [d] and [ð], which are allophones of a single phoneme (i.e., /d/) in Spanish, can distinguish word meaning in English. Doing so is assumed to entail the updating of internalized knowledge about the distribution of the phones in the L2 (i.e., learning that the phones are not restricted to particular environments in the target language, but instead can occur in the same phonological environments). Eckman et al. (2001; 2003, and subsequent work) referred to this learning scenario, in which sounds that are allophones of one phoneme in a learner's native language constitute separate phonemes in the target language, as an ‘allophonic split’³. It is this L2 learning scenario that is the focus of the present study⁴.

L1 context-dependent allophones present unique challenges for the L2 learner from the perspective of production and perception. The learner must learn to detect the target language phonemic contrasts in perception, and suppress L1 positional variants in L2 production (even when the phonological context is appropriate for their production). Both anecdotal and experimental evidence from speech production (Lado, 1957; Hammerly, 1982; Hardy, 1993; Zampini, 1996; Eckman et al.,

¹Waltmunson (2005) reports that the intervocalic spirantization of /d/ occurs 99% of the time.

²Patterson and Connine (2001) report that the flapped variant occurs 94% of the time in its conditioning environment.

³It is worth noting that to attain truly native-like competence in English, the Spanish learner must also learn to treat the phones [d] and [r] as allophones of the same phoneme in English. Since this would involve the joining of L1 allophones, we might call this learning scenario ‘allophonic union.’

⁴Other related work in L2 phonology has investigated the acquisition of positional variants in the target language by L2 learners in production (Zampini, 1994; Shea and Curtin, 2011) and perception (Shea and Curtin, 2010; Shea and Renaud, 2014).

2001, 2003) suggests that this learning situation presents considerable difficulty for second language learners. However, the acquisition of novel target language contrasts between L1 context-dependent allophones has not been adequately explored from the perspective of L2 speech perception.

Research with adult native listeners has revealed that speech perception is not only influenced by listeners' experience (or lack of experience) with the phones in question; the phonological status of a sound contrast also affects listeners' perception. Several behavioral studies have reported differences in the perception of familiar phones (i.e., phones that occur regularly in the native language of the listener) depending on whether the sounds in the pair function as contrastive phonemes or non-contrastive allophones in the listener's native language. In particular, these studies report that sounds which are contrastive are discriminated more readily, and are rated less perceptually similar than allophonically related phones (Pegg and Werker, 1997; Whalen et al., 1997; Harnsberger, 2001; Peperkamp et al., 2003; Boomershine et al., 2008). For example, Whalen et al. (1997) used a categorical AXB task to investigate the discriminability of the phones [p], [p^h], and [b] by adult English listeners. They found that English listeners more readily discriminated the distinction between phonemic contrasts [b]-[p] and [b]-[p^h] than the allophonic contrasts [p]-[p^h] in a word-medial syllable initial position.

A similar pattern was also reported by Pegg and Werker (1997), who used an AX discrimination task to compare listeners' sensitivity to the voiced and the voiceless unaspirated alveolar stop pair, [d]-[t], relative to the [d]-[t^h] pair. Crucially, while adult English listeners have extensive experience with all three phones, [d]-[t^h] serve to distinguish words in word-initial position in English, whereas [d]-[t] do not. In line with Whalen et al. (1997), the phonemic pair was discriminated more accurately than the allophonic pair, despite the listeners' extensive experience with both phones in perception and production.

Peperkamp et al. (2003) used an AX discrimination task to investigate French listeners' perception of phonemic [m]-[n] and allophonic [ʋ]-[χ] contrasts. In French, [χ] is a predictable variant of the phoneme /ʋ/ which precedes a voiceless consonant. Like the other studies mentioned above, the authors found better discrimination for the phonemic [m]-[n] pair than the allophonic distinction between [ʋ] and [χ], when the latter were presented in a preconsonantal environment (i.e., [aʋ.CV]-[aχ.CV]). Interestingly, poorer discrimination was observed for the allophonic contrast regardless of whether the voicing of the consonant in the context syllable was phonotactically legal (matched the phone in question in voicing) or not, suggesting that allophonic variants are represented as a single phonological category.

In a recent study, Boomershine et al. (2008) used a similarity rating and a speeded AX discrimination task to investigate the impact of contrast and allophony on the perception of the phones [d], [ð], and [r] in intervocalic contexts by native English and Spanish listeners. The authors hypothesized that, if the phonological status of these segments in the listeners' native language determines the perceived similarity of the pair, we should expect relatively more discrimination difficulty

(longer RTs on a speeded AX discrimination task) and greater perceived similarity (higher similarity ratings on a similarity rating task) for the allophonic than for phonemic contrasts. These predictions were borne out. Spanish listeners produced higher similarity ratings and longer RTs than English speakers for the [d]-[ð] contrast, which are allophones of the same phoneme in Spanish. In contrast, English listeners had more difficulty discriminating [d]-[r], which are phonologically related in their native language. The pair was also rated by English listeners as being perceptually less distinct than the other two contrasts⁵. These findings are consistent with those reported earlier and provide additional evidence that listeners' perception is shaped by the phonology of their native language. In particular, the phonological status of pairs of phones in a listener's native language is an important factor in determining the discriminability and perceived similarity of a pair of phones (see also Johnson and Babel, 2010 who report data for Dutch listeners' perception of fricatives, Shea and Renaud, 2014 for Spanish listeners' perception of the palatal obstruent alternation, and Harnsberger, 2001 for Malayalam listeners' perception of allophonically-related dental and alveolar nasal consonants).

In addition to the behavioral studies reviewed above, research using neurophysiological techniques has also reported important differences in the processing of contrastive vs. non-contrastive sound pairs (Näätänen et al., 1997; Kazanina et al., 2006). Unlike behavioral measures, which may reflect late conscious processes, time-sensitive measures such as electroencephalography (EEG) and magnetoencephalography (MEG) measure neuronal activity in the brain directly and can be collected continuously without the necessity of an overt behavioral response on the part of the participant. They have thus proven useful for studying language processing and acquisition in a wide range of participant populations, including infants, and clinical populations. They also hold promise for studying language learners, since they may provide a measure of stimulus processing even in the absence of a behavioral change. For example, McLaughlin et al. (2004) demonstrated that ERPs to L2 words and pseudowords provide early evidence for word learning before changes in overt judgments were evident on lexical decision tasks. Therefore, it is possible that the learner's neural response will provide evidence of sound category learning that is not yet evident in her behavioral response.

A negative component of the event-related potential known as the mismatch negativity (MMN), and its magnetic counterpart, the mismatch field (MMF) response recorded using MEG, provide an early automatic, change detection response (Näätänen, 1992) which has proven useful for the study of auditory processing. The MMN is typically elicited in an oddball paradigm in which a stream of frequent repeated auditory stimulus (i.e., the standard in an experimental block) is interrupted by an oddball (i.e., an infrequent deviant acoustic event) which may differ in frequency, duration, intensity, phoneme category, etc. The MMN, which is obtained by

⁵It is worth noting that these results were observed despite the fact that [d] does not occur naturally in an intervocalic environment in either Spanish or English.

subtracting the event-related response to the standard event from the response to the deviant event, typically peaks at 150–250 ms from the onset of an infrequent detectable change and can be elicited in the absence of attention (i.e., in passive listening conditions). Moreover, by making use of a paradigm in which participants are presented with multiple non-orthogonally varying tokens from each category (as opposed to an acoustic standard), an MMN serves as a measure of category identification (Phillips et al., 2000).

A number of studies have demonstrated that aspects of a listener's native phonology modulate MMN amplitude (Näätänen et al., 1997; Phillips et al., 2000; Kazanina et al., 2006). In a seminal study, Näätänen et al. (1997) investigated the role of experience with language-specific vowel categories by studying the MMN responses of Finnish and Estonian listeners to the Estonian vowels /e, ö, õ, o/. Crucially, the Finnish language has the vowels /e, ö, o/, but lacks /õ/. Finnish and Estonian listeners were presented with the vowel /e/ as the frequent standard stimulus and /ö, õ, o/ as deviants in an oddball paradigm. The authors reported larger MMN responses for vowel contrasts involving native language vowel prototypes than contrasts involving nonnative vowel prototypes. That is, the Finnish participants showed an enhanced MMN response when the deviant vowel existed in Finnish, but the response was unexpectedly small (given the size of the acoustic difference in the F2 dimension) when it was elicited by a vowel that doesn't exist in the Finnish vowel inventory (i.e., /õ/), suggesting that the MMN response is influenced by experience with native language phoneme categories.

The MMN response has also been used as an index of nonnative vowel phoneme acquisition by second language listeners. Winkler et al. (1999) investigated whether novel vowel phoneme representations can be learned by recording the MMN responses of three groups of listeners, Finnish native speakers, proficient L1 Hungarian-L2 Finnish listeners, and naive L1 Hungarian listeners. The MMN responses of these groups were compared for two vowel contrasts, one that is phonemic in Finnish only (i.e., /e/-/æ/), and one that is phonemic in both languages (i.e., /e/-/y/). While an MMN was observed for all three groups for the /y/ deviants when presented in the context of the /e/ standard, the responses to /æ/ deviants differed as a function of experience. An MMN was observed for the /æ/ deviants for the native Finnish and the L1 Hungarian-L2 Finnish listeners, but not for the naive Hungarian listeners. This finding is taken to suggest that the proficient Hungarians had developed a new phonemic vowel representation for the Finnish vowel /æ/ as a result of their experience.

In a study which looked specifically at the pre-attentive processing of phonemes vs. allophones, Kazanina et al. (2006) investigated whether the MMN response is sensitive to the functional significance of native language sound categories. The authors examined the processing of the phones [t] and [d] in word initial position by Russian listeners, for whom the contrast is phonemic, and by Korean listeners, for whom the contrast is allophonic. That is, while both [t] and [d] naturally occur in word-initial position in Russian, only [t] is found word-initially in Korean. The voiced variant [d] occurs

in intervocalic position in Korean. Thus, [t] and [d] do not distinguish meaning in Korean. Russian participants showed both behavioral evidence of categorical perception (i.e., a classic step-like identification function for the /ta/-/da/ VOT continuum and better between-category than within-category discrimination) and neurophysiological evidence of change detection in auditory cortex. In contrast, Korean participants showed neither behavioral, nor neurophysiological evidence of perceptual sensitivity to the pair. These results suggest that adult native listeners' auditory cortex groups sounds based on phonemic categories, and that the functional significance of sounds factors into speech perception at a very early stage of processing. Moreover, the amplitude of the MMN response can be used as an early automatic index of perceptual categorization.

In a recent training study with L2 learners, Herd (2011) made ERP recordings both prior to and following perception training in order to investigate the effects of training on the L1 English-L2 Spanish listeners' automatic, pre-attentive processing of auditory stimuli containing the Spanish /d/-/r/ contrast. The author examined the processing of the phones [d] and [r] in an intervocalic context (i.e., [ede] and [ere]) by Spanish listeners, for whom the contrast is phonemic, and by L1 English learners of Spanish, for whom the target language contrast is allophonic in their L1. As expected, native Spanish listeners showed a significant MMN response, with deviant stimuli eliciting a more negative response than their standard counterparts. This pattern was observed both when [ede] standard was compared to [ede] deviant and when [ere] standard was compared to [ere] deviant. L1 English learners of Spanish also showed a significant MMN for both pairs at post-test. Unexpectedly, however, an MMN response was also present at pre-test for [ede] standard vs. [ede] deviant for the L1 English learner group, suggesting that an [ede] deviant is detected in a stream of [ere] standards even before perception training. These results are difficult to interpret, however, since the author does not report the performance of a monolingual English control group. As a result, it is unclear how much learning has occurred, either prior to the training, or as a result of the training. More work is needed to understand the role of L1 context-dependent allophones in second language speech perception and phonological development.

A related question in bilingual speech perception has been whether early stages of speech representation which are indexed by the MMN can be affected by the language being used. For instance, in a follow up to their earlier study, Winkler et al. (2003) investigated whether Hungarian-Finnish bilinguals would show different patterns of neural activity in response to the same stimulus pairs as a function of language context. The authors elicited MMN responses with two oddball sequences in which the Finnish word /pæti/ "was qualified" served as the frequent standard stimulus and /peti/ "bed" the infrequent deviant, first in a Hungarian language context, and later in a Finnish language context. The Hungarian-Finnish bilingual participants exhibited an MMN response to the /pæti/-/peti/ pairs in both the Hungarian and Finnish contexts, and the responses elicited in the two contexts did not differ from one

another. Based on these findings, the authors concluded that language context does not affect the automatic change detection response elicited by auditory deviance. Instead, the acquisition of a second language results in new phonemic categories that are used regardless of language context.

In contrast, a recent study by García-Sierra et al. (2012) demonstrated that language context can influence the pre-attentive detection of auditory deviance. The authors investigated Spanish-English bilinguals' MMN responses to two different pairings of three stimulus tokens from a synthetic VOT continuum in both a Spanish and an English language context. The language context was manipulated by having Spanish-English bilingual participants silently read magazines in either Spanish or English while ERPs were recorded. In the *phonemic in English* condition participants heard a stimulus token with +50 ms VOT as standard and +15 ms VOT as deviant. In the *phonemic in Spanish* condition participants heard a stimulus token with -20 ms VOT as standard and +15 ms VOT as deviant. As predicted, an MMN was elicited for the phonemic in English condition when the participants were in an English language context, but not a Spanish language context. Likewise, an MMN was observed for the phonemic in Spanish condition in the Spanish language context, but not the English language context. The authors take these findings to suggest that language context can indeed affect pre-attentive auditory change detection. While the present study did not set out to investigate the role of language context, the results of Winkler et al. (1999), Winkler et al. (2003) and García-Sierra et al. (2012) do suggest that sounds that are non-contrastive in a listener's L1 may be perceived differently as a result of experience. Moreover, bilingual listeners may demonstrate flexibility in their perceptual abilities as a result of the language context.

In sum, listeners' perception of speech sounds is strongly and systematically constrained by the native language phonology, with the discriminability of pairs of phones being influenced by phonological status in the native language. This pattern of relative insensitivity to phone pairs which are allophones of a single phoneme category in the listener's native language is observed both in behavioral and neural responses. While these patterns of perception may be optimal for listeners when listening to their native language, such learned, early, and automatic insensitivity to L1 allophones may present challenges for L2 learners who are faced with the task of establishing a novel contrast among familiar pairs of target language phones. These findings prompt the question of whether and to what extent these patterns of perception can be overcome with experience. In particular, do L1 context-dependent allophones continue to play a role in L2 perception?

In this study we further investigate the acquisition of novel target language contrasts among L1 context-dependent allophones by L2 learners. We take advantage of the cross-linguistic differences in the mappings between the phones [d], [ð], and [r] and their respective phoneme categories in English and Spanish. To this end, two experiments were conducted to investigate the representation and processing of three sound contrasts [d]-[ð], [d]-[r], and [ð]-[r] by three participant groups:

English native speakers, Spanish native speakers, and advanced L1 Spanish late-learners of English.

We used an AX discrimination task as a behavioral measure of participants' sensitivity to various tokens of three nonword pairings [idi]-[iði], [idi]-[iri], and [iði]-[iri]. Following Boomersshine et al. (2008) (among others), it was expected that the same phonetic contrast would be perceived more readily by listeners for whom the pair is phonemic in their native language than by listeners for whom the pair is allophonically related, and that this difference in sensitivity should be reflected in participants' d' scores. Thus, higher d' scores are expected for Spanish listeners than English listeners for the [idi]-[iri] contrast which is phonemic in Spanish, and allophonic in English, whereas English listeners were expected to outperform the Spanish listeners on the [idi]-[iði] pair which is phonemic in English and allophonic in Spanish. Finally, both native English and Spanish speakers were expected to demonstrate comparable sensitivity to the [iði]-[iri] control contrast which is phonemic in both languages. Of particular interest is the performance of the advanced L1 Spanish late-learners of English for the [idi]-[iði] contrast which is allophonic in the listeners' L1. If learners have overcome the learned insensitivity to the phonetic distinction between [idi]-[iði] and have established a novel contrast between /d/ and /ð/ in English, we expect no difference in their performance for this pair from the performance of the English speaker group. However, if learners have not yet established a novel target language contrast among L1 positional variants in perception, then we expect they may continue to have difficulty discriminating the pair.

Magnetoencephalographic (MEG) recordings were also used to measure the detailed time-course of brain activity in each of the three listener groups. By making a three-way comparison of pre-attentive processing to the three phones of interest by Spanish, English, and L2 listeners we can gain insight into the interlanguage phonological representations of the L2 learners. By using the presence of an MMN as an index of category identification, we will be able to show whether L2 learners represent the phones [d], [ð], and [r] as English speakers or Spanish speakers do. If early auditory brain responses are shaped by the functional significance of the sound categories in the listeners' native language (Kazanina et al., 2006), then we should observe a different pattern of results as a function of listener group. A significant MMN response is expected for both Spanish and English listeners for the control contrast (i.e., [iði]-[iri]). For the English group, a MMN response is also expected for the phonemic pair [idi]-[iði], but not for the [idi]-[iri] pair, which is allophonic in the language. In contrast, an MMN should be observed for Spanish listeners for the [idi]-[iri] pair, but not for the allophonically related pair [idi]-[iði]. With respect to the performance of the advanced late learners of English, we expect that if they have acquired the English /d/-/ð/ contrast, they will show evidence of perceptual sensitivity in their pre-attentive brain response. However, if they have not yet acquired the target language contrast, we expect them to perform like the native Spanish speaker group.

MATERIALS AND METHODS

Participants

Three groups of participants were recruited to participate in these experiments for monetary compensation; 15 English native speakers (Female = 5, Male = 10, mean age = 22.3 years, range = 19–28), 15 Spanish native speakers (Female = 8, Male = 7, mean age = 34.7 years, range = 23–45), and 15 advanced L1 Spanish late-learners of English (Female = 8, Male = 7, mean age = 30.1 years, range = 24–38). The learner group had a mean age of exposure of 10.1 yrs ($SD = 3.5$), had lived in the US for 6.2 yrs on average ($SD = 5$) and had 8.6 yrs of formal training in English ($SD = 4.7$). All participants tested strongly right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971) and reported no history of hearing or neurological disorder. All participants were recruited from the University of Maryland, College Park and the surrounding area. English speaking participants and the majority of the Spanish speaking learners of English were undergraduate and graduate students who studied or worked at the University of Maryland campus. The Spanish speakers with little/no experience with English were recruited from a neighboring community with a large Spanish speaking population. This group was largely comprised of immigrants from Central America who had recently arrived to the area and continue to use Spanish as their primary mode of communication. They report having had little exposure to English aside from what is heard on TV and the radio⁶.

The proficiency of each of the listener groups was assessed by self report. Participants were asked to rate their abilities in the areas of speaking, listening, reading, and writing on a scale of 1–10 (where 1 = poor and 10 = excellent) in both Spanish and English. The English speaker means were 10 ($SD = 0$) speaking, 9.9 ($SD = 0.3$) listening, 9.9 ($SD = 0.3$) reading, 9.9 ($SD = 0.3$) writing in English and 1.7 ($SD = 0.9$) speaking, 1.9 ($SD = 0.9$) listening, 2.1 ($SD = 1.3$) reading, and 1.6 ($SD = 1.1$) writing in Spanish. The Spanish speaker means were 2.7 ($SD = 2.1$) speaking, 3.5 ($SD = 2.4$) listening, 3.5 ($SD = 2.4$) reading, and 2.7 ($SD = 1.8$) writing for English and 10 ($SD = 0$) speaking, 10 ($SD = 0$) listening, 9.9 ($SD = 0.3$) reading, and 10 ($SD = 0$) writing in Spanish. The mean ratings for the Learner group in English were 8.0 ($SD = 1.3$) speaking, 8.5 ($SD = 1.1$) listening, 9.1 ($SD = 0.9$) reading, and 8.1 ($SD = 1.4$) writing. The means of the Learner group in Spanish were 9.9 ($SD = 0.4$) speaking, 10 ($SD = 0$) listening, 10 ($SD = 0$) reading, and 9.8 ($SD = 0.6$) writing.

⁶It is worth noting that, in addition to language experience, the participants in the Spanish speaking group likely differ from the listeners in the other two groups in a number of other respects, including SES, level of education, experience and level of comfort working with computers, etc. While it may have been possible to find a better matched group of Spanish speakers elsewhere, we were constrained by location of accessible MEG equipment. This is not an obvious concern for our MEG data (which requires no behavioral response), but could impact the quality of our behavioral data which required participants to respond by pressing buttons on a computer keyboard.

Stimuli

Materials for our experiments consisted of 10 natural tokens of each of the following VCV sequences: [idi], [iði], [iri] spoken by a single female speaker of American English with phonetic training. Multiple instances of each stimulus type were recorded using a head-mounted microphone in a soundproof room. The vowel [i] was chosen for the vowel context because Spanish [i] and English [i] have the greatest perceived similarity by listeners of both groups (Flege et al., 1994). The resulting stimulus set did not result in words in either Spanish or English. Because the phones [d] and [ð] and [d] and [r] are in complementary distribution in Spanish and English, respectively, it was not possible to find a context in which all three phones occur naturally. For this reason, it should be noted that the [idi] tokens may not sound particularly natural to either speaker group. All [idi] tokens were produced with care by a native English speaker with phonetic training so as to avoid flapping. Each was later inspected by two additional trained phoneticians to ensure that intervocalic [d] was not produced as [r]. To ensure that any observed differences in the MMN response could only be attributed to differences in the consonant (as opposed to the preceding vowel), the initial [i] from each token was removed and replaced with an identical [i] recorded in a neutral context (i.e., [isi]). The ten best stimulus tokens of each type were chosen on the basis of their perceived naturalness to native speakers of Spanish and English to ensure that each stimulus token was perceived as acceptable by native speakers of both languages. All experimental stimuli were normalized for intensity using Praat (Boersma and Weenink, 2009) and were presented to participants at a comfortable listening level (~70 dB).

One challenge for this kind of design is ensuring that the tokens used are relatively natural exemplars across both languages. We examined a number of acoustic parameters to determine to what extent this was true of the current stimuli. The initial [i] of each token had a duration of 160 ms, intensity of 77 dB, F0 of 190 Hz, F1 of 359 Hz, F2 of 2897 Hz, and F3 of 3372 Hz. The initial [i] was cross-spliced with the natural consonant and final [i] productions. The files were matched from positive going zero-crossing to positive going zero-crossing. The final [i] tokens had a mean duration of 177 ms ($SD = 20$), intensity of 75 dB ($SD = 1.8$), F0 of 172 Hz ($SD = 8$), F1 of 350 Hz ($SD = 14$), F2 of 2826 Hz ($SD = 72$), and F3 of 3278 Hz ($SD = 66$). These formant values for initial and final [i] tokens fall within the range of values for female speakers of American English reported by Hillenbrand et al. (1995) (F0 = 227 Hz ($SD = 24$), range = 155–275 Hz; F1 = 437 Hz ($SD = 41$), range = 331–531 Hz; F2 = 2761 Hz ($SD = 147$), range = 2359–3049 Hz; F3 = 3372 Hz ($SD = 237$), range = 2958–3831 Hz)). The vowel duration reported by Hillenbrand et al. (1995) for [i] is longer (306 ms ($SD = 46$), range = 222–433 ms) than the duration of the [i] tokens reported here. However, this is expected given that their recordings were elicited in a h_d context. The formant values also match fairly closely the values reported by Quilis and Esgueva (1983) for Spanish [i] (F1 = 241 Hz ($SD = 32$), range = 202–324 Hz; F2 = 2839 Hz ($SD = 237$), range = 2349–3321 Hz; F3 = 3358 Hz ($SD = 249$), range = 2632–3726 Hz), with the exception that the Spanish [i] has a lower F1

than English [i]. The mean duration of the consonant segments of interest measured from the F2 offset of V1 to the onset of F1 of V2 were 76 ms ($SD = 10$, range = 63–96 ms) for [d], 78 ms ($SD = 13$, range = 59–99 ms) for [ð], and 41 ms ($SD = 5$, range = 34–51 ms) for [r]. These values are comparable to those reported for English by Lavoie (2001) and Stathopoulos and Weismer (1983) for initial and medial non-prestressed /d/ (i.e., 70, 80 and 37, 41 ms, respectively). Our speaker's [ð] productions were on average longer than those reported Lavoie (2001) for initial and medial non-prestressed environments (57 and 48 ms). For Mexican Spanish, Lavoie (2001) reports durations of 51, 24, and 55 ms for medial non-prestressed /d/ and /r/ and initial non-prestressed /d/, respectively. The mean duration of the stimulus tokens measured from word onset to word offset from the Praat waveform was 426 ms ($SD = 29$, range = 384–480 ms) for [idi], 416 ms ($SD = 9$, range = 402–429 ms) for [iði], and 363 ms ($SD = 20$, range = 319–397 ms) for [iri]. Following Carrasco et al. (2012), we also computed a ratio of the minimum intensity of the consonant/maximum intensity of the following vowel as a measure of the relative intensity/degree of constriction of the consonant productions. A ratio that is close to one indicates a more open vowel-like production of the consonant, and a ratio that is closer to zero indicates a more stop-like realization. The ratio for the [d] was 0.70 ($SD = 0.02$), for [ð] was 0.77 ($SD = 0.04$) and for [r] was 0.81 ($SD = 0.03$). While the ratios shouldn't be compared directly to those reported in Carrasco et al. (2012), since vowel contexts are known to affect these measures (Simonet et al., 2012)⁷ and the vowel contexts differ from those used in their study, what is worth noting is that the most vowel like production is the [r] and the least vowel-like production is the [d]. The [ð] lies in between those two.

Post-study Identification Task

To ensure that participants in the study also identified the stimuli as instances of the intended category, each performed a brief identification task following the MEG recording and the AX

⁷Simonet et al. (2012) report several continuous measurements of relative intensity. The authors argue that even among intervocalic tokens of /d/, the height of the preceding vowel conditions the degree of constriction of the consonant. Importantly, they report that in Iberian Spanish /d/ is more constricted after a high vowel than after a mid or low vowel.

discrimination task. Participants were presented with 40 stimuli (each of the 30 experimental items and 10 filler items) and were instructed to use the keys 1, 2, and 3 to identify the stimulus they heard. Naturally, the labels for the identification task had to vary across language, such that the English speakers were asked to label stimuli as an instance of a nonword “eithe,” “eady,” or “other” and the Spanish speakers as the nonwords “idi,” “iri,” or “other.” In order to implement the task in a similar way across groups we had to decide which labeling to request from the Learners. Given that our primary interest in the identification task was to learn if our stimulus tokens would be categorized as instances of the expected stimulus type in the listeners' L1, we opted to use L1 labeling options for all three listener groups.

Figure 2 shows the frequency of each response by stimulus type for each of the three language groups. All participants chose “other” predominantly for the filler items. English listeners chose “eithe” predominantly for the [iði] tokens, and “eady” for both the [idi] and [iri] tokens. Both Spanish listeners and the Learner group primarily chose “idi” as the label for [idi] and [iði] tokens, whereas [iri] tokens were predominantly identified as “iri” by Spanish listeners. Thus, the stimulus tokens used in the study can be heard as instances of the expected stimulus type in the listeners' L1 (at least on a conscious-labeling task).

PROCEDURES

AX Discrimination Task

During the AX discrimination task participants wore headphones and were seated in a quiet room in front of a computer. The presentation of experimental stimuli was controlled by DMDX (Forster and Forster, 2003). In the AX discrimination task participants were presented two of the experimental stimuli which were either different tokens of the same nonword (i.e., [idi]-[idi], [iði]-[iði], [iri]-[iri]) or one of the six possible ordered pairings of different nonwords (i.e., [idi]-[iri], [idi]-[iði], [iði]-[iri], [iði]-[idi], [iri]-[iði], [iri]-[idi]). Participants responded to 32 same (16 AA, 16 BB) and 32 different (16 AB, 16 BA) trials per contrast, for a total of 192 test trials. Each stimulus was presented with an interstimulus interval (ISI) of 500 ms. Participants were instructed to press the “F” key on the keyboard with their left

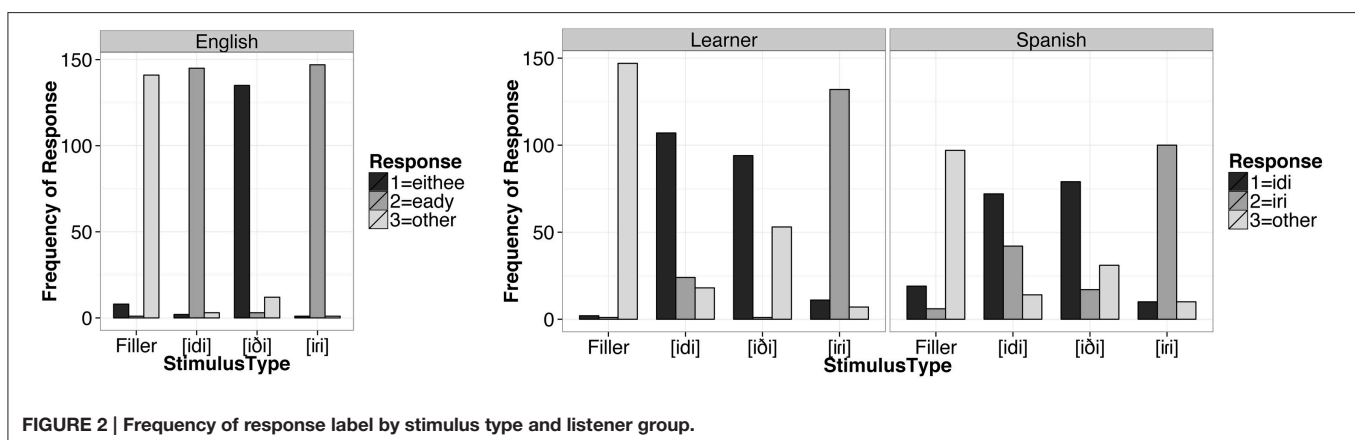


FIGURE 2 | Frequency of response label by stimulus type and listener group.

index finger if the two stimuli were two pronunciations of the same “word” and to press the “J” key with their right index finger if the paired stimuli corresponded to two different “words.” Participants were asked to respond as quickly and accurately as possible and had a maximum of 4 s to respond on each trial. Written instructions were provided in the native language of each listener group, as well as orally by the experimenter. Six practice trials without feedback preceded the test trials to ensure that participants understood and were comfortable performing the experimental task. These practice trials were repeated a second time in the case that participants still appeared uncertain about the task or uncomfortable with providing their response on the computer keyboard. The AX discrimination task lasted approximately 15 min and was divided into four blocks of 48 items with three self-timed breaks between each block.

MEG Recordings

Magnetic fields were recorded in DC (no high-pass filter) using a whole-head MEG device with 157 axial gradiometers (Kanazawa Institute of Technology, Kanazawa, Japan) at a sampling rate of 1 kHz. An online low pass filter of 200 and a 60 Hz notch filter were applied during data acquisition. All stimuli were presented binaurally via Etymotic ER3A insert earphones at a comfortable listening level (~70 dB). MEG recording sessions included 4 runs: 1 screening run and 3 experimental blocks which are described in greater detail below. Participants passively viewed a silent movie during the experimental runs to avoid fatigue. Each MEG recording session lasted approximately 90 min in total.

In the screening run, participants were presented approximately 100 repetitions of a 1 kHz sinusoidal tone. Each tone was separated by a randomly chosen ISI of 1000, 1400, or 1800 ms. Data from the screening run were averaged and examined to verify a canonical M100 response. The M100 is an evoked response which is produced whenever an auditory stimulus has a clear onset and is observed regardless of attentional state (Näätänen and Picton, 1987). Data from 45 participants run across the three participant groups showed a reliable bilateral M100 response with a source/sink reversal between anterior and posterior channels in the left and right hemisphere. Three additional participants were recruited and run on the screening task, but were excluded because they did not show a strong bilateral M100 response elicited by a 1-kHz pure tone at pretest. The M100 response elicited to non-speech tone stimuli were additionally used to select the auditory channels of interest for each of our participants for the MMN amplitude analysis.

In the experimental blocks, stimuli were presented using a modified version of the optimal passive oddball paradigm (Näätänen et al., 2004). In each of the three experimental blocks one of the three stimulus types (i.e., [idi], [iði], or [iri]) was presented frequently (i.e., the standard) and was followed by infrequent stimuli of the other two types. For example, in **Figure 3**, the first block shows [idi] as the frequent standard and [iði] and [iri] as the less frequent intervening deviant stimulus types. Following Phillips et al. (2000), there was no acoustic standard. Instead, participants were presented multiple non-orthogonally varying tokens from each category. This was done

Experimental Blocks:

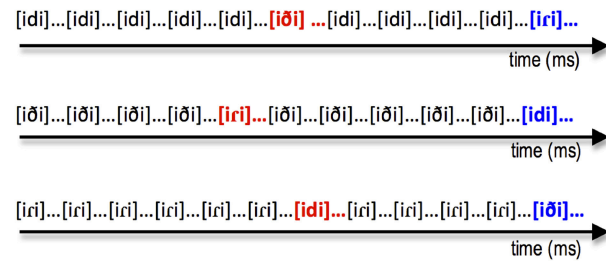


FIGURE 3 | Illustration of the structure of each of the three experimental blocks in our modified passive oddball paradigm. Stimuli shown in black correspond to the stimulus type that served as the standard for that experimental block. The two types of deviants for a particular block are shown in red and blue.

to avoid a purely acoustic interpretation of the elicited responses. Thus, the presence of an MMN serves as a measure of grouping of different acoustic tokens into phoneme or allophone categories. Each block consisted of 882 standards and 168 deviants (84 of each deviant type). A deviant was presented after a minimum of 4 and a maximum of 6 standards with the probability of deviant (either deviant type A or B) = 0.167. Each stimulus token was separated by an ISI that varied randomly between 600 and 1000 ms. Each of the three experimental blocks lasted approximately 20 min. Participants were given a short break after each 10 min of recording. Block order was counterbalanced across participants. **Figure 3** shows the structure of each of the three blocks.

The experimental procedures were completed in the following order for all participants: [1] participants were provided an overview of the procedures and provided their informed consent, [2] participants completed a language background and handedness questionnaire to ensure they met the study requirements, [3] MEG recordings were made, and [4] AX discrimination and identification data were collected.

DATA ANALYSIS

AX Discrimination Data

Data from four Spanish participants (S003, S004, S011, S014) whose performance was at or below chance (i.e., 50% accuracy) on the control contrast (i.e., [iði]-[iri]) were excluded from subsequent AX discrimination analyses. For the remaining participants, d' scores were computed for each individual and each different pair according to the Same-Different Independent Observations Model (Macmillan and Creelman, 2005) using the `dprime.SD()` function from the *psyphy* package in R (Koblauch, 2007). The result is a measure of sensitivity which factors out participants' response bias. The “hit rate” was computed as the proportion of “different” responses when the words in the pair were different. The “false alarm rate” was the proportion of “different” responses when the words in the pair were the same. To correct for extreme proportions (i.e., hit rates and false alarm rates of 0 or 1), we applied Laplace smoothing (Jurafsky and Martin, 2009). In probability theory, Laplace's Rule of Succession

is used to estimate underlying probabilities when there are few observations, or for events that have not been observed to occur at all in some finite sample of data. The rule states that if we repeat an experiment that we know can result in a success or failure (in our case hit or false alarm), n times independently, and observe s successes, then the probability of success on the next repetition of the experiment is $(s + 1)/(n + 2)$. Thus, our best estimate of a participant's hit rate when 32 hits and 0 misses are observed across 32 different trials is $(32 + 1)/(32 + 2)$ (or 0.97). For a participant with a false alarm rate of 0, our best estimate of the false alarm rate is $(0 + 1)/(32 + 2)$ (or 0.03). As a result, the largest d' score that may be observed given our experimental materials with 32 different trials was 4.34. The d' values obtained for each test pair per subject ranged from 4.34 to 0. Two participants achieved the maximum d' score for one of the conditions (E015 for [iði]-[iri] and L009 for [idi]-[iði]). **Figure 4** shows the mean d' score by language group and contrast. These d' scores were subsequently analyzed using linear mixed effects modeling.

MEG Data Pre-processing

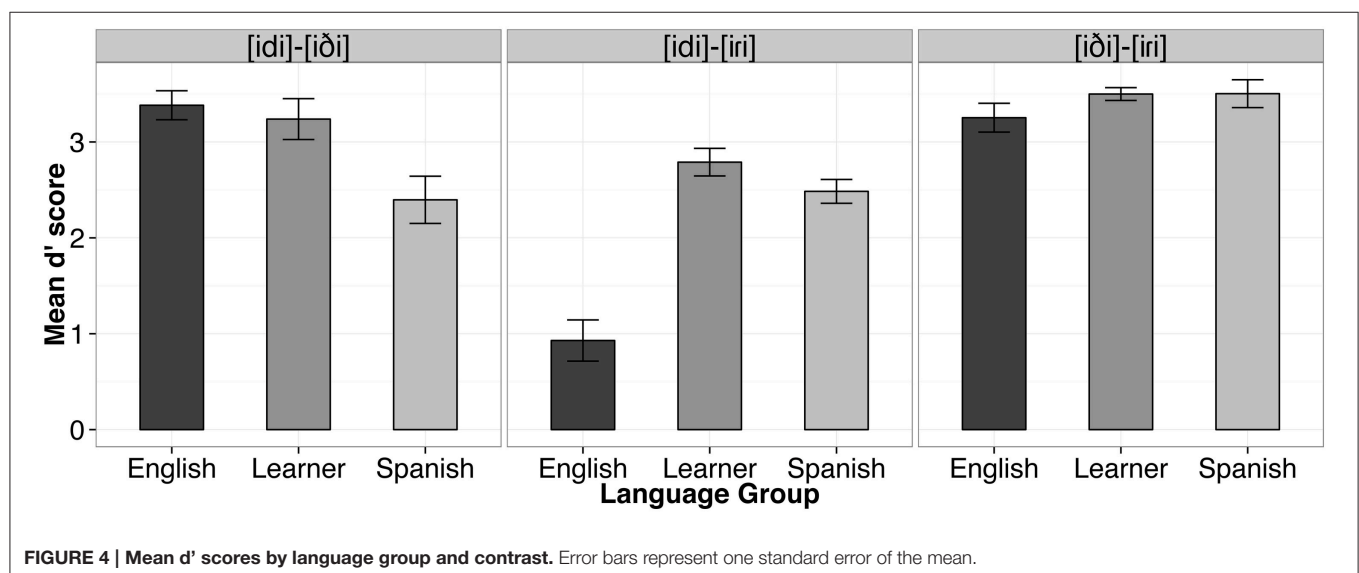
MEG data were imported into Matlab and de-noised using a multi-shift PCA noise reduction algorithm (de Cheveigné and Simon, 2007, 2008). Epochs included 100 ms pre-stimulus onset to 800 ms post-stimulus onset. Artifact rejection was conducted manually in MEG160 to exclude trials containing muscle and eye-related artifacts. All epochs were then averaged, baseline corrected over a 100 ms pre-stimulus interval, and filtered using a 0.03 to 30-Hz band-pass filter.

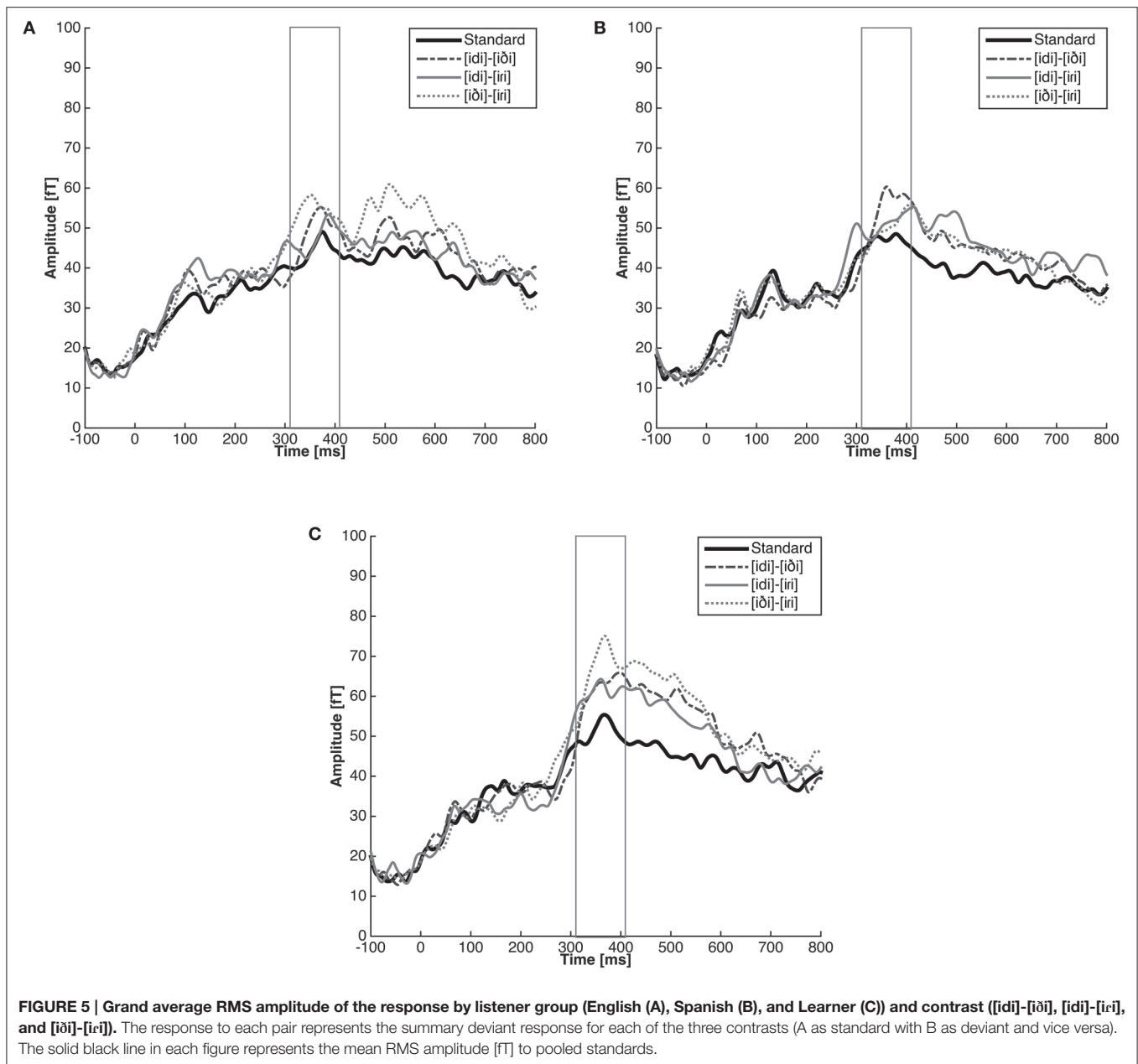
For each participant, the 10 strongest left hemisphere channels (5 from left anterior, 5 from left posterior) were identified and selected visually in MEG160 from the peak of the average M100 response to 1 kHz tones elicited during the auditory localizer pre-screening test. Because the MMNm to phoneme prototypes has been found to be stronger in the left hemisphere than in the right (Näätänen et al., 1997), we calculated the root mean square

(RMS) amplitude of the MEG temporal waveforms over the left hemisphere channels selected on the basis of the pre-screening test. Trials were averaged separately for each participant and for each condition (i.e., three standard and six deviant types).

We created a single summary deviant response for each of the three contrasts by averaging together the two relevant deviant responses. For example, for the [iri]-[iði] control contrast, we averaged together the response to [iði] deviants in an [iri] block and the response to [iri] deviants in an [iði] block. The averaged responses elicited by standards were also pooled, resulting in a single summary standard response. The grand average waveform from -100 ms pre-stimulus to 800 ms post-stimulus was then computed for language group by averaging across participants ($n = 15$ per group) for each condition (i.e., [idi]-[iði], [idi]-[iri], [iði]-[iri], and Standard). These are shown in **Figure 5**. Although in this analysis we collapse across data from both directions of a given contrast (A as standard with B as deviant and vice versa), it is worth noting that in certain cases such as phonological underspecification, directionality impacts the size of the MMN response (Eulitz and Lahiri, 2004). In the current case, we had no a priori reason to expect a systematic impact of directionality and therefore we collapsed across directions to ensure sufficient power. However, for the interested reader we include a supplementary analysis of the MMN data separated by direction in the Supplementary Materials.

The mean RMS power over a single 100 ms time window from 310 to 410 ms for each of the participants for each of the experimental conditions was computed. This time window was chosen because the vowel offset and consonant onset occurred at 160 ms and the MMN is expected to occur about 150–250 ms following the onset of a detectable change. Our statistical comparisons used linear mixed effects modeling to examine whether the difference in the mean RMS of the response to deviants and the response to standards reached significance over the MMN time window (310–410 ms).





RESULTS

d' Scores

Statistical analyses of d' scores were performed with linear mixed effects modeling using R package *lme4* (Bates et al., 2015) with factors Language Group (English, Learner, Spanish), Contrast ([idi]-[iri], [idi]-[iði], [iði]-[iri]), and the Language Group \times Contrast interaction as fixed effects and subject as a random effect in order to account for inter-subject variability. P-values were computed using the Satterthwaite's approximation for denominator degrees of freedom with the *lmerTest* package (Kuznetsova et al., 2014). We observed a main effect of Language Group [$F_{(2, 37)} = 10.07, p < 0.001$], and of Contrast [$F_{(2, 74)} =$

$54.40, p < 0.001$], as well as a Language Group by Contrast interaction [$F_{(4, 74)} = 19.20, p < 0.001$].

We conducted nine planned tests of our experimental hypotheses regarding listeners' sensitivity to allophonic vs. phonemic contrasts using simultaneous tests for general linear hypotheses with the *multcomp* package in R (Hothorn et al., 2008). P-values were adjusted using the single-step method. First, it was hypothesized that our three listener groups should not differ in performance on the control contrast (i.e., [iði]-[iri]), as the contrast is phonemic in both Spanish and English. This prediction was borne out. English listeners did not differ from Spanish listeners for this contrast ($\beta = 0.25, SE = 0.25, z = 1.02, p = 0.91$), nor did the d' scores of the English group and the

Learner group ($\beta = 0.25$, $SE = 0.23$, $z = 1.07$, $p = 0.89$) or the Spanish group and the Learner group differ for this contrast ($\beta = -0.003$, $SE = 0.25$, $z = -0.02$, $p = 1.00$).

For the [d]-[r] contrast, which is phonemic in Spanish, but allophonic in English, it was expected that the L1 Spanish listeners would outperform the English listeners. This prediction was also borne out. The English listeners performed significantly worse than both the Spanish listeners ($\beta = 1.56$, $SE = 0.25$, $z = 6.31$, $p < 0.001$) and the Learner group ($\beta = 1.86$, $SE = 0.23$, $z = 8.06$, $p < 0.001$), supporting the hypothesis that phonological status influences perception on our AX discrimination task. No difference was observed between the Spanish and Learner group for this contrast ($\beta = 0.30$, $SE = 0.25$, $z = 1.22$, $p = 0.82$).

For us the most important question is what level of discrimination performance Spanish late-learners of English would show on a contrast that is phonemic in English but allophonic in Spanish (i.e., [d]-[ð]). First, as expected, the Spanish group performed significantly poorer on this contrast than the English listeners ($\beta = -0.99$, $SE = 0.25$, $z = -4.00$, $p < 0.001$), again providing support for differential processing of the contrast as a function of phonemic status in the language. Interestingly, with respect to our primary research question, a significant difference was observed for the L1-allophonic contrast [d]-[ð] for the Spanish and Learner listener groups ($\beta = 0.84$, $SE = 0.25$, $z = 3.37$, $p < 0.01$), with larger d' scores observed for the Learners than Spanish listeners. Moreover, no significant difference in d' was observed between the English listener group and the Learner group ($\beta = -0.14$, $SE = 0.23$, $z = -0.63$, $p = 0.99$), suggesting that the participants in the Learner group may have acquired a target language contrast among the phones [d]-[ð] which function as context-dependent allophones in their L1.

Mean RMS Amplitude of MMN

We again used linear mixed effects modeling in R to conduct the statistical analyses of mean RMS amplitude over the 310–410 ms time window. Our first linear mixed effects analysis was designed to confirm that there were no reliable differences between the responses to the different standards. This is important to establish because we would like to collapse across the response to standards in our subsequent critical planned comparisons of the MMN response by contrast. Analyses of mean RMS amplitude for the response elicited by the standards consisted of fixed effects Language Group (English, Learner, Spanish) and Standard Type ([idi] standard, [iði] standard, [iri] standard), as well as Language Group \times Standard Type interaction and subject as random effect. These statistical analyses revealed no significant results, suggesting that the mean power elicited by standard stimuli did not differ by Language Group [$F_{(2, 42)} = 0.43$, $p = 0.66$] or Standard Type [$F_{(2, 84)} = 2.12$, $p = 0.13$], nor did these factors interact [$F_{(4, 84)} = 0.13$, $p = 0.96$]. We take this to suggest that listeners are able to form a coherent representation for the standard stimuli and that we are justified in comparing responses elicited by deviants against pooled standards.

Figure 6 shows the mean RMS amplitude of the MMN for each of the three contrasts for each listener group. Analyses of the MMN amplitude consisted of fixed effects Language Group

(English, Learner, Spanish) and Stimulus Type (Allophonic, Phonemic, Control, Standard), as well as Language Group \times Stimulus Type interaction and subject as random effect. There was no main effect of Language Group [$F_{(2, 42.14)} = 1.01$, $p = 0.37$]. However, the main effect of Stimulus Type reached significance [$F_{(3, 351)} = 7.21$, $p < 0.001$]. No interaction between Language Group and Stimulus Type was observed [$F_{(6, 351)} = 1.32$, $p = 0.25$].

In our statistical analyses of the listeners' responses to deviants, we conducted three planned comparisons separately for each listener group using simultaneous tests for general linear hypotheses with the *multcomp* package in R (Hothorn et al., 2008). P -values were adjusted using the single-step method. We compared each groups' response to the pooled standard (i.e., responses to the stimuli [idi], [iði], and [iri] when they are presented as standards in a block) to the groups' responses to the summary deviant response for each of the three contrasts.

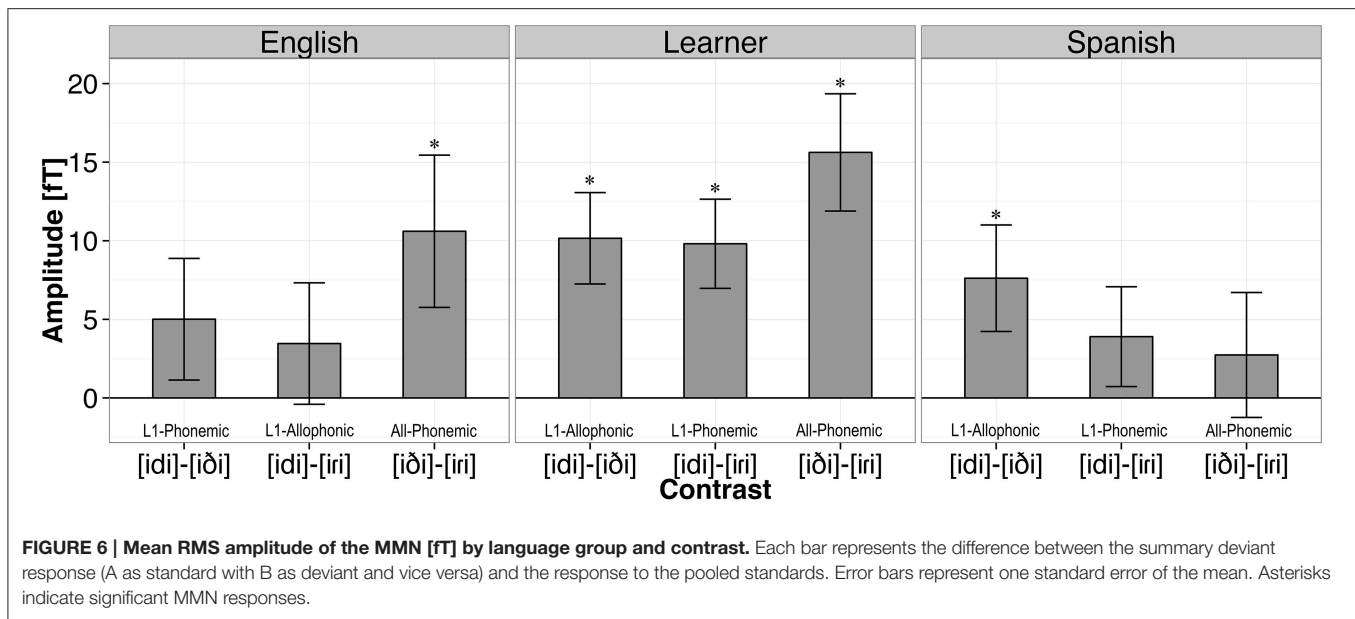
As expected for the English listeners, the response to the control contrast [iði]-[iri] was larger than the response to the standard stimuli ($\beta = 21.21$, $SE = 8.18$, $z = 2.59$, $p < 0.05$). Again as expected, we found no difference between the magnitude of the response elicited by the standards and the allophonic pair [idi]-[iri] ($\beta = 6.92$, $SE = 8.18$, $z = 0.85$, $p = 0.75$). Unexpectedly, we found no difference between the response to the standard and the response to the English phonemic contrast [idi]-[iði] ($\beta = 10.02$, $SE = 8.18$, $z = 1.22$, $p = 0.49$).

Unfortunately, the MMN responses for the Spanish listeners followed none of our predictions. We found a marginal difference between the response to the standard stimuli and the [idi]-[iði] pair which are phonologically related in the language ($\beta = 15.23$, $SE = 6.51$, $z = 2.34$, $p = 0.05$). We also found no difference between the standards and either the [iði]-[iri] control pair ($\beta = 5.47$, $SE = 6.51$, $z = 0.84$, $p = 0.76$) or the phonemic [idi]-[iri] pair ($\beta = 7.80$, $SE = 6.51$, $z = 1.20$, $p = 0.51$).

For the critical learner group, the MMN results followed the pattern predicted according to the hypothesis that learners successfully implemented the phonological knowledge of their second language at an early, pre-attentive stage of processing. A significant difference was observed between the standards and L1 allophonic contrast [idi]-[iði] ($\beta = 20.31$, $SE = 8.19$, $z = 2.48$, $p < 0.05$), the phonemic contrast [idi]-[iri] ($\beta = 19.62$, $SE = 8.19$, $z = 2.40$, $p = 0.05$), and the control contrast [iði]-[iri] ($\beta = 31.24$, $SE = 8.19$, $z = 3.82$, $p < 0.001$). These results suggest that the learners' ability to distinguish the contrasts that were observed in the behavioral data is also apparent at the stage of early pre-attentive processing, regardless of the pairs phonological status in the L1.

DISCUSSION

In this study we explored the impact of phonological knowledge on perceptual categorization, particularly in cases in which the phonemic status in a late-learned second language directly conflicts with the native language. Our Spanish and English listeners demonstrated greater sensitivity for nonword pairs



distinguished by phonemic than by allophonic contrasts on an AX discrimination task, mirroring previous findings. Interestingly, Spanish late-learners demonstrated sensitivity (large d' scores and MMN responses) to all three contrasts, suggesting that these L2 learners may have established a novel [d]-[ð] contrast despite the phonological relatedness of these sounds in the L1. We discuss each of these findings in turn.

Phoneme-Based Equivalence Classes in the L1

Our behavioral findings from the native speaker groups provide support for the hypothesis that listeners form equivalence classes on the basis of phoneme categories. In particular, we observed better discrimination of the [idi]-[iri] contrast by Spanish listeners for whom the pair are phonemic than by English listeners for whom the pair is allophonic in their L1. Similarly, English listeners outperformed Spanish listeners in the discrimination of the [idi]-[iði] pair which is phonemic in English, but allophonic in Spanish. Finally, both Spanish and English listener groups performed comparably well on the [iði]-[iri] control contrast which is a phonemic distinction in both languages. These results replicate previous behavioral findings from Boomershteyn et al. (2008), and provide additional evidence that phonological relatedness among sounds reduces their perceptual similarity in native listeners.

The MEG data also provides partial support for the hypothesis that listeners establish equivalence classes on the basis of phonemes. Given this hypothesis, we expected to observe an MMN when the stimulus presented as the deviant is in contrast in the listener's native language with the stimulus serving as the standard in an experimental block, but not when the standard and deviant are phonologically related as allophones of the same phoneme in the listeners' L1. As expected, a significant MMN

was observed for the [iði]-[iri] control contrast, but not for the allophonic [idi]-[iri] contrast for English listeners. Contrary to our expectations, however, no MMN was observed for the phonemic [idi]-[iði] pair. In contrast with the data from the English listeners, the results for the Spanish listeners did not provide support for our hypothesis. A significant MMN was observed for the [idi]-[iði] contrast, which is allophonic in Spanish, while no MMN was observed for either the [idi]-[iri] or the [iði]-[iri] pair which are phonemic in Spanish.

It is not clear how to explain the unexpected MMN patterns observed in the two native listener groups. First, any explanation based on poor stimulus quality seems inconsistent with the behavioral data, which showed the predicted pattern of discrimination across groups for all contrasts (although it is of course logically possible that the behavioral responses were based on a late-stage process that the early MMN does not reflect). Second, it is not clear how any simple explanation based on the acoustic properties of the stimuli could explain the cross-linguistic differences in responses. However, we note that the only surprising datapoint in the English listener data was the absence of a significant MMN in the phonemic [idi]-[iði] contrast, but that the response was trending in the right direction. Therefore, we might speculatively attribute this result to a Type II error.

One factor that may have reduced our power to detect MMN differences in the current paradigm is that the position of the deviant within the standard stream was somewhat more predictable than in many MMN studies. In our experiment, a deviant always occurred after either 4, 5, or 6 intervening standards. Previous work has demonstrated that when the position of a deviant within the standard stream is completely predictable, the MMN is almost completely neutralized (see Sussman et al., 2014 for review) and therefore the partial predictability may have reduced the overall strength of MMN effects. Although in the current case this increase in predictability was partly driven by our desire to examine three different

contrasts (forcing a smaller standard to deviant ratio with fewer trials between deviants), in future work it would be useful to investigate the same contrast with greater unpredictability.

In addition, the slightly non-canonical status of the speech stimuli as neither perfectly English-like nor perfectly Spanish-like may have caused some of the unexpected MMN patterns observed in the Spanish and English groups. In an active task like AX discrimination, increased attention might mitigate the impact of slightly non-canonical tokens on categorization, but in a passive listening mode, as in the MMN paradigm, participants might not have automatically perceived and grouped the tokens according to their native speech categories. On the other hand, the bilingual participants might be more permissive of irregularities even in passive listening, based on their exposure to different distributions of sounds across the two languages. Strange and Shafer (2008), in their Automatic Selective Perception (ASP) model, have suggested that the perception of nonnative contrasts is dependent on task demands that determine the degree of attentional focus that is placed on the phonetic details of the stimuli. In support of the model, Hisagi et al. (2010) demonstrated that selective attention enhanced the magnitude of the MMN responses of the American English listeners to the nonnative Japanese vowel length contrast. Attention may likewise be required for listeners to categorize familiar phonological contrasts when the contrasts are specified by slightly different acoustic-phonetic parameters. Future work could examine the potential role of attention by manipulating attention directly and by incorporating active tasks which allow the researcher to monitor the participants' focus of attention. We believe that an additional related factor in the unexpected MMN pattern for Spanish and English speaker groups may be the fact that the [idi] token does not match prior language experience for either Spanish or English speakers, as [d] does not occur intervocally in either language. Addressing either of these factors in future work will be challenging however, because it is not possible to create tokens for the full set of contrasts that are fully and equally natural tokens of Spanish and English, and as allophonic variation is context dependent, contrasting allophones in the MMN design necessarily requires one of the allophones to be presented in an unnatural context.

Acquiring New Mappings among Familiar Phones

Our primary research question asked whether advanced L1 Spanish late-learners of English overcome learned insensitivities to L1 context-dependent allophones and acquire a new target-language contrast among familiar phones [d] and [ð]. The behavioral and neural data from L2 learners which we report here converge to suggest that the answer to this question is affirmative. On both tasks we observed no difference between learners' ability to discriminate between phone pairs which are L1 allophones and L1 phonemes, suggesting that they do not classify the two phones as allophones of the same underlying phoneme category. That is, with experience, the advanced L2 learners in our study have acquired adequate knowledge of the L2 phonological system to distinguish the English /d/-/ð/ contrast in perception. Moreover,

this learned sensitivity is observable both behaviorally, and in listeners' early, pre-attentive brain responses. We note that the neural data must be interpreted somewhat more cautiously than the behavioral data. Although the MMN pattern observed in the late-learner group was exactly what was predicted if they had successfully acquired the L2 phonological system, the two native listener groups did not show the MMN patterns predicted based on their L1 phonology, as described above. Therefore, further replication will be needed to confirm the interpretation of the MMN pattern in the late-learner group.

Given that our behavioral and MEG data from our Learner group was elicited in an English language context (all testing was conducted in an English speaking environment and all interactions and instructions were given in English), we might have expected the Learners' neural and behavioral responses to look maximally English-like (i.e., discriminating [d]-[ð] and [ð]-[r], but not [d]-[r]). However, this was not what was observed (contra García-Sierra et al., 2012 and in line with Winkler et al., 2003). We note, however, that we did not actively attempt to manipulate language context in our study. It is possible that Learners' performance on [d]-[r] would have been different had we done so. It is also possible that other factors contribute to the observed effects, such as the language dominance of the participants or the proportion of L1/L2 use. These interesting possibilities should be taken up in future research.

A question that arises naturally from our learner data is: how do L2 learners acquire the ability to perceive novel target language contrasts among familiar phones? In particular, what is the role of the input in shaping the learners' hypotheses about the phonological system they are acquiring, and how do learners' expectations about the characteristics of the target language influence the learning process. With respect to mechanisms, three possibilities have been discussed in the infant literature (Seidl and Cristia, 2012, provide a more detailed review). First, it has been proposed that some information about phonological status may be available in the acoustic signal. That is, allophones may be more acoustically similar than phonemes. Some support for the plausibility of a phonetic mechanism comes from acoustic analyses of nasal and oral vowel allophones and phonemes in corpora of infant directed speech (Seidl et al., 2014). However, more research is needed to demonstrate the extent to which reliable information of this sort is available in the input to learners and to investigate whether infants and adults actually can and do use this information when learning about phonological status.

Another possibility is that listeners' make use of distributional information (Maye et al., 2002), such as the phonological context in which phones occur to learn phonological status. It is well established that both infants and adults can track the distribution of phones in acoustic space (Maye and Gerken, 2000; Hayes-Harb, 2007) and other phonological units, such as syllables (Saffran et al., 1996). Distribution-based learning mechanisms are also assumed to play a role in infants learning of allophones (Peperkamp et al., 2003, 2006; White et al., 2008), and have been invoked in the acquisition of L2 allophonic alternations (Shea and Curtin, 2010, 2011; Shea, 2014). In such cases, L2 learners are thought to acquire knowledge about the phonological patterning of L2 allophonic variants by tracking the distribution of target

language phones with respect to their conditioning contexts. In the case of allophonic splits, it would seem that distributional learning is also required. Learners must learn that phones which are contextually licensed in their L1 are not restricted to the same phonological environments in their L2. For example, [ð] is permitted word-initially and word-finally in English, in addition to in word-medial post-vocalic non-prestress environments as in Spanish. More work is needed to investigate how the input is processed by adult L2 learners and to demonstrate that novel contrasts among L1 context-dependent allophones can indeed be learned by tracking phones and their respective phonological contexts.

Lexical mechanisms of various sorts have also been proposed, such as knowledge of word meanings and knowledge of words' phonological forms. For example, the availability of minimal pairs has been shown to enhance the perception of nonnative phonetic contrasts in both infants and adults (Hayes-Harb, 2007; Yeung and Werker, 2009). More recently, acquisition researchers have investigated the role of word contexts in phonetic category learning, demonstrating that infants and adults are sensitive to and can use distinct word forms (in the absence of visual referents or knowledge of word meanings) to constrain their interpretation of phonetic variability (Feldman et al., 2013).

Finally, in addition to the implicit learning mechanisms mentioned above, it has also been suggested that adults might avail themselves of explicit learning mechanisms and that these may serve to initiate the acquisition process (Shea, 2014). The effectiveness of various types of explicit input to L2 learners should be taken up in future research.

In sum, the behavioral and neural results presented here suggest that phonological relatedness influences perceived similarity, as evidenced by the results of the native speaker groups, but may not cause persistent difficulty for advanced L2 learners in perception. Instead, L2 learners overcome learned insensitivities to L1 allophones in perception as they gain experience with the target language. These findings provide a starting point to investigate when and how this learning takes

place, as well as determine the respective contributions of the proposed mechanisms to the acquisition of novel target language contrasts among L1 context-dependent allophones in the L2.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: SB, AN, EL, NF, WI. Performed the data collection: SB, AN. Analyzed the data: SB, EL, WI. Wrote the manuscript: SB. Revised the manuscript critically for important intellectual content: SB, AN, EL, NF, WI. Provided final approval of the version to be published: SB, AN, EL, NF, WI. Agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: SB, AN, EL, NF, WI.

FUNDING

This research was partially supported by a University of Utah University Research Committee Faculty Research and Creative Grant awarded to SB, as well as a University of Maryland, Department of Linguistics, Baggett Scholarship awarded to AN.

ACKNOWLEDGMENTS

We are grateful to the audiences of the 2014 Second Language Research Forum, the University of Maryland Cognitive Neuroscience of Language Lab, the University of Utah Speech Acquisition Lab, and the reviewers and editor for their comments and suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00995>

REFERENCES

- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). *lme4: Linear Mixed-Effects Models Using Eigen and S4*. R package version 1.1-1.10.
- Best, C. T. (1995). "A direct realist perspective on cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed W. Strange (Timonium, MD: York Press), 171–204.
- Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: commonalities and complementarities," in *Language Experience and Second Language Speech Learning: In Honor of James Emil Flege, Volume 17 of Language Learning and Language Teaching*, eds O.-S. Bohn and M. J. Munro (Philadelphia, PA: John Benjamins Publishing Company), 13–34.
- Boersma, P., and Weenink, D. (2009). *Praat: Doing Phonetics by Computer [Computer Program]*. Version 5.1.13.
- Boomershine, A., Currie Hall, K., Hume, E., and Johnson, K. (2008). "The impact of allophony versus contrast on speech perception," in *Contrast in Phonology: Theory, Perception, Acquisition, Volume 13 of Phonology & Phonetics*, eds P. Avery, B. E. Dresher, and K. Rice (Berlin: Mouton de Gruyter), 145–172.
- Carrasco, P., Hualde, J. I., and Simonet, M. (2012). Dialectal differences in Spanish voiced obstruent allophony: Costa Rican versus Iberian Spanish. *Phonetica* 69, 149–179. doi: 10.1159/000345199
- de Cheveigné, A., and Simon, J. Z. (2007). Denoising based on time-shift PCA. *J. Neurosci. Methods* 165, 297–305. doi: 10.1016/j.jneumeth.2007.06.003
- de Cheveigné, A., and Simon, J. Z. (2008). Sensor noise suppression. *J. Neurosci. Methods* 168, 195–202. doi: 10.1016/j.jneumeth.2007.09.012
- Eckman, F. R., Elreyes, A., and Iverson, G. K. (2001). Allophonic splits in L2 phonology: the question of learnability. *Int. J. English Stud.* 1, 21–51.
- Eckman, F. R., Elreyes, A., and Iverson, G. K. (2003). Some principles of second language phonology. *Second Lang. Res.* 19, 169–208. doi: 10.1191/0267658303sr2190a
- Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization*. Ph.D. thesis, Utrecht University, Netherlands.
- Eulitz, C., and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16, 577–583. doi: 10.1162/0899892904323057308

- Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., and Morgan, J. L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition* 127, 427–438. doi: 10.1016/j.cognition.2013.02.007
- Flege, J. E. (1995). “Second language speech learning: theory, findings, and problems,” in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed W. Strange (Timonium, MD: York Press), 233–276.
- Flege, J. E., Munro, M. J., and Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *J. Acoust. Soc. Am.* 95, 3623–3641.
- Forster, K. I., and Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behav. Res. Methods* 35, 116–124. doi: 10.3758/BF03195503
- García-Sierra, A., Ramírez-Esparza, N., Silva-Pereyra, J., Siard, J., and Champlin, C. A. (2012). Assessing the double phonemic representation in bilingual speakers of Spanish and English: an electrophysiological study. *Brain Lang.* 121, 194–205. doi: 10.1016/j.bandl.2012.03.008
- Goto, H. (1971). Auditory perception by normal Japanese adults with sounds “L” and “R.” *Neuropsychologia* 9, 317–323.
- Hammerly, H. (1982). Contrastive phonology and error analysis. *Int. Rev. Appl. Linguist. Lang. Teach.* 20, 17–34.
- Hardy, J. E. (1993). “Phonological learning and retention in second language acquisition,” in *Confluence: Linguistics, L2 Acquisition and Speech Pathology, Volume 4 of Language Acquisition & Language Disorders*, ed F. R. Eckman (Philadelphia, PA: John Benjamins Publishing Company), 235–247.
- Harnsberger, J. D. (2001). The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: a multidimensional scaling analysis. *J. Phon.* 29, 303–327. doi: 10.1006/jpho.2001.0140
- Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Lang. Res.* 23, 65–94. doi: 10.1177/0267658307071601
- Herd, W. (2011). *The Perception and Production Training of /d/, /t/, and /r/ in L2 Spanish: Behavioral, Psycholinguistic, and Neurolinguistic Evidence*. Ph.D. thesis, University of Kansas.
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97:3099. doi: 10.1121/1.411872
- Hisagi, M., Shafer, V. L., Strange, W., and Sussman, E. S. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: behavioral and electrophysiological measures. *Brain Res.* 1360, 89–105. doi: 10.1016/j.brainres.2010.08.092
- Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biom. J.* 50, 346–363. doi: 10.1002/bimj.200810425
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for nonnative phonemes. *Cognition* 83, B47–B57. doi: 10.1016/S0010-0277(02)00198-1
- Johnson, K., and Babel, M. (2010). On the perceptual basis of distinctive features: evidence from the perception of fricatives by Dutch and English speakers. *J. Phon.* 38, 127–136. doi: 10.1016/j.wocn.2009.11.001
- Jurafsky, D., and Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Series In Artificial Intelligence, Pearson Education, Inc., 2nd Edn.* Upper Saddle River, NJ: Prentice Hall.
- Kazanina, N., Phillips, C., and Idsardi, W. J. (2006). The influence of meaning on the perception of speech sounds. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11381–11386. doi: 10.1073/pnas.0604821103
- Kenstowicz, M. (1994). *Phonology in Generative Grammar*. Cambridge, MA: Blackwell Publishers.
- Knoblauch, K. (2007). *Psyphy: Functions for Analyzing Psychophysical Data in R*. R package version 0.1–7.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–608. doi: 10.1126/science.1736364
- Kuznetsova, A., Bruun Brockhoff, P., and Haubo Bojesen Christensen, R. (2014). *lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models (lmer Objects of lme4 Package)*. R package version 2.0–11.
- Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, MI: University of Michigan Press.
- Lavoie, L. M. (2001). *Consonant Strength: Phonological Patterns and Phonetic Manifestations*. New York, NY: Garland Publishing, Inc.
- Lisker, L., and Abramson, A. S. (1970). “The voicing dimension: some experiments in comparative phonetics,” in *Proceedings of the sixth International Congress of Phonetic Sciences* (Prague: Academia), 563–567.
- Macmillan, N. A., and Creelman, C. (2005). *Detection Theory: A User's Guide, 2nd Edn.* Mahwah, NJ: Lawrence Erlbaum Associates.
- Maye, J., and Gerken, L. (2000). “Learning phonemes without minimal pairs,” in *Proceedings of the 24th Annual Boston University Conference on Language Development*, Vol. 2, eds S. C. Howell, S. A. Fish, and T. Keith-Lucas (Somerville, MA: Cascadia Press), 522–533.
- Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect discrimination. *Cognition* 82, B101–B111. doi: 10.1016/S0010-0277(01)00157-3
- McLaughlin, J., Osterhaut, L., and Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat. Neurosci.* 7, 703–704. doi: 10.1038/nn1264
- Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huottilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434. doi: 10.1038/385432a0
- Näätänen, R., Pakarinen, S., Rinne, T., and Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clin. Neurophysiol.* 115, 140–144. doi: 10.1016/j.clinph.2003.04.001
- Näätänen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Patterson, D., and Connine, C. (2001). Variant frequency in flap production. *Phonetica* 58, 254–275. doi: 10.1159/000046178
- Pegg, J. E., and Werker, J. F. (1997). Adult and infant perception of two English phones. *J. Acoust. Soc. Am.* 102, 3742–3753.
- Peperkamp, S., Le Calvez, R., Nadal, J.-P., and Dupoux, E. (2006). The acquisition of allophonic rules: statistical learning with linguistic constraints. *Cognition* 101, B31–B41. doi: 10.1016/j.cognition.2005.10.006
- Peperkamp, S., Pettinato, M., and Dupoux, E. (2003). “Allophonic variation and the acquisition of phoneme categories,” in *Proceedings of the 27th Annual Boston University Conference on Language Development*, Vol. 2, eds B. Beachley, A. Brown, and F. Conlin (Somerville, MA: Cascadia Press), 650–661.
- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *J. Cogn. Neurosci.* 12, 1038–1055. doi: 10.1162/089892900051137567
- Polka, L., and Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 421–435. doi: 10.1037/0096-1523.20.2.421
- Quilis, A., and Esgueva, M. (1983). Realización de los fonemas vocálicos españoles en posición fonética normal. *Estud. Fonética* 1, 159–252.
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.
- Seidl, A., and Cristia, A. (2012). Infants’ learning of phonological status. *Front. Psychol.* 3:448. doi: 10.3389/fpsyg.2012.00448
- Seidl, A., Onish, K. H., Alamian, G., and Cristia, A. (2014). Acoustic correlates of allophonic versus phonemic dimensions in monolingual and bilingual infants’ input. *J. Phonet.* 45, 43–51. doi: 10.1016/j.wocn.2014.03.004
- Shea, C. E. (2014). Second language learners and the variable speech signal. *Front. Psychol.* 5:1338. doi: 10.3389/fpsyg.2014.01338
- Shea, C. E., and Curtin, S. (2010). Discovering the relationship between context and allophones in a second language: evidence for distribution-based learning. *Stud. Second Lang. Acquis.* 32, 581–606. doi: 10.1017/S0272263110000276
- Shea, C. E., and Curtin, S. (2011). Experience, representations and the production of second language allophones. *Second Lang. Res.* 27, 229–250. doi: 10.1177/0267658310375753

- Shea, C. E., and Renaud, J. (2014). L2 perception of Spanish palatal variants across different tasks. *Bilingualism* 17, 203–221. doi: 10.1017/S1366728913000047
- Simonet, M., Hualde, J. I., and Nadeu, M. (2012). “Lenition of /d/ in spontaneous Spanish and Catalan,” in *Proceedings of Interspeech* (Portland, OR: International Speech Communication Association), 1416–1419.
- Stathopoulos, E. T., and Weismer, G. (1983). Closure duration of stop consonants. *J. Phonet.* 11, 395–400.
- Strange, W., and Shafer, V. L. (2008). “Speech perception in second language learners: the re-education of selective attention,” in *Phonology in Second Language Acquisition, Vol. 36 in Studies in Bilingualism*, eds J. G. Hansen Edwards and M. L. Zampini (Philadelphia, PA: John Benjamins Publishing Company), 153–191.
- Sussman, E. S., Chen, S., Sussman-Fort, J., and Dinces, E. (2014). The five myths of MMN: redefining how to use MMN in basic and clinical research. *Brain Topogr.* 27, 553–564. doi: 10.1007/s10548-013-0326-6
- Waltmunson, J. (2005). *The Relative Difficulty of Spanish/t, d/, Trill and Tap by L1 English Speakers: Auditory and Acoustic Methods of Defining Pronunciation Accuracy*. Ph.D. thesis, University of Massachusetts, Amherst, MA.
- Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Dev.* 5, 349–355.
- Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63.
- Whalen, D. H., Best, C. T., and Irwin, J. R. (1997). Lexical effects in the perception and production of American English /p/ allophones. *J. Phonet.* 25, 501–528. doi: 10.1006/jpho.1997.0058
- White, K. S., Peperkamp, S., Kirk, C., and Morgan, J. L. (2008). Rapid acquisition of phonological alternations by infants. *Cognition* 107, 238–265. doi: 10.1016/j.cognition.2007.11.012
- Winkler, I., Kujala, T., Alku, P., and Nääätänen, R. (2003). Language context and phonetic change detection. *Cogn. Brain Res.* 17, 833–844. doi: 10.1016/S0926-6410(03)00205-2
- Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36, 638–642.
- Yeung, H. H., and Werker, J. F. (2009). Learning words’ sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorization speech information. *Cognition* 113, 234–243. doi: 10.1016/j.cognition.2009.08.010
- Zampini, M. L. (1994). The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania* 77, 470–481.
- Zampini, M. L. (1996). Voiced stop spiratization in the ESL speech of native speakers of Spanish. *Appl. Psycholinguist.* 17, 335–354.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Barrios, Namyst, Lau, Feldman and Idsardi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



More Limitations to Monolingualism: Bilinguals Outperform Monolinguals in Implicit Word Learning

Paola Escudero^{1,2*}, Karen E. Mulak^{1,2}, Charlene S. L. Fu³ and Leher Singh³

¹ The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith, NSW, Australia, ² Centre of Excellence for the Dynamics of Language, Australian Research Council, Canberra, ACT, Australia, ³ Department of Psychology, National University of Singapore, Singapore, Singapore

OPEN ACCESS

Edited by:

Annie Tremblay,
University of Kansas, USA

Reviewed by:

Margarita Kaushanskaya,
University of Wisconsin-Madison,
USA

Henrike Katina Blumenfeld,
San Diego State University, USA

*Correspondence:

Paola Escudero
paola.escudero@westernsydney.
edu.au

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 12 February 2016

Accepted: 02 August 2016

Published: 15 August 2016

Citation:

Escudero P, Mulak KE, Fu CSL and
Singh L (2016) More Limitations
to Monolingualism: Bilinguals
Outperform Monolinguals in Implicit
Word Learning.
Front. Psychol. 7:1218.
doi: 10.3389/fpsyg.2016.01218

To succeed at cross-situational word learning, learners must infer word-object mappings by attending to the statistical co-occurrences of novel objects and labels across multiple encounters. While past studies have investigated this as a learning mechanism for infants and monolingual adults, bilinguals' cross-situational word learning abilities have yet to be tested. Here, we compared monolinguals' and bilinguals' performance on a cross-situational word learning paradigm that featured phonologically distinct word pairs (e.g., BON-DEET) and phonologically similar word pairs that varied by a single consonant or vowel segment (e.g., BON-TON, DEET-DIT, respectively). Both groups learned the novel word-referent mappings, providing evidence that cross-situational word learning is a learning strategy also available to bilingual adults. Furthermore, bilinguals were overall more accurate than monolinguals. This supports that bilingualism fosters a wide range of cognitive advantages that may benefit implicit word learning. Additionally, response patterns to the different trial types revealed a relative difficulty for vowel minimal pairs than consonant minimal pairs, replicating the pattern found in monolinguals by Escudero et al. (2016) in a different English accent. Specifically, all participants failed to learn vowel contrasts differentiated by vowel height. We discuss evidence for this bilingual advantage as a language-specific or general advantage.

Keywords: monolinguals, simultaneous bilinguals, implicit word learning, minimal pairs, phonetic detail, bilingual advantage

INTRODUCTION

Typically, a person has learned 10s of 1000s of words by adulthood. While many of these words are learned explicitly, through instruction or clear, coinciding presentation of the word and its referent, not all words are learned in this manner. Some words are learned implicitly, by tracking the occurrence of an auditory word across multiple presentations in the context of multiple candidate referents. Humans are powerful statistical learners, and through this ability can implicitly derive the most likely referent of a novel word based on the likelihood of a candidate referent occurring simultaneously with an auditory word.

This type of learning, commonly termed *cross-situational word learning* (XSWL), appears staggering when one considers that the world presents learners with a seemingly infinite number of candidate referents for a single word in any one moment in time (Quine, 1960). Nonetheless, evidence shows that both infants (Smith and Yu, 2008; Vouloumanos and Werker, 2009;

Vlach and Johnson, 2013) and adults (Yu and Smith, 2007; Smith et al., 2011; Suanda and Namy, 2012; Yurovsky et al., 2013; Dautriche and Chemla, 2014) can learn novel words through XSWL.

In a typical XSWL experiment, participants are presented with a series of ambiguous learning trials consisting of multiple objects and multiple words, with no explicit indication of word-object correspondences. During the learning phase, participants are not given instruction with regard to the nature of the task, and instead are simply asked to view the trials. After the learning phase, participants are presented with a forced-choice test in which they are asked to identify object-label mappings.

Studies on XSWL have typically included words that contained gross phonological differences (e.g., BLICKET vs. GAX; Smith and Yu, 2008; Vlach and Johnson, 2013). For pairs like this, listeners do not need to pay attention to fine phonological detail to differentiate competitor words and therefore do not need to pay attention to such information to allow learning. However, real-world word learning requires that words be encoded with fine phonological detail due to the presence of many phonologically overlapping words. The most extreme case of phonological overlap is seen in minimal pairs, in which words differ by only a single segment (e.g., TIP-DIP or TIP-TAP).

Recently, Escudero et al. (2016) asked whether adults in Sydney, Australia could learn novel words produced in Australian English via XSWL while simultaneously encoding fine phonological detail. In their experiment, participants viewed two side-by-side novel images during training, and heard the novel name associated with each image, without indication as to whether the words were named left-to-right or right-to-left. The words comprised eight CVC words in which four words differed by only one consonant (BON, DON, PON, TON), and the other four differed by one vowel (DEET, DIT, DOOT, DUT). During the test, in each trial the named image was paired with a distractor image. Based on the word associated with each image, this target-distractor pair formed either a non-minimal pair, in which two or all three segments differed (e.g., BON-DEET, DON-DEET), a consonant minimal pair, in which the initial consonant differed (e.g., BON-DON), or a vowel minimal pair, in which the vowel differed (e.g., DEET-DIT). Escudero et al. (2016) found that adults were able to learn all pair types via XSWL, but that performance was weakest in the context of a vowel minimal pair, indicating that phonological encoding of vowels was weaker than encoding of consonants.

Like monolinguals, bilinguals most certainly can and do learn words via cross-situational learning. However, it is unclear whether or how exposure or mastery of more than one language affects their learning relative to monolinguals. Bilingualism is often associated with greater performance on tests of executive function, selective attention and inhibitory control (e.g., Bialystok et al., 2006; Carlson and Meltzoff, 2008; Bialystok and Viswanathan, 2009; Bialystok and Craik, 2010). For instance, in the Stroop task, the names of colors are presented on a screen, and the color of the text either matches or mismatches the written color. Participants are then asked to name the color of the text, rather than to read the written word. Compared to monolinguals, bilinguals named the color of the text more

quickly when the color of the text did not match the written color (Bialystok et al., 2008).

Bilingual advantages have been found in the linguistic domain as well. Using an explicit novel word learning paradigm, Kaushanskaya and Marian (2009) taught monolingual English speakers and early Spanish-English and Mandarin-English bilinguals 48 novel auditory words constructed from an artificial phonological system unfamiliar to all groups. After hearing each word, participants were shown its English orthographic translation. During the test phase, participants heard one of the novel words and were asked to select its English orthographic translation from five options. Both the English-Spanish and English-Mandarin bilinguals outperformed the English monolinguals when tested immediately after the learning phase, and 1 week later. In a follow-up study, Spanish-English bilinguals also outperformed English monolinguals when the words were comprised of phonemes that occurred in both English and Spanish (Kaushanskaya, 2012).

Kaushanskaya (2012) proposed that bilinguals' advantage in novel word learning may be due to an enhanced phonological short term memory. Indeed, this proposal corresponds to research demonstrating that bilingualism confers gains in phonological working memory (Service et al., 2002; Majerus et al., 2008; Adesope et al., 2010), and also to research showing that multilinguals demonstrate better performance in digit-span and non-word repetition tasks (Papagno and Vallar, 1992). To test this proposal, Kaushanskaya (2012) divided monolinguals into high- and low-span phonological memory groups and tested them alongside bilinguals in their learning of novel phonologically familiar and unfamiliar words. Bilinguals outperformed both groups of monolinguals, suggesting that the bilingual advantage on this task may not be sufficiently explained by differences in phonological memory span.

But at a conceptual level, bilingualism might be expected to result in poorer or slower performance in some language abilities relative to monolinguals due to increased competition. During the course of spoken word recognition, competitor words are activated. For instance, the word CAT is activated during perception of the word CATALOG (e.g., Norris and McQueen, 2008). Because bilinguals have a lexicon in each language, there are more potential words that could be activated in the bilingual lexicon relative to monolinguals. Spoken word recognition is more difficult with increasing activation of competitors (Luce and Pisoni, 1998), and in the same way, the enlarged lexical space of bilinguals could be expected to interfere more with novel word learning. However, as described above, bilinguals typically show equal or enhanced word learning relative to monolinguals, possibly suggesting that they are able to suppress competitor activation in the non-target language. The general advantages in executive control discussed above may emerge from the need to control access and parallel activation between the bilinguals' two languages, which takes place through enhanced attention to one language and/or inhibition of the other (e.g., Bialystok and DePape, 2009; Costa et al., 2009; Festman et al., 2010; Blumenfeld and Marian, 2011; Kroll and Bialystok, 2013; Duncan et al., 2016). Indeed, the areas of the brain involved in domain general executive control significantly overlap with the areas used in

language control in bilinguals (Bialystok et al., 2012; Pliatsikas and Luk, 2016).

Experimental support for the suggestion that the executive control advantages commonly found in bilinguals are linked to their negotiation of access to their two languages comes from a word learning experiment in which Spanish–English bilinguals and English monolinguals learned novel translations for pictures of known items. At test, participants heard a newly learned word while viewing a target and distractor item, and were asked to click on the corresponding item. In some trials, the familiar word associated with the distractor image overlapped phonologically with the target word (e.g., the new word SHUNDO was associated with a picture of an acorn, and during test the acorn image was paired with a picture of a shovel, which shares the same onset as SHUNDO). Although, bilinguals had more phonological competitor words compared to monolinguals through their knowledge of words in two languages rather than just one, bilinguals looked less to the competitor images than did monolinguals, and mouse-tracking results showed that they tracked more directly to the target image (Bartolotti and Marian, 2012).

While bilingual advantages have been demonstrated in explicit word learning tasks, it is not clear how bilingualism might affect *implicit* word learning. Bartolotti et al. (2011) compared monolingual and bilinguals' ability to extract and learn novel words composed of pure tones based on Morse code by tracking transitional probabilities in a continuous auditory stream. Participants with high bilingual experience, defined as higher reported L2 proficiency, earlier age of L2 acquisition, and higher frequency of L2 use, were better at learning words through tracking transitional probabilities than those with low bilingual experience. Inhibitory control strength (as measured by the Simon task) did not affect performance. When the bilingual participants were subsequently exposed to a different Morse code auditory stream containing conflicting transitional probabilities compared to the first stream, strength of inhibitory control (but not bilingual experience) aided performance, presumably through participants' ability to suppress the influence from the former Morse code "language." The authors proposed that the contribution of bilingual experience was perhaps due to increased phonological working memory. Although, this does not appear to explain the bilingual advantage for explicit word learning, it may have more of an effect on implicit learning (Bartolotti et al., 2011). Alternatively, while effects of increased phonological memory and enhanced executive functioning do not reliably explain the bilingual advantage when compared to skill-matched monolingual peers, bilingualism may nonetheless support these cognitive skills such that they are stronger in the bilingual population as a whole compared to monolinguals (see Kaushanskaya, 2012).

Importantly, bilingual advantages are not always found. With regard to the ability to form pairings between stimuli – a skill inherent to cross-situational word learning – there have been instances of finding no bilingual advantages in learning of non-linguistic tone-symbol pairings (Blumenfeld and Adams, 2014) and novel word-abstract referent pairings (Kaushanskaya and Reetzgel, 2012). As well, a review of the existing literature

investigating bilingual advantages in enhanced executive control found inconsistent evidence of such an advantage (Hilchey and Klein, 2011), and this has been further supported through subsequent empirical research (Kousaie and Phillips, 2012a,b; Paap and Greenberg, 2013). Paap (2014) recently proposed that the generally accepted notion of a bilingual advantage, at least in executive functioning, may be the result of a publication bias. This is supported by a meta-analysis of subsequent publication rates of studies submitted as conference abstracts, based on whether their findings supported or challenged the notion of a bilingual advantage in executive functioning (de Bruin et al., 2015). Alarming, the analysis showed a clear publication bias. While the number of conference abstracts supporting and challenging the bilingual advantage in executive functioning were similar (54 vs. 50, respectively), 63% of the studies in support of the bilingual advantage went on to be published as full journal articles, compared to only 36% of the studies that challenged the bilingual advantage. Thus, bilingual advantages in executive functioning, and perhaps in other areas, are very likely not as pervasive, and are likely weaker, than has been generally understood.

One important factor that may influence whether a bilingual advantage is measured in the linguistic domain is the relationship between the linguistic stimuli and the listeners' phonological space. Models such as PAM (Perceptual Assimilation Model; Best, 1994, 1995), its extension to non-native and second language (L2) learning (PAM-L2; Best and Tyler, 2007) and L2LP (Second Language Linguistic Perception model; Escudero, 2005, 2009; van Leussen and Escudero, 2015) say that perception of non-native contrasts that do not exist in a learner's native language is generally expected to be worse than perception of non-native contrasts that have a counterpart in the learner's native language (though both models claim that the relationship between native and non-native phones predicts perception of specific non-native contrasts). Thus, infants, children, and adults who learn two languages from birth may have more difficulty or fail to show an advantage if a contrast is absent in one or both of their languages. By extension of this proposal, research comparing novel word learning in monolingual and bilingual infants has shown that when bilinguals are familiar with phonological contrasts in both test languages, they outperform monolinguals (Mattock et al., 2010; Singh et al., 2016). However, other research has found that bilingual infants exposed to English and another language are delayed in novel word learning of minimal pairs relative to English monolinguals, and that this delay is independent of the similarity of the English phonological contrast being tested with the analogous contrast in their second language (Fennell et al., 2007). It remains an open question whether the phonological status of a contrast affects a possible bilingual advantage in adulthood, and whether factors such as language dominance and age of acquisition of the L2 correlate with any such effect.

In Escudero et al. (2016) examination of cross-situational word learning of minimal pair words, 40 of the 71 total participants reported proficiency in one or more languages in addition to English; however, no effect of bi/multilingualism was found. While this lack of a bilingual advantage may reflect a lack of a bilingual advantage in implicit word learning, the null

result may instead stem from the heterogeneity of the bilingual sample with regard to several factors that may be related to cognitive advantages associated with bilingualism. For instance, age of acquisition of a second language (L2) has been shown to affect performance on the flanker task (Eriksen and Eriksen, 1974), which measures response inhibition. Early bilinguals who acquired their L2 before the age of 10 outperformed late bilinguals and monolinguals (Luk et al., 2011). As well, bilinguals who switch between their languages more frequently outperform those who switch less frequently in measures of executive control (e.g., Prior and Gollan, 2011; Verreyt et al., 2016).

To test whether a bilingual advantage occurs in implicit word learning, we compared performance by Australian English monolinguals from Sydney, Australia with a homogeneous population of Singaporean English–Mandarin simultaneous bilinguals from Singapore. We tested their XSWL of the same non-minimal and minimal pair words used by Escudero et al. (2016), but produced by an American English speaker, so that the accent would not be native to either group, but would be familiar to both groups (e.g., through media). Thus, in line with research demonstrating a bilingual advantage in explicit word learning (Kaushanskaya and Marian, 2009; Kaushanskaya, 2012), and based on our supposition that XSWL would be aided by executive functioning features, for which bilinguals have often been found to have an advantage over monolinguals (e.g., Adesope et al., 2010; Kroll and Bialystok, 2013), it was predicted that our bilingual participants would outperform our monolingual participants when learning novel word pairs in an implicit learning paradigm, at least when tested in a non-minimal pair context. Secondly, we predicted that accuracy by both groups would be poorest for vowel minimal pair trials, which would replicate the finding by Escudero et al. (2016). Lastly, all words in the present study were comprised of phonemes present in (or analogous to) Australian English (Cox and Palethorpe, 2007), Singaporean English (Wee, 2004, 2010; Deterding, 2007) and Standard Mandarin (Duanmu, 2002; which is phonologically similar to Singaporean Mandarin), with the exception that the vowels /ɪ/ and /ʊ/, found in the novel words DIT and DUT are not present in Singaporean English or Standard Mandarin. The L2LP model (Escudero, 2005; van Leussen and Escudero, 2015) predicts that non-native vowels may be perceived as acoustically proximate native vowels. While Mandarin does contain the vowel /ɣ/, which is acoustically proximate to /ʊ/, the most acoustically proximate vowel to /ɪ/ is /i/, as in DEET, which may lead to confusion in learning and discriminating DIT–DEET, which is differentiated by vowel height. Thus, we predicted that our Singaporean bilinguals would show poorer performance for vowel contrasts differentiated by height relative to monolinguals.

MATERIALS AND METHODS

Participants

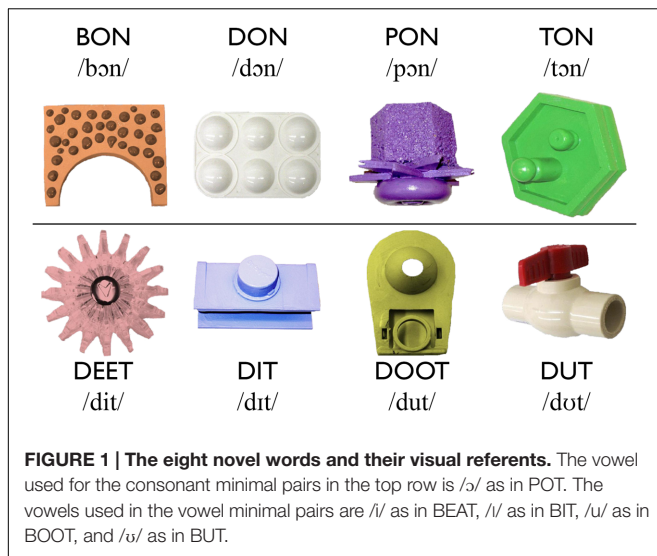
All participants were native English speakers at English-language universities. Monolingual participants were 16 monolingual Australian English (henceforth, AusE) speakers

aged 17.1–37.0 years ($M_{\text{age}} = 24.3$, $SD = 5.9$, 10 females) who were primarily undergraduate students at Western Sydney University. These participants received course credit or \$10 travel compensation for their participation. Bilingual participants were 15 simultaneous Singaporean English–Mandarin (henceforth, SE–SM) bilinguals aged 20.0–23.5 years who were undergraduate students from the National University of Singapore ($M_{\text{age}} = 21.6$, $SD = 1.1$, 9 females). These participants received \$5 SGD compensation for participation. None of the participants reported a history of hearing or language impairment. Participants' language background was determined via a language background questionnaire administered at the beginning of the session. Participants were determined to be AusE monolinguals if all parents or caretakers were born in Australia and were native speakers of AusE, and if the participant reported that during childhood they did not regularly spend time with someone whose native language was not AusE (e.g., a close relative or family friend, and/or someone who lived with them). Participants were determined to be SE–SM simultaneous bilinguals if they received exposure to both SE and SM by 2 years of age, and reported current proficiency in both SE and SM. When asked to rate their oral comprehension and productive proficiency on a seven-point Likert scale (7 = native), monolinguals' average rating for their English comprehension ability was 7.0 ($SD = 0.0$), and was 6.9 ($SD = 0.2$) for their productive ability. On average, bilinguals reported their English comprehension ability as 6.7 ($SD = 0.7$), and their production ability as 6.6 ($SD = 0.7$), and these values did not significantly differ from monolinguals' ratings. Bilinguals rated their Mandarin comprehension ability as 5.8 ($SD = 1.4$), and their production ability as 5.1 ($SD = 1.7$). Participants gave informed consent prior to participation in accordance to the Western Sydney University Human Research and Ethics Committee and National Singapore University Institutional Review Board.

Stimuli

Novel Words

Eight monosyllabic nonsense words were recorded by a female native speaker of American English. As shown in **Figure 1**, the words followed a CVC structure, and adhered to English phonotactics. The words have been used in previous research on the acquisition of minimal pairs (Curtin et al., 2009; Fikkert, 2010), including in a cross-situational word learning context, which used the same set of words recorded by a native female speaker of AusE, produced with the same intonation contours as the present study (Escudero et al., 2016, under review). Four of the words differed minimally in their first consonant, whereas the other four differed in their vowel. All words were comprised of phonemes present in (or analogous to) AusE (Cox and Palethorpe, 2007), SE (Deterding, 2007; Wee, 2010) and Standard Mandarin (Duanmu, 2002; which is phonologically similar to SM), with the exception that the vowels /ɪ/ and /ʊ/ found in the novel words DIT and DUT are not present in SE or Mandarin, though Mandarin does contain the vowel /ɣ/, which is acoustically proximate to /ʊ/. Two tokens of each of the eight



spoken words were selected for use in the experiment so that intonation contours were comparable across words.

Novel Visual Referents

The visual referents for the words were pictures of novel items used in previous studies on XSWL (Vlach and Sandhofer, 2014; Escudero et al., 2016, under review). Each nonsense word was randomly paired once with a visual referent (**Figure 1**). The same word-referent pairings were presented to all participants, and were the same pairs used in previous studies on cross-situational word learning (Escudero et al., 2016, under review). Each image measured 280×274 pixels. Slides were created in which two of the eight visual referents were placed on an 800×600 -pixels white background with the top-left corner of the left images positioned at 20×163 pixels, and the top-left corner of the right image positioned at 500×163 pixels.

Attention Videos

Each attention video consisted of a looped cartoon animation measuring 170×170 pixels, which was centered on the monitor between every third trial in the learning phase and between each trial in the testing phase. Each animation was paired with a non-linguistic sound.

Procedure

The procedure was identical to that reported in Escudero et al. (2016), and consisted of a learning phase and testing phase. Examples of learning and testing phase trials can be seen in **Figure 2**. At the beginning of the experiment, participants were seated in front of a 19-in. display and were told that they would watch some images on the screen and hear some words. Participants were not told that the words were names for the images, nor were they asked to try and discover which word was paired with which image.

Learning Phase

The learning phase consisted of 36 trials, across which participants were presented with each word-referent pairing nine

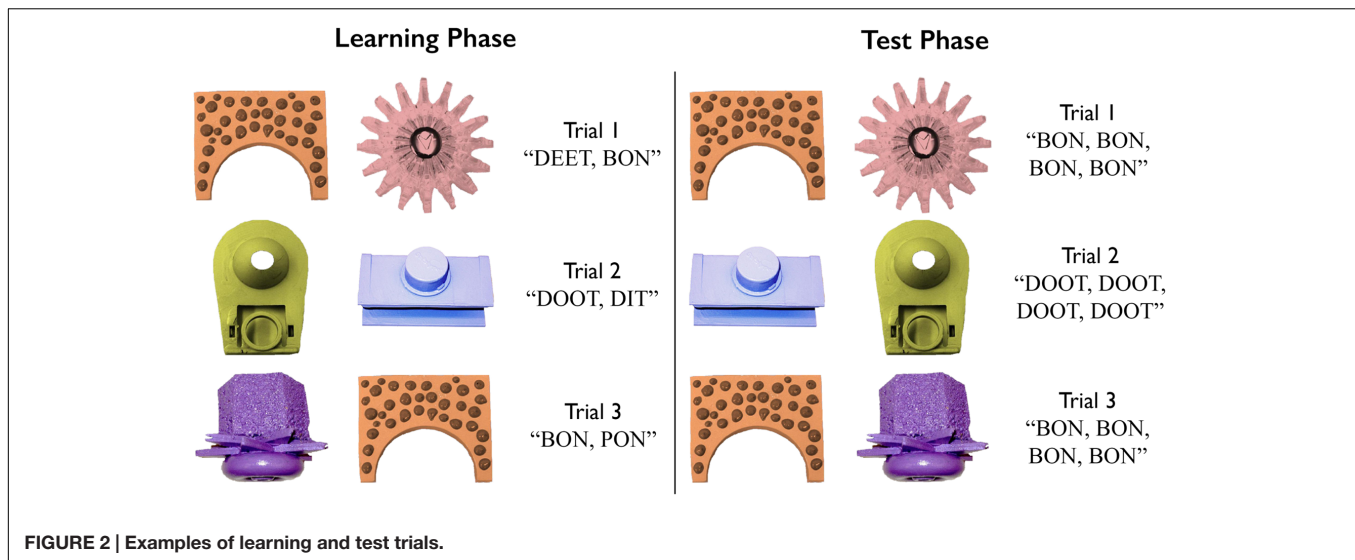
times. In each learning trial, two of the eight visual referents were displayed on the screen. After 500 ms, the word corresponding to each item was spoken so that each picture was named once, either left-to-right, or right-to-left, with 500 ms between spoken words. There were no cues that signaled whether the visual referents were named left-to-right or right-to-left.

The presentation order of the paired trials was randomized for each participant and the pairings were controlled such that each visual referent occurred with every other visual referent at least once, and no more than twice. If the same pairing occurred more than once, the designations of the left and right image were swapped so that participants never saw the exact same visual pairing more than once. As each word appeared nine times, the occurrence of an image in the left or right position was balanced such that half of the words appeared five times on the left and four times on the right, while the other half appeared in the opposite pattern. Whether a visual referent was named first or second, and the number of times each of the two tokens of each nonsense word were heard, were also balanced.

The two words presented in each trial belonged to one of three possible phonological relationships when paired: non-minimal pairs (non-MPs) differed in two or all three segments (e.g., BON-DEET, DON-DEET); consonant minimal pairs (cMPs) differed in their initial consonant (BON-TON), and; vowel minimal pairs (vMPs) differed in their vowel (DEET-DIT). Further, cMPs differed either by place (BON-DON, PON-TON), voicing (BON-PON, DON-TON), or both place and voicing (DON-PON, TON-BON), and vMPs differed by height (DEET-DIT, DOOT-DUT), backness (DEET-DOOT, DUT-DIT), or both height and backness (DUT-DEET, DIT-DOOT). During training, participants were exposed to 24 non-MPs, and all 6 cMPs and 6 vMPs, for a total of 36 pairs. Each learning trial lasted 3.5 s and an attention getter comprising a 170×170 centrally presented looped video paired with a non-linguistic sound, played between every third trial until the participant's gaze was centrally fixed. The total duration of the learning phase was approximately 3 min. Examples of training trials are presented in **Figure 2**.

Test Phase

After completion of the learning phase, participants were seated in front of a laptop computer with a 15-in. monitor. Participants were instructed that they would see two images on the screen from the same set of images they had just watched. They were told they would hear the name corresponding to one of the images, and should indicate by pressing the left or right ALT key whether they believed the word corresponded to the image on the left or right, respectively. Test trials contained the same pairs of two words and visual referents as the learning phase, but the left and right designations of the images were randomized once such that for half of the trials, the order of the images was swapped relative to the training phase. Each participant received the same test trials, presented in three counterbalanced blocks of 12, with the trials within each block occurring in a random order. For each trial, once the two images had been on the screen for 500 ms, two tokens of the spoken word corresponding to one of the images (the target object) played twice each in an alternating fashion with 500 ms between each repetition, such



that the participant heard the word a total of four times. Each word served as the target four or five times. As in the training phase, the test consisted of 24 non-MP trials, 6 cMP trials, and 6 vMP trials. Each trial lasted 6.5 s, resulting in a test phase duration of approximately 4 min. Examples of test trials are presented in **Figure 2**.

RESULTS

No-response trials, which comprised 1.3% of the total sample, were removed from analysis. To examine whether there were differences in word learning performance between non-MP and MP trials, and to compare bilinguals' performance relative to monolinguals, participants' correct and incorrect responses were analyzed in a mixed-effects binary logistic model with pair type (non-MP, MP) and language background (monolingual vs. bilingual) as fixed variables, and subject, order, target, distractor, and target location as random variables. A separate independent-samples *t*-test revealed a trend such that the monolingual group was marginally older than the bilingual group ($t[16.17] = 1.76$, $p = 0.098$, $[-0.5, 5.8$ years]). Thus, age was entered in the mixed-effects model as a random covariate. As can be seen in **Figure 3**, the model revealed a main effect of pair type [$\chi^2(1, n = 1101) = 4.49$, $p = 0.034$], with greater accuracy for non-MP than MP trials. There was also an effect of language background [$\chi^2(1, n = 1101) = 5.02$, $p = 0.025$]. Overall, bilinguals were more accurate than monolinguals. There was no interaction of language background and pair type [$\chi^2(1, n = 1101) = 1.66$, $p = 0.198$]. One-sample *t*-tests against chance showed that proportion fixation to the named image was above chance for both non-MP and MP trials, for both monolinguals (non-MP: $t[15] = 4.67$, $p < 0.001$, 95% CI $[0.12, 0.32]$; MP: $t[15] = 4.52$, $p < 0.001$, $[0.11, 0.31]$) and bilinguals (non-MP: $t[14] = 12.49$, $p < 0.001$, $[0.29, 0.41]$; MP: $t[14] = 7.61$, $p < 0.001$, $[0.21, 0.38]$). Thus, all learners were able to infer word-object pairings for both non-MP and MP trials.

Participants' reaction times (RTs) for correct responses, which are shown in **Figure 4**, were analyzed in a mixed-effects linear model with the same fixed and random factors and random covariate as in the accuracy analysis. There was no main effect of pair type [$\chi^2(1, n = 857) = 2.40$, $p = 0.121$], or language background [$\chi^2(1, n = 857) = 1.24$, $p = 0.266$], and no interaction between the two [$\chi^2(1, n = 857) = 1.79$, $p = 0.181$].

To answer our next question of whether performance differed depending on whether the MP differed in one consonant or one vowel, and whether participants' language background affected performance, participants' correct and incorrect responses for MP trials were analyzed in a mixed-effects binary logistic model with MP type (cMP, vMP) and language background (monolingual vs. bilingual) as fixed variables, and subject, order, target, distractor, and target location as random variables, and age entered as a random covariate. As shown in **Figure 5**, the model revealed a main effect of MP type [$\chi^2(1, n = 368) = 5.01$, $p = 0.025$], with greater accuracy for cMP than vMP trials. There was no effect of language background [$\chi^2(1, n = 368) = 0.82$, $p = 0.366$], and no interaction of language background and minimal pair type [$\chi^2(1, n = 368) = 0.61$, $p = 0.435$]. One-sample *t*-tests against chance showed that proportion fixation to the named image was above chance for both cMP and vMP trials, for both monolinguals (cMP: $t[15] = 7.27$, $p < 0.001$, $[0.22, 0.40]$; vMP: $t[15] = 2.51$, $p = 0.024$, $[0.03, 0.33]$) and bilinguals (cMP: $t[14] = 7.89$, $p < 0.001$, $[0.24, 0.42]$; vMP: $t[14] = 4.65$, $p < 0.001$, $[0.15, 0.40]$).

Participants' RTs for correct responses to MP trials were analyzed in a mixed-effects linear model with the same fixed and random factors and random covariate as in the minimal pair accuracy analysis. As seen in **Figure 6**, while there was no main effect of MP type [$\chi^2(1, n = 284) = 1.35$, $p = 0.245$], or language background [$\chi^2(1, n = 284) = 1.19$, $p = 0.275$], the interaction of MP type and language background was significant [$\chi^2(1, n = 284) = 6.50$, $p = 0.011$]. LSD-corrected pairwise comparisons showed that while monolinguals' RT did not differ for cMP and vMP trials ($p = 0.323$, $[-161.62$ ms, 490.63 ms]),

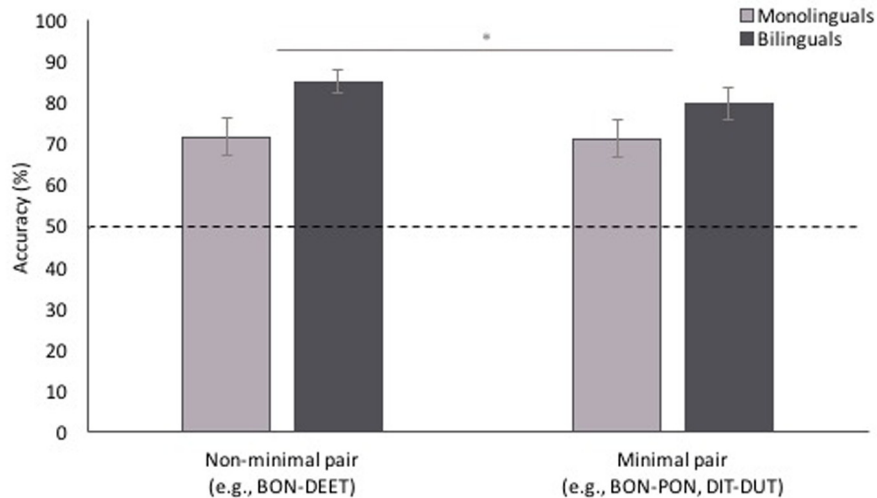


FIGURE 3 | Accuracy for non-minimal pair and minimal pair test trials. Participants were more accurate for non-minimal pair than minimal pair trials, and bilingual participants were more accurate overall than monolingual participants. Error bars represent one standard error. * $p < 0.05$.

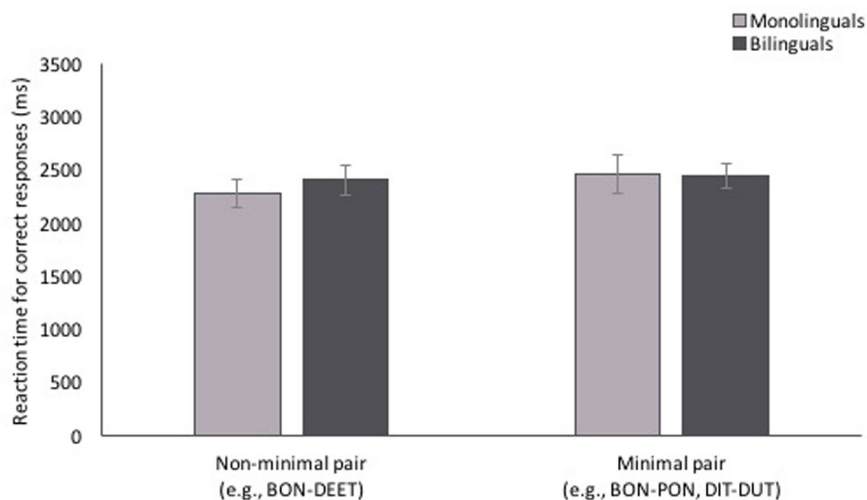


FIGURE 4 | Reaction time (RT) to non-minimal pair and minimal pair test trials. RTs did not differ between pair types or language background. Error bars represent one standard error.

bilinguals were slower to respond to vMP trials than cMP trials ($p = 0.009$, [108.90 ms, 771.14 ms]). Further, while RT for cMP trials did not differ between monolinguals and bilinguals ($p = 0.689$, [−358.51 ms, 542.19 ms]), bilinguals were slower to respond to vMP trials than monolinguals ($p = 0.022$, [75.39 ms, 949.98 ms]).

Finally, to determine whether participants' performance for MP trials differed depending on the feature difference between the MPs, we analyzed participants' correct and incorrect responses for cMPs and vMPs in two separate mixed-effects binary logistic models with contrast type (cMPs: place contrast, voicing contrast, place+voicing contrast; vMPs: height contrast, backness contrast, height+backness contrast) and language

background as fixed effects, and subject, order, target, distractor, and target location as random factors, and age included as a random covariate. As can be seen in **Figure 7**, performance for cMPs did not differ depending on the contrast type [$\chi^2(2, n = 183) = 1.14, p = 0.564$], or language background [$\chi^2(1, n = 183) = 0.13, p = 0.723$], and there was no interaction of contrast type and language background [$\chi^2(2, n = 183) = 3.57, p = 0.168$]. However, for vMP trials, performance differed depending on the vowel contrast [$\chi^2(2, n = 185) = 9.57, p = 0.008$]. LSD-corrected pairwise comparisons showed that participants were less accurate for vowel contrasts differing in height only than for contrasts differing in both height and backness ($p = 0.032$, [−0.38, −0.02]) or backness only

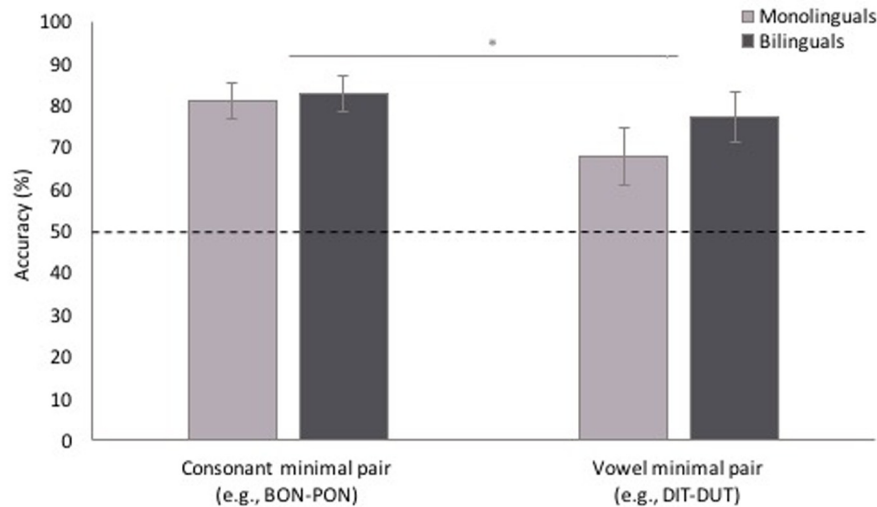


FIGURE 5 | Accuracy for consonant minimal pair and vowel minimal pair test trials. Participants were less accurate for vowel minimal pairs than consonant minimal pairs. Error bars represent one standard error. * $p < 0.05$.

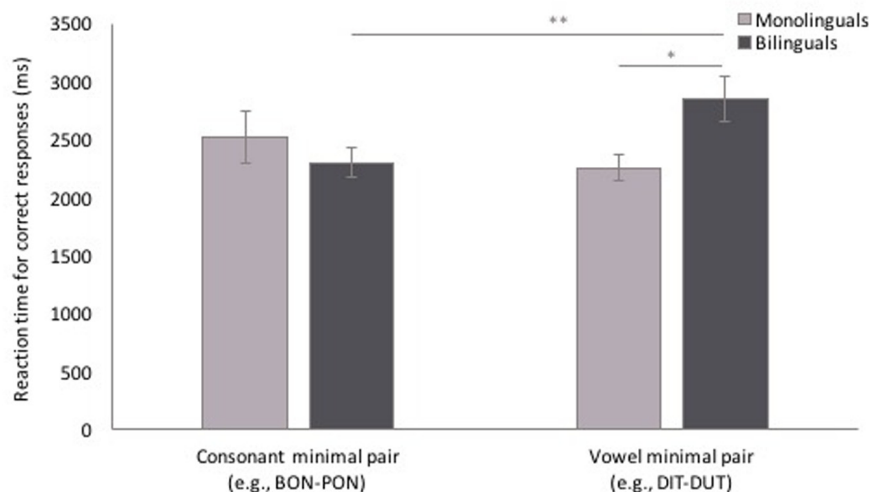
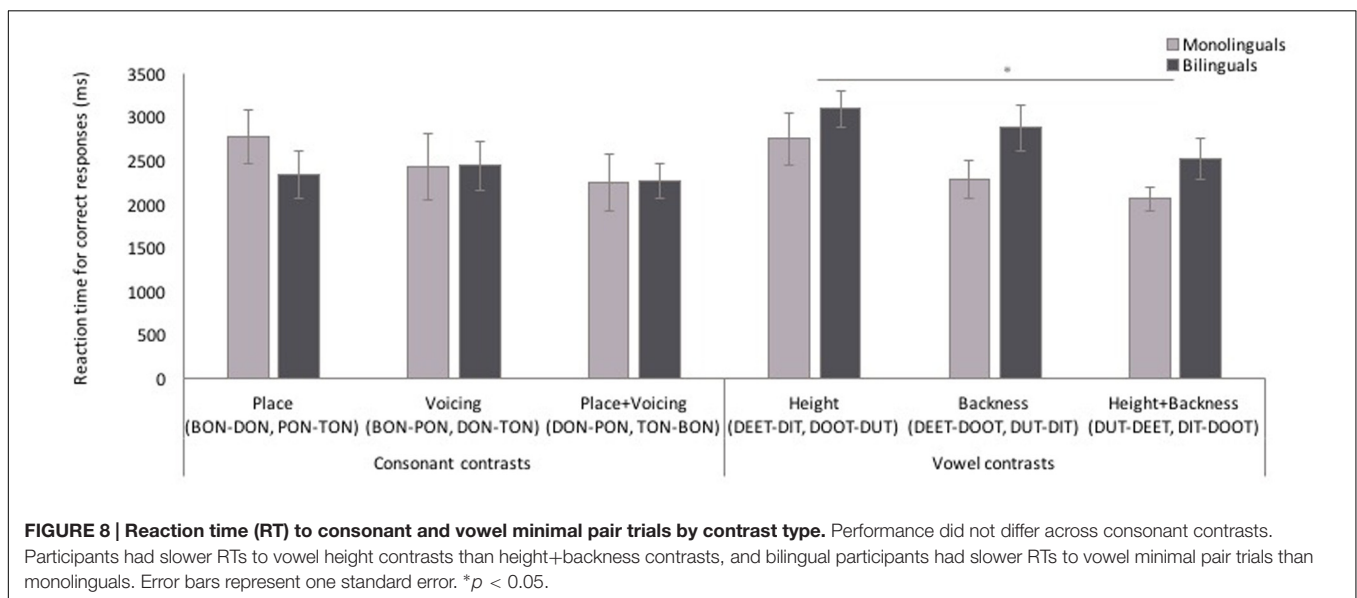
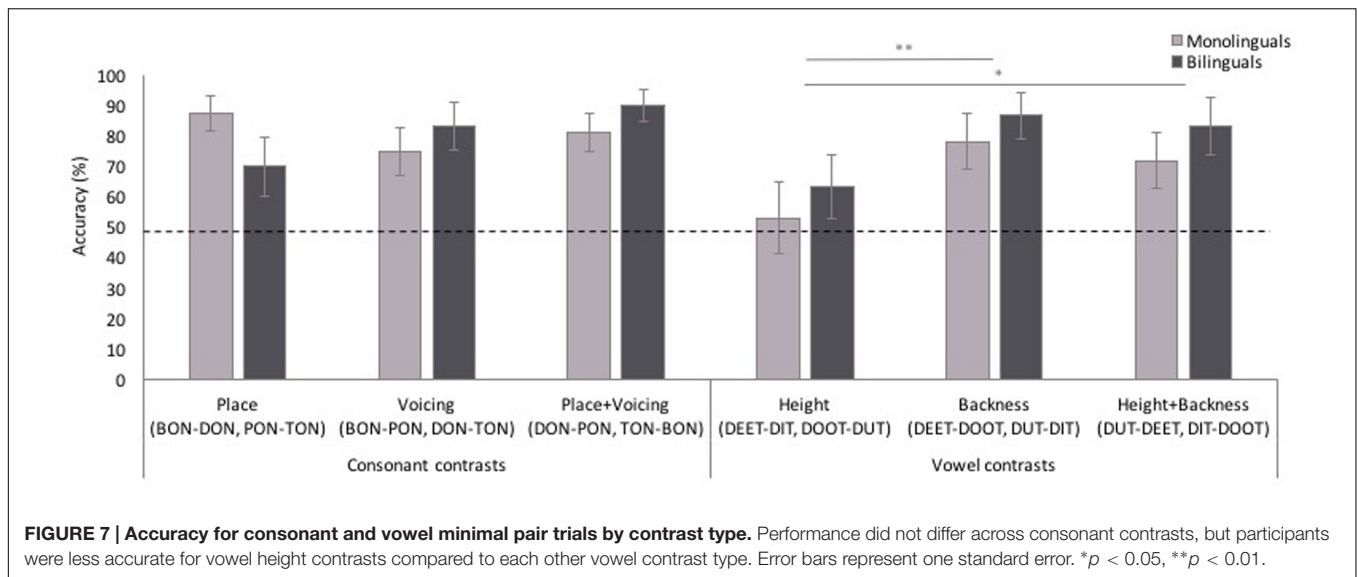


FIGURE 6 | Reaction time (RT) to consonant minimal pair and vowel minimal pair test trials. Bilinguals had slower RTs for vowel minimal pair trials than consonant minimal pair trials, and had slower RTs to vowel minimal pair trials than monolinguals. Error bars represent one standard error. * $p < 0.05$, ** $p < 0.01$.

($p = 0.004$, $[-0.41, -0.08]$). There were no effects of language background [$\chi^2(1, n = 185) = 1.34$, $p = 0.247$], nor was there an interaction of contrast type and language background [$\chi^2(2, n = 185) = 0.08$, $p = 0.959$]. One-sample t -tests against chance showed that both monolinguals and bilinguals demonstrated above chance performance for vowel contrasts differentiated by height+backness (monolinguals: $t[15] = 2.41$, $p = 0.029$, 95% CI $[0.25, 0.41]$; bilinguals: $t[14] = 3.57$, $p = 0.003$, $[0.13, 0.53]$) and backness only (monolinguals: $t[15] = 3.09$, $p = 0.007$, $[0.09, 0.48]$; bilinguals: $t[14] = 4.79$, $p < 0.001$, $[0.20, 0.53]$), but were both at chance for vowel contrasts differentiated by height only (monolinguals: $t[15] = 0.27$, $p = 0.791$, $[-0.22, 0.28]$; bilinguals: $t[14] = 1.29$, $p = 0.217$, $[-0.09, 0.35]$).

Participants' RTs for correct responses to cMP and vMP trials by contrast type (Figure 8) were analyzed in a mixed-effects linear model with the same fixed and random factors and random covariate as the accuracy analysis. RT did not differ based on the contrast type for the cMPs [$\chi^2(2, n = 150) = 2.40$, $p = 0.301$], and there was no effect of language background [$\chi^2(1, n = 150) = 0.11$, $p = 0.742$] or interaction with contrast type and language background [$\chi^2(2, n = 150) = 1.21$, $p = 0.547$]. However, for vMPs, RT did differ based on the contrast type [$\chi^2(2, n = 134) = 6.60$, $p = 0.037$], and language background of the participant [$\chi^2(1, n = 134) = 5.21$, $p = 0.022$], but there was no interaction between the two [$\chi^2(2, n = 185) = 0.193$, $p = 0.908$]. Overall, bilinguals were slower to respond to



vMPs than monolinguals, and participants were slower to respond to contrasts differing in height only than contrasts differing on both height and backness ($p = 0.013$, [123.60 ms, 1072.24 ms]).

DISCUSSION

In this study, we compared Australian English monolinguals and simultaneous Singaporean English–Mandarin bilinguals in their learning of phonologically overlapping novel words in an implicit, cross-situational paradigm, comparing vowel and consonant minimal pairs and non-minimal pairs produced in American English. Participants from both groups were significantly above chance in their recognition of new words across all pair types, consistent with successful learning within

this paradigm by adult listeners. While not unexpected, this result reinforces that this mechanism of word learning at least remains available to adult listeners, and also reassures as to the validity of the experimental paradigm.

Bilinguals outperformed monolinguals overall, which was consistent with our hypothesis, and also consistent with the interpretation that bilinguals have increased abilities in language-based tasks, hypothesized to be through enhanced phonological working memory (e.g., Kaushanskaya, 2012) and/or enhanced executive functioning skills (e.g., Bartolotti and Marian, 2012). In cross-situational word learning of phonologically overlapping words, enhanced phonological memory may allow for better implicit tracking of word-referent co-occurrence probabilities across trials, and augmented inhibitory control may allow for reduced activation of phonological neighbors in the lexicon.

Crucially, bilinguals' greater accuracy relative to monolinguals in our experiment is unlikely to be due to increased exposure to or familiarity with the American English target accent, as it is likely that there is more exposure to American English in Australia than in Singapore. For instance, American television programs account for 26.94% of free-to-air television broadcast hours in Australia, comprising 694 of 2576 weekly broadcast hours across 23 channels (Australian Tv guide for free-to-air television, 2016), whereas American programming accounts for only 10.08% of free-to-air broadcast hours in Singapore, comprising 79 of 784 h weekly across 7 MediaCorp TV channels (Toggle Tv Guide, 2016).

Although, we did not measure socioeconomic status (SES) of participants in this experiment, we believe that it is unlikely that there were differences between our participant groups that could account for the pattern of results found. A difference in SES between groups has been shown to lead to differences in performance in the Simon task (a measure of inhibitory control strength), such that children from higher SES families performed better on the task compared to children from lower SES families (Morton and Harper, 2007). The authors reasoned that because bilingual families are typically of lower SES than monolingual families, studies that have not controlled for SES may be conflating SES with bilingualism. While this is more likely to be the case in predominantly monolingual communities, both of our participant groups represented the predominant homogeneous language group of their community. As well, both of our participant groups comprised students at major urban public universities in developed countries with lifelong educational instruction in English. Tuition fees for undergraduate psychology degrees (the degree undertaken by the majority of participants) at each university are comparable¹, and both universities offer subsidized fees for nationals.

Moving to linguistic, rather than general cognitive aspects of this research, participants struggled more for minimal pair trials relative to non-minimal pairs, and in vowel minimal pair trials relative to consonant minimal pairs. Lower accuracy for vowel contrasts has previously been shown in native Australian English listeners' cross-situational learning of minimal pair words in Australian English (Escudero et al., 2016), and the present study extends this finding to other varieties of Modern English. There are several factors that may contribute to this. Firstly, while consonants tend to be perceived categorically (Liberman et al., 1967), vowels are perceived in a more continuous manner in many languages, including English (Fry et al., 1962; Stevens et al., 1969; Beddor and Strange, 1982; Polka, 1995). This may make it more difficult to perceive differences between vowel minimal pairs relative to word pairs that contain consonant differences. In English, vowels are also proposed to play less of a lexical role than consonants in speech perception, and instead play more of a role in conveying suprasegmental and syntactic information to the perceiver (Nespor et al., 2003). Supporting

this, research typically finds a perceptual bias toward consonants in tasks involving lexical access and processing (e.g., Cutler et al., 2000; Bonatti et al., 2005; Toro et al., 2008), including in explicit word learning by adults (Havy et al., 2014). Specifically, both monolinguals and bilinguals failed to discriminate vowel contrasts differing by height only, and also displayed slower RTs for these contrasts. Bilinguals were expected to have difficulty in discriminating the height contrast DEET and DIT due to the lack of the vowel in DIT in their phonological space. While it is not clear at this point whether a different factor accounted for monolinguals' failure, or whether failure by both groups was due to an unforeseen factor is at this point unclear. Interestingly, Escudero et al. (2014) found that Australian English-learning infants could not discriminate an Australian English vowel height contrast embedded in a minimal pair in an explicit word learning task, perhaps suggesting that vowel height may be a more difficult cue to perceive through the lifespan. Ongoing research in our laboratory comparing Australian English monolinguals, Singaporean English-Mandarin simultaneous bilinguals, and Mandarin-Australian English late sequential bilinguals in their cross-situational learning of minimal pairs produced in Australian English and American English will further address this question.

Although, bilinguals did not demonstrate an overall difference in accuracy for minimal pair types compared to monolinguals, they were slower to respond to vowel minimal pair trials compared to monolinguals. As mentioned above, this may have been due to difficulty in perceiving the vowel /i/ in DIT, and in particular, discriminating it from DEET (/i/). As DEET and DIT were involved separately in all vowel contrast types, this may have led to bilinguals' overall slower RTs for vowel minimal pair trials relative to monolinguals. Alternatively, this difference in performance may be due to differences between English and Mandarin. While experiments in English typically find a consonant bias, there is evidence that in Mandarin, vowels may contribute more to lexical identity than consonants. For instance, Chen et al. (2015) found that native Mandarin listeners showed better identification for Mandarin words made up of a consonant and vowel (CV words) when the consonant was replaced with noise (V-only words) than when the vowel was replaced with noise (C-only words). They also found that adding a proportion of the vowel aided identification of C-only words, while adding a proportion of the consonant to V-only words did not aid their identification, perhaps due to the fact that tone information is coupled with vowel information. It is therefore possible that apart from the consonant minimal pairs, in which every vowel was the same, for vowel minimal pairs, bilingual participants may have waited longer to respond in order to process the vowel information that is more lexically important to them compared to the monolinguals. Notably, this interpretation of the finding implies cross-talk between both of the bilinguals' languages. Future work could test this interpretation by comparing bilinguals' performance here with bilinguals whose languages both demonstrate a consonant bias rather than a vowel bias in word identification, such as English and French. Another possibility is that the same bias for

¹For instance, many participants were undertaking a Bachelor of Arts degree. The annual cost of this degree at the National University of Singapore for 2016/2017 was \$29,350 SGD (Fees for Undergraduate Programmes, 2016), and was \$22,000 AUD for the 2016 academic year at Western Sydney University (Fees and Costs, 2016).

consonants over vowels does not exist in native speakers of Singaporean English, which may be due to the local influence of Mandarin.

CONCLUSION

Our study is the first to demonstrate that bilinguals can also learn words via cross-situational statistical learning, and can do so while encoding fine phonological detail. Both Singaporean English–Mandarin simultaneous bilinguals and Australian English monolinguals learned phonologically overlapping word–object pairings sufficiently as to identify visual referents corresponding to words spoken in American English in the context of minimal, as well as non-minimal pairs. Thus, the finding also generally replicates Escudero et al. (2016), who found that a separate set of Australian English speakers than those tested here could learn minimal pair words produced in their accent. More importantly, although research on explicit word learning has often found a bilingual advantage relative to monolinguals (e.g., Kaushanskaya and Marian, 2009; Bartolotti and Marian, 2012; Kaushanskaya, 2012), our findings demonstrate for the first time that bilinguals also outperform monolinguals in a cross-situational word learning task. Future research can now explore whether this bilingual advantage in word learning accuracy lies in general cognitive attributes such as increased verbal working memory or attention, or cultural factors (e.g., Yang et al., 2011). Alternatively, the advantage here may be specific to the linguistic background of the bilinguals (i.e., simultaneous English–Mandarin) and test language used (American English).

REFERENCES

- Adesope, O. O., Lavin, T., Thompson, T., and Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Rev. Educ. Res.* 80, 207–245. doi: 10.3102/0034654310368803
- Australian Tv guide for free-to-air television (2016). *Australian TV Guide for Free-to-Air Television*. Available at: <http://www.ebroadcast.com.au/tv/static/SydneyNight.html> (accessed May 11, 2016)
- Bartolotti, J., and Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cogn. Sci.* 36, 1129–1147. doi: 10.1111/j.1551-6709.2012.01243.x
- Bartolotti, J., Marian, V., Schroeder, S. R., and Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Front. Psychol.* 2:324. doi: 10.3389/fpsyg.2011.00324
- Beddor, P. S., and Strange, W. (1982). Cross-language study of perception of the oral–nasal distinction. *J. Acoust. Soc. Am.* 71, 1551–1561. doi: 10.1121/1.387809
- Best, C. T. (1994). “Learning to perceive the sound pattern of english,” in *Advances in Infancy Research*, Vol. 9, eds C. Rovee-Collier and L. Lipsitt (Norwood, NJ: Ablex Publishing Corporation), 217–305.
- Best, C. T. (1995). “A direct realist perspective on cross-language speech perception,” in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, eds W. Strange and J. J. Jenkins (Timonium, MD: York Press), 171–204.
- Best, C. T., and Tyler, M. D. (2007). “Nonnative and second-language speech perception: commonalities and complementaries,” in *Language Experience in Second Language Speech Learning: in Honor of James Emil Flege*, eds O.-S. Bohn and M. J. Munro (Amsterdam: John Benjamins), 13–34.
- Bialystok, E., Craik, F., and Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 859–873. doi: 10.1037/0278-7393.34.4.859

Ongoing work in our lab will begin to address this latter issue by measuring performance in this task by late sequential Mandarin–English bilinguals and by bilinguals with different linguistic backgrounds, as well as in the same task but using a different English accent as the stimulus.

AUTHOR CONTRIBUTIONS

PE conceived the project and discussed bilingual testing with L.S. KM programmed and set up the experiment in Sydney and sent instructions as well as checked for comparability of set up and results with CF, who implemented the set up in Singapore. CF collected data in Singapore. KM completed analyses. All authors wrote the paper.

FUNDING

This research was funded by the ARC (Australian Research Council) Centre of Excellence for the Dynamics of Language (project number CE140100041) where PE is Chief Investigator and KM is Postdoctoral Fellow.

ACKNOWLEDGMENT

The authors would like to thank Hana Zjakic, Christina Quattropani, Nicole Traynor, and Valeria Peretokina for help with data collection in Sydney.

- Bialystok, E., and Craik, F. I. M. (2010). Cognitive and linguistic processing in the bilingual mind. *Curr. Dir. Psychol. Sci.* 19, 19–23. doi: 10.1177/0963721409358571
- Bialystok, E., Craik, F. I. M., and Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends Cogn. Sci.* 16, 240–250. doi: 10.1016/j.tics.2012.03.001
- Bialystok, E., Craik, F. I. M., and Ryan, J. (2006). Executive control in a modified antisaccade task: effects of aging and bilingualism. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 1341–1354. doi: 10.1037/0278-7393.32.6.1341
- Bialystok, E., and DePape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 565–574. doi: 10.1037/a0012735
- Bialystok, E., and Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition* 112, 494–500. doi: 10.1016/j.cognition.2009.06.014
- Blumenfeld, H. K., and Adams, A. M. (2014). Learning and processing of nonverbal symbolic information in bilinguals and monolinguals. *Front. Psychol.* 5:1147. doi: 10.3389/fpsyg.2014.01147
- Blumenfeld, H. K., and Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition* 118, 245–257. doi: 10.1016/j.cognition.2010.10.012
- Bonatti, L. L., Peña, M., Nespor, M., and Mehler, J. (2005). Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychol. Sci.* 16, 451–459. doi: 10.1111/j.0956-7976.2005.01556.x
- Carlson, S. M., and Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Dev. Sci.* 11, 282–298. doi: 10.1111/j.1467-7687.2008.00675.x
- Chen, F., Wong, M. L. Y., Zhu, S., and Wong, L. L. N. (2015). Relative contributions of vowels and consonants in recognizing isolated Mandarin words. *J. Phon.* 52, 26–34. doi: 10.1016/j.wocn.2015.04.003

- Costa, A., Hernández, M., Costa-Faidella, J., and Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: now you see it, now you don't. *Cognition* 113, 135–149. doi: 10.1016/j.cognition.2009.08.001
- Cox, F., and Palethorpe, S. (2007). An illustration of the IPA: Australian english. *J. Int. Phon. Assoc.* 37, 341–350. doi: 10.1017/S0025100307003192
- Curtin, S. A., Fennell, C., and Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Dev. Sci.* 12, 725–731. doi: 10.1111/j.1467-7687.2009.00814.x
- Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., and van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: cross-linguistic comparisons. *Mem. Cogn.* 28, 746–755. doi: 10.3758/BF03198409
- Dautriche, I., and Chemla, E. (2014). Cross-situational word learning in the right situations. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 892–903. doi: 10.1037/a0035657
- de Bruin, A., Treccani, B., and Sala, S. D. (2015). Cognitive advantage in bilingualism an example of publication bias? *Psychol. Sci.* 26, 99–107. doi: 10.1177/0956797614557866
- Deterding, D. (2007). *Singapore English*. Edinburgh: Edinburgh University Press.
- Duanmu, S. (2002). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.
- Duncan, H. D., Segalowitz, N., and Phillips, N. A. (2016). Differences in L1 linguistic attention control between monolinguals and bilinguals. *Bilingualism* 19, 106–121. doi: 10.1017/S136672891400025X
- Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16, 143–149. doi: 10.3758/BF03203267
- Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization*. Doctoral dissertation, Utrecht University, Utrecht.
- Escudero, P. (2009). “Linguistic perception of “similar” L2 sounds,” in *Phonology in Perception*, eds P. Boersma and S. Hamann (Berlin: Mouton de Gruyter), 151–190.
- Escudero, P., Best, C. T., Kitamura, C., and Mulak, K. E. (2014). Magnitude of phonetic distinction predicts success at early word learning in native and non-native accents. *Front. Psychol.* 5:1059. doi: 10.3389/fpsyg.2014.01059
- Escudero, P., Mulak, K. E., and Vlach, H. A. (2016). Cross-situational learning of minimal word pairs. *Cogn. Sci.* 40, 455–465. doi: 10.1111/cogs.12243
- Fees and Costs (2016). *Fees and Costs*. Available at: http://www.uws.edu.au/international/home/admissions/course_fees (accessed July 18, 2016).
- Fees for Undergraduate Programmes (2016). *Fees for Undergraduate Programmes*. Available at: <http://www.nus.edu.sg/registrar/education-at-nus/undergraduate-education/fees.html> (accessed July 18, 2016).
- Fennell, C. T., Byers-Heinlein, K., and Werker, J. F. (2007). Using speech sounds to guide word learning: the case of bilingual infants. *Child Dev.* 78, 1510–1525. doi: 10.1111/j.1467-8624.2007.01080.x
- Festman, J., Rodriguez-Fornells, A., and Münte, T. F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behav. Brain Funct.* 6:5. doi: 10.1186/1744-9081-6-5
- Fikkert, P. (2010). “Developing representations and the emergence of phonology: evidence from perception and production,” in *Laboratory Phonology 10: Variation, Phonetic Detail and Phonological Representation*, Vol. 10, eds C. Fougerson, B. Kühnert, and M. D'Imperio (Berlin: De Gruyter Mouton), 227–258.
- Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Lang. Speech* 5, 171–189.
- Havy, M., Serres, J., and Nazzi, T. (2014). A consonant/vowel asymmetry in word-form processing: evidence in childhood and in adulthood. *Lang. Speech* 57, 254–281. doi: 10.1177/0023830913507693
- Hilchey, M. D., and Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychon. Bull. Rev.* 18, 625–658. doi: 10.3758/s13423-011-0116-7
- Kaushanskaya, M. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: the role of phonological memory. *Bilingualism* 15, 470–489. doi: 10.1017/S1366728911000472
- Kaushanskaya, M., and Marian, V. (2009). The bilingual advantage in novel word learning. *Psychon. Bull. Rev.* 16, 705–710. doi: 10.3758/PBR.16.4.705
- Kaushanskaya, M., and Rehtzgel, K. (2012). Concreteness effects in bilingual and monolingual word learning. *Psychon. Bull. Rev.* 19, 935–941. doi: 10.3758/s13423-012-0271-5
- Kousaie, S., and Phillips, N. A. (2012a). Ageing and bilingualism: absence of a “bilingual advantage” in Stroop interference in a nonimmigrant sample. *Q. J. Exp. Psychol.* 65, 356–369. doi: 10.1080/17470218.2011.604788
- Kousaie, S., and Phillips, N. A. (2012b). Conflict monitoring and resolution: are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Res.* 1446, 71–90. doi: 10.1016/j.brainres.2012.01.052
- Kroll, J. F., and Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *J. Cogn. Psychol.* 25, 497–514. doi: 10.1080/20445911.2013.799170
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461. doi: 10.1037/h0020279
- Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear.* 19, 1–36. doi: 10.1097/00003446-199802000-00001
- Luk, G., De Sa, E., and Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism* 14, 588–595. doi: 10.1017/S1366728911000010
- Majerus, S., Poncelet, M., Van der Linden, M., and Weekes, B. S. (2008). Lexical learning in bilingual adults: the relative importance of short-term memory for serial order and phonological knowledge. *Cognition* 107, 395–419. doi: 10.1016/j.cognition.2007.10.003
- Mattock, K., Polka, L., Rvachew, S., and Krehm, M. (2010). The first steps in word learning are easier when the shoes fit: comparing monolingual and bilingual infants. *Dev. Sci.* 13, 229–243. doi: 10.1111/j.1467-7687.2009.00891.x
- Morton, J. B., and Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Dev. Sci.* 10, 719–726. doi: 10.1111/j.1467-7687.2007.00623.x
- Nespor, M., Peña, M., and Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue E Linguaggio* 2, 203–229.
- Norris, D., and McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychol. Rev.* 115, 357–395. doi: 10.1037/0033-295X.115.2.357
- Paap, K. R. (2014). The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *J. Cogn. Psychol.* 26, 242–255. doi: 10.1080/20445911.2014.891597
- Paap, K. R., and Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cogn. Psychol.* 66, 232–258. doi: 10.1016/j.cogpsych.2012.12.002
- Papagno, C., and Vallar, G. (1992). Phonological short-term memory and the learning of novel words: the effect of phonological similarity and item length. *Q. J. Exp. Psychol.* 44A, 47–67. doi: 10.1080/14640749208401283
- Pliatsikas, C., and Luk, G. (2016). Executive control in bilinguals: a concise review on fMRI studies. *Bilingualism* 19, 699–705. doi: 10.1017/S1366728916000249
- Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *J. Acoust. Soc. Am.* 97, 1286–1296. doi: 10.1121/1.1412170
- Prior, A., and Gollan, T. H. (2011). Good language-switchers are good task-switchers: evidence from Spanish-English and Mandarin-English bilinguals. *J. Int. Neuropsychol. Soc.* 17, 682–691. doi: 10.1017/S1355617711000580
- Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA: MIT Press.
- Service, E., Simola, M., Metsänheimo, O., and Maury, S. (2002). Bilingual working memory span is affected by language skill. *Eur. J. Cogn. Psychol.* 14, 383–408. doi: 10.1080/09541440143000140
- Singh, L., Poh, F. L. S., and Fu, C. S. L. (2016). Limits on monolingualism? A comparison of monolingual and bilingual infants' abilities to integrate lexical tone in novel word learning. *Front. Psychol.* 7:667. doi: 10.3389/fpsyg.2016.00667
- Smith, K., Smith, A. D. M., and Blythe, R. A. (2011). Cross-situational learning: an experimental study of word-learning mechanisms. *Cogn. Sci.* 35, 480–498. doi: 10.1111/j.1551-6709.2010.01158.x

- Smith, L. B., and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106, 1558–1568. doi: 10.1016/j.cognition.2007.06.010
- Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., and Öhman, S. (1969). Cross-language study of vowel perception. *Lang. Speech* 12, 1–23.
- Suanda, S. H., and Namy, L. L. (2012). Detailed behavioral analysis as a window into cross-situational word learning. *Cogn. Sci.* 36, 545–559. doi: 10.1111/j.1551-6709.2011.01218.x
- Toggle Tv Guide (2016). *Toggle TV Guide*. Available at: <http://www.toggle.sg/en/toggle/channelguide> (accessed May 11, 2016).
- Toro, J. M., Nespor, M., Mehler, J., and Bonatti, L. L. (2008). Finding words and rules in speech stream: functional differences between vowels and consonants. *Psychol. Sci.* 19, 137–144. doi: 10.1111/j.1467-9280.2008.02059.x
- van Leussen, J.-W., and Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Front. Psychol.* 6:1000. doi: 10.3389/fpsyg.2015.01000
- Verreyt, N., Woumans, E., Vandelandotte, D., Szmalec, A., and Duyck, W. (2016). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism* 19, 181–190. doi: 10.1017/S1366728914000352
- Vlach, H. A., and Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition* 127, 375–382. doi: 10.1016/j.cognition.2013.02.015
- Vlach, H. A., and Sandhofer, C. M. (2014). Retrieval dynamics and retention in cross-situational statistical word learning. *Cogn. Sci.* 38, 757–774. doi: 10.1111/cogs.12092
- Vouloumanos, A., and Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Dev. Psychol.* 45, 1611–1617. doi: 10.1037/a0016134
- Wee, L. (2004). "Singapore English: phonology," in *Africa, South and Southeast Asia*, Vol. 4, ed. R. Mesthrie (Berlin: Mouton de Gruyter), 259–277.
- Wee, L. (2010). "Eurasian Singapore English" in *The Lesser-Known Varieties of English: An Introduction*, eds D. Schreier, P. Trudgill, E. W. Schneider, and J. P. Williams (Cambridge: Cambridge University Press).
- Yang, S., Yang, H., and Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism* 14, 412–422. doi: 10.1017/S1366728910000611
- Yu, C., and Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychol. Sci.* 18, 414–420. doi: 10.1111/j.1467-9280.2007.01915.x
- Yurovsky, D., Yu, C., and Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cogn. Sci.* 37, 891–921. doi: 10.1111/cogs.12035

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Escudero, Mulak, Fu and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

