L2 PHONOLOGY MEETS L2 PRONUNCIATION

EDITED BY: John Archibald, Mary Grantham O'Brien and Andrew Sewell

PUBLISHED IN: Frontiers in Communication and Frontiers in Psychology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-061-1
DOI 10.3389/978-2-88974-061-1

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: frontiersin.org/about/contact

L2 PHONOLOGY MEETS L2 PRONUNCIATION

Topic Editors:
John Archibald, University of Victoria, Canada
Mary Grantham O'Brien, University of Calgary, Canada
Andrew Sewell, Lingnan University, Hong Kong, SAR China

Citation: Archibald, J., O'Brien, M. G., Sewell, A., eds. (2022). L2 Phonology Meets L2 Pronunciation. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-061-1

Table of Contents

frontiers in Communication

EDITORIAL
published: 25 November 2021
doi: 10.3389/fcomm.2021.804062

Editorial: L2 Phonology Meets L2 Pronunciation

John Archibald[1]*, *Mary Grantham O'Brien*[2] *and Andrew Sewell*[3]

[1] *Department of Linguistics, University of Victoria, Victoria, BC, Canada,* [2] *School of Languages, Linguistics, Literatures, and Cultures, University of Calgary, Calgary, AB, Canada,* [3] *Department of English, Lingnan University, Hong Kong SAR, China*

Keywords: L2 phonology, L2 pronunciation, L2 speech, L2 perception and production, L2 pedagogy

Editorial on the Research Topic

L2 Phonology Meets L2 Pronunciation

The theme of this collection is "*L2 Phonology Meets L2 Pronunciation*." Such an interdisciplinary approach, of course, runs the risk of any gathering of friends at which you discover that your chess-club friends having nothing to say to your ultimate-frisbee teammates. We are sure, however, that these papers reveal this not to be the case here. In various guises as researchers and teachers, the three editors have tackled the question of what is easy and what is difficult in both learning and teaching (the sub-theme of the collection). We hope you find our introductory mini-reviews helpful in setting the stage. One is on pronunciation teaching (O'Brien), one on functional load (Sewell), and one on L2 phonology (Archibald). There are clear similarities in what L2 phonologists are interested in, and what L2 pronunciation teachers are interested in. All three of these themes are intertwined in the collection.

What we have assembled here are papers written by people who have been fascinated by these same questions. In these nine papers, there are some which focus on consonants (Cardoso et al.; Stefanich and Cabrelli; Zhang and Levis), some on vowels (Cebrian et al.; Munro), some on prosody (Ghosh and Levis; Liu and Reed), and some on teaching (Colantoni et al.; Kostromitina and Kang). We group them in this way to reflect the Commentaries by eminent scholars that appear after the papers. We thank Shea, Thomson, McGregor, and Sonsaat Hegelheimer for accepting our invitation to round out the Research Topic. Of course, many of the papers reveal that the boundary line between phonology and pronunciation is really quite blurry. Such is the reality of scholarly life.

Stefanich and Cabrelli look at the production of the Spanish alveopalatal nasal/ɲ/by L1 English speakers. They illustrate the complex developmental path of acquisition of this new sound.

Zhang and Levis look at a less-studied consonantal pattern: the effects of a merger of/n/and/l/in the Southwestern Mandarin dialect of Chinese. They demonstrate that this L1 property affects production in Standard Mandarin differently than it affects English production.

Cardoso et al. look at the effects of different types of instruction on the acquisition of consonantal sequences. They show that the group which received instruction on the most marked structure fares the best.

Munro presents a fascinating data set of Cantonese learners of English tense/lax vowels, and shows that there is a great deal of individual and lexical variation which makes it very challenging to talk of a monolithic notion of difficulty.

Cebrian et al. probe the relationship between perceived similarity judgments of English tense/lax vowels by Spanish/Catalan native speakers and their perception and production. They discover that perceived similarity is not always a good predictor of discrimination ability.

Colantoni et al. draw on the pronunciation literature, and set out and illustrate some design principles for enhancing L2 Spanish intelligibility in the classroom.

OPEN ACCESS

Edited and reviewed by:
Manuel Carreiras,
Basque Center on Cognition, Brain and Language, Spain

****Correspondence:***
John Archibald
johnarch@uvic.ca

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *28 October 2021*
Accepted: *09 November 2021*
Published: *25 November 2021*

Citation:
Archibald J, O'Brien MG and Sewell A (2021) Editorial: L2 Phonology Meets L2 Pronunciation. Front. Commun. 6:804062. doi: 10.3389/fcomm.2021.804062

Kostromitina and Kang document pronunciation development in ESL immersion learners. They demonstrate the differential effects that immersion can have on learners of differing levels of proficiency, and discuss the curricular implications.

Ghosh and Levis examine the effects on intelligibility that different types of stress errors can have. The propose a Gravity Hierarchy to classify error types and suggest that errors affecting vowel quality should be prioritized in training.

Liu and Reed probe the complexity of the L2 intonational system via both production and processing measures. They explore possible bridges between L2 phonology and L2 pronunciation teaching.

As has been said many times before, there is no such thing as the L2 learner or the L2 classroom; learners vary, and classrooms vary. But just as in the broader fields of linguistics (i.e., across languages) and language acquisition (i.e., across learners), we seek to account for both global uniformity and local variation, we see that this is an issue in our collection as well. Some of the papers reveal commonalities in either the learner population (Stefanich and Cabrelli; Zhang and Levis) or the learner behavior (Ghosh and Levis) or the learning environment (Cardoso et al.; Colantoni et al.; Kostromitina and Kang). Other papers highlight the local variation within a given population (Cebrian et al.; Liu and Reed; Munro). We may find different profiles when we look at different levels of proficiency within a language class, or different similarity judgements within an L1 group. This clearly is an area of interest for both the L2 phonologist and the L2 pronunciation teacher or curriculum designer.

This is a collection that probes some difficult questions but will provide no easy answers. Nonetheless, we feel its insights are numerous for theory, research methodology, and classroom practice.

So, take down this e-book and explore it like you'd explore parallel sessions in a conference, or streets from a workshop venue you visited years ago. You may come across some old friends, you will make some new and exciting discoveries, you might make a note to come back to certain things later, and you could find yourself at times nodding, at times questioning, this is as it should be. L2 phonology meet L2 pronunciation.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

We would like to thank all the authors, reviewers, and editors who worked so diligently to help us bring this collection together.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 February 2021
doi: 10.3389/fpsyg.2021.640354

The Effects of L1 English Constraints on the Acquisition of the L2 Spanish Alveopalatal Nasal

Sara Stefanich[1]* *and Jennifer Cabrelli*[2]

[1] *Department of Spanish and Portuguese, Northwestern University, Evanston, IL, United States,* [2] *Department of Hispanic and Italian Studies, University of Illinois at Chicago, Chicago, IL, United States*

This study examines whether L1 English/L2 Spanish learners at different proficiency levels acquire a novel L2 phoneme, the Spanish palatal nasal /ɲ/. While alveolar /n/ is part of the Spanish and English inventories, /ɲ/, which consists of a tautosyllabic palatal nasal+glide element, is not. This crosslinguistic disparity presents potential difficulty for L1 English speakers due to L1 segmental and phonotactic constraints; the closest English approximation is the heterosyllabic sequence /nj/ (e.g., "canyon" /kænjn/ ['k^{h}æn.jn], cf. Spanish *cañón* "canyon" /kaɲon/ [ka.'ɲon]). With these crosslinguistic differences in mind, we ask: (1a) Do L1 English learners of L2 Spanish produce acoustically distinct Spanish /n/ and /ɲ/ and (1b) Does the distinction of /n/ and /ɲ/ vary by proficiency? In the case that learners distinguish /n/ and /ɲ/, the second question investigates the acoustic quality of /ɲ/ to determine (2a) if learners' L2 representation patterns with that of an L1 Spanish representation or if learners rely on an L1 representation (here, English /nj/) and (2b) if the acoustic quality of L2 Spanish /ɲ/ varies as a function of proficiency. Beginner ($n = 9$) and advanced ($n = 8$) L1 English/L2 Spanish speakers and a comparison group of 10 L1 Spanish/L2 English speakers completed delayed repetition tasks in which disyllabic nonce words were produced in a carrier phrase. English critical items contained an intervocalic heterosyllabic /nj/ sequence (e.g., ['p^{h}an.jə]); Spanish critical items consisted of items with either intervocalic onset /ɲ/ (e.g., ['xa.ɲa]) or /n/ ['xa.na]. We measured duration and formant contours of the following vocalic portion as acoustic indices of the /n/∼/ɲ/ and /ɲ/ ∼/nj/ distinctions. Results show that, while L2 Spanish learners produce an acoustically distinct /n/ ∼ /ɲ/ contrast even at a low level of proficiency, the beginners produce an intermediate /ɲ/ that falls acoustically between their English /nj/ and the L1 Spanish /ɲ/ while the advanced learners' Spanish /ɲ/ and English /nj/ appear to be in the process of equivalence classification. We discuss these outcomes as they relate to the robustness of L1 phonological constraints in late L2 acquisition coupled with the role of perceptual cues, functional load, and questions of intelligibility.

Keywords: second language acquisition, phonology, phonetics, spanish, english, nasals

OPEN ACCESS

Edited by:
John A. Archibald,
University of Victoria, Canada

Reviewed by:
Antonio A. González-Poot,
Autonomous University of
Campeche, Mexico
Mary Zampini,
Le Moyne College, United States

***Correspondence:**
Sara Stefanich
sara.stefanich@northwestern.edu

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

Received: *11 December 2020*
Accepted: *21 January 2021*
Published: *15 February 2021*

Citation:
Stefanich S and Cabrelli J (2021) The Effects of L1 English Constraints on the Acquisition of the L2 Spanish Alveopalatal Nasal.
Front. Psychol. 12:640354.
doi: 10.3389/fpsyg.2021.640354

INTRODUCTION

A lasting question that has occupied a central role in the study of second language (L2) phonology across several decades asks which factors modulate the acquisition of L2 contrastive sounds that are not part of the first language (L1) grammar. A look at the collective body of research reveals that, while there is robust evidence that novel L2 sounds are acquirable (see e.g., Broselow and Kang, 2013, for a review), it is clear that not all sounds are equal when it comes to their acquirability.

A sound's degree of difficulty can depend on a number of variables, which span the existence or absence of a phonologically similar L1 sound, functional load of the L2 sound, markedness, articulatory complexity, and language-specific constraints (featural and suprasegmental alike), among other factors.

In the present study, we examine L1 American English speakers' acquisition of the alveopalatal nasal /ɲ/ in L2 Spanish. This sound is a challenge for L1 American English speakers for a number of reasons. First, this is a scenario in which the L2 sound does not exist in the L1. Second, it is the least frequent phoneme in the Spanish inventory (Melgar de González, 1976) and has low functional load in Spanish. Third, L1 segmental and phonotactic constraints complicate the L2 learning task: American English does not permit complex palatal segments and the closest approximation in the English inventory is the sequence /nj/ (e.g., "canyon" /kænjn/ ['k^{h}æn.jn], which is derived from Spanish *cañón* "canyon" /kaɲon/ [ka.'ɲon]), which is restricted to heterosyllabic position. With these crosslinguistic differences in mind, to converge on the L2 Spanish target, the learner's grammar must come to allow a single alveopalatal nasal segment. The question, then, that follows, is whether these constraints can be overcome in the L2. While there are no L2 Spanish /ɲ/ acoustic or perception data to inform our predictions for the current study[1], we can look to a body of work that has examined L1 English speakers' acquisition of L2 Russian palatalized consonants to inform predictions for L2 Spanish learners. Specifically, L2 Russian learners have been reported to persistently rely on L1 /Cj/ sequences in both perception and production (e.g., Diehm, 1998), which leads to the prediction that L2 Spanish learners will pattern similarly to L2 Russian speakers and fail to reliably produce an alveopalatal nasal segment. In this study, we report acoustic data from a delayed repetition task completed in English and Spanish by L1 English learners of L2 Spanish at beginner and advanced levels of proficiency and in Spanish by a baseline comparison group of L1 Spanish/L2 English speakers. While L2 Spanish learners produce an acoustically distinct /n/ $\sim$ /ɲ/ contrast even at a low level of proficiency, the beginners produce an intermediate /ɲ/ that falls acoustically between their English /nj/ and the L1 Spanish /ɲ/ while the advanced learners' Spanish /ɲ/ and English /nj/ appear to be in the process of equivalence classification. We discuss these outcomes as they relate to the robustness of L1 phonological constraints in late L2 acquisition coupled with the role of functional load and questions of intelligibility. In the remainder of this section, we outline the phonetic and phonological properties of the nasal segments and sequences in English and Spanish and the L2 learning task, followed by a brief overview of the L2 research that informs our research questions and predictions and the questions and predictions themselves.

Nasal Consonants in English and Spanish

Spanish and English each have three nasal phonemes that contrast by place of articulation; however, while the Spanish nasal inventory (/m/ /n/ /ɲ/) includes an alveopalatal nasal that contrasts with the other nasals in word-medial onset position[2], the English inventory (/m/ /n/ /ŋ/, the latter of which is limited to coda position) does not.

1.	/m/	*cama*	/'ka**m**a/	'bed'
	/n/	*cana*	/'ka**n**a/	'gray hair'
	/ɲ/	*caña*	/'ka**ɲ**a/	'cane'

According to Martínez Celdrán and Fernández Planas (2007), the alveopalatal /ɲ/ is comprised of an alveolar nasal segment and a "partial" glide element. This glide element is posited to be phonologically associated with the nasal segment (e.g., Colina, 2009; Bongiovanni, 2019). An alveopalatal onset is illicit in American English due to two phonological constraints. First, as noted, American English does not allow palatal consonants with complex (simultaneous or sequential) points of articulation; instead, consonantal palatalization is realized non-contrastively as a sequence of distinct consonantal and glide segments (e.g., "music" [mju:zik], cf. [m^{j}u:zik]) (e.g., Antonova, 1988, cited in Diehm, 1998). As a result, /nj/ will be the closest American English approximation to Spanish /ɲ/. Second, American English[3] /nj/ cannot occupy onset position due to a ban on onset clusters that consist of a coronal segment and /j/ (see Kulikov, 2011). Rather, /nj/ is limited to a heterosyllabic context in which /n/ occupies a syllable coda and /j/ is phonologically associated with the following syllable onset (consider, for example, "canyon" /kænjn/ ['k^{h}æn.jn], which is derived from Spanish *cañón* /kaɲon/ [ka.'ɲon]). Together, these L1 constraints yield a learning task in which the grammar must come to allow a single nasal segment alveolar and palatal places of articulation in syllable onset position.

To determine whether L2 Spanish learners produce a single segment (/ɲ/) or a sequence (nj), we follow Bongiovanni (2019) by acoustically examining the vocalic portion that follows the nasal segment[4]. Specifically, we measure duration and first and second formant (F1 and F2) contours as correlates of the phonological association of a glide element. In her comparison of the production of /nj/[5] and /ɲ/ in Buenos Aires Spanish, Bongiovanni examined reported differences in gestural timing, specifically, sequential and quasi-simultaneous alveolar and palatal contact in /nj/ and /ɲ/, respectively (see e.g., Recasens and Romero, 1997). While some speakers evidenced neutralization of /nj/ and /ɲ/, the speakers who preserved the contrast exhibited

[1] To our knowledge, there is only a single study of L2 Spanish /ɲ/ (Díaz-Campos, 2004), which relies on impressionistic data. Although the author reports target-like production of the segment, it is not clear whether the target criteria differentiated between a single segment versus a two-segment sequence. That is, it is possible that the learners were producing heterosyllabic /nj/ rather than /ɲ/.

[2] While the sound appears word-initially, the limited inventory of lexical items is largely composed of loans from indigenous languages or onomatopoeia.

[3] We distinguish the variety of English spoken by the learners in the current study from varieties of English that permit /nj/ in onset position in words such as 'nuclear'.

[4] In line with Bongiovanni (2019), we avoid acoustic analysis of the nasal segment due to its reported unreliability (see, e.g., Fujimura, 1962, cited in Bongiovanni, 2019, p. 4).

[5] For ease of exposition, we follow Bongiovanni's (2019) phonemic notation of /nj/ and acknowledge that doing so conflates phonological and phonetic representations since the glide is not phonemic in Spanish.

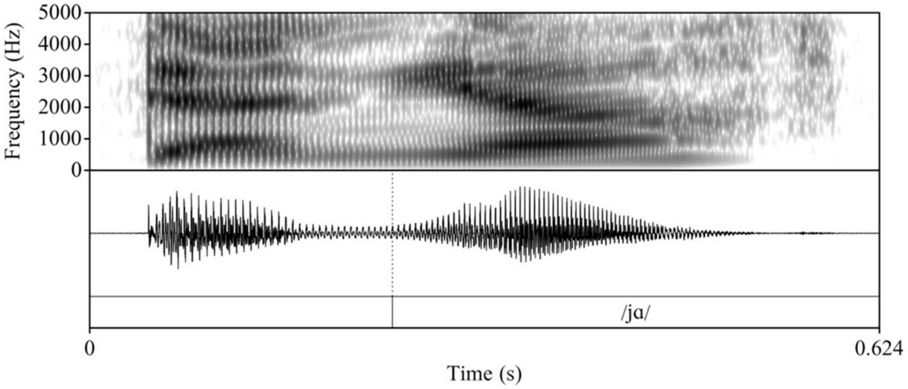


FIGURE 1 | Waveform and spectrogram of an advanced L2 Spanish participant's production of the English nonce item /dɛnja/ "denya."

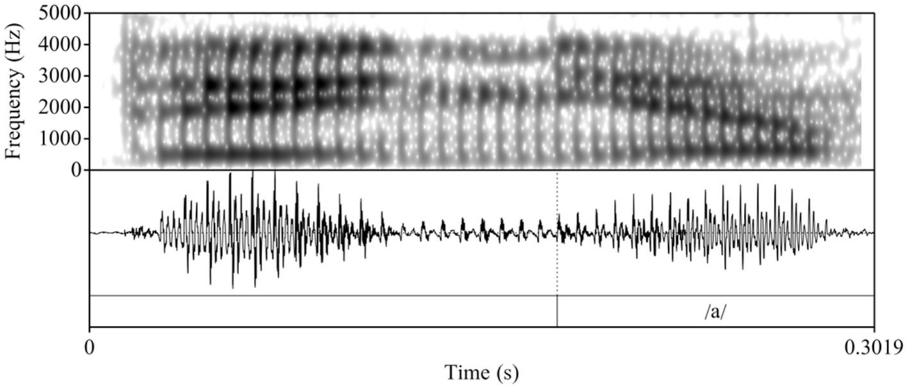


FIGURE 2 | Waveform and spectrogram of an L1 Spanish participant's production of the Spanish nonce item /deɲa/ "deña."

formant contour trajectories that differed in the first (F1) and second (F2) formants, with /nj/ showing a rise in F2 and lowering of F1 and a later F1 minimum and F2 maximum compared to /ɲ/. Duration of the vocalic portion in /nj/ was longer than in /ɲ/, given the glide's independent status. To our knowledge, there are no crosslinguistic comparisons of Spanish /ɲ/ and English /nj/[6]. Therefore, we rely on the tautosyllabic Spanish data to form the logical prediction that the distinction between heterosyllabic English /nj/ and Spanish /ɲ/ will be qualitatively similar to Spanish /nj/ vs. /ɲ/ and potentially more pronounced given the association of English /j/ with the onset of the following syllable. **Figures 1–3** illustrate the differences in formant trajectory and duration of the following vocalic portion by an advanced L2 Spanish participant of English /nj/ (**Figure 1**) compared with an L1 Spanish /ɲ/ (**Figure 2**) and L1 Spanish /n/ (**Figure 3**), the latter of which we include as a baseline for comparison with /nj/ and /ɲ/.

L2 Acquisition of Complex Palatal(ized) Consonants

As mentioned, although this is the first study to our knowledge to examine L1 English speakers' acquisition of the L2 Spanish palatal nasal, we can look to a small body of research that has examined L1 English speakers' acquisition of L2 Russian palatalized consonants to inform predictions for the current study. Similarly to Spanish /ɲ/, and unlike the English /nj/ sequence, Russian palatalized consonants are single complex segments with dual places of articulation that contrast phonemically with non-palatalized counterparts. Therefore, the L2 learning task for L1 English speakers is similar. In onset position[7], it seems as though L2 learners are able to perceive Russian /C^j/∼/C/ contrasts (Larson-Hall, 2004; Kulikov, 2011). However, it is not clear from these studies whether learners accurately perceive the distinction as /C^j/∼/C/ or whether the operation of L1 constraints instead persists in driving perception of the contrast

[6] As far as we are aware, the only comparison of English /Cj/ with a language that has a complex palatal segment comes from Diehm (1998); her examination of F2 trajectories found that L1 English /$C_{[labial]}$j/ formant transitions at consonantal release were longer and contained a shallower slope than L2 Russian /C^j/.

[7] Other studies such as Hacking (2011) and Hacking et al. (2016) have also examined L2 Russian palatalized consonants, but focus on coda position, where, as Hacking et al. (2016) notes, /Cj/ decomposition is not a possible compensatory strategy.

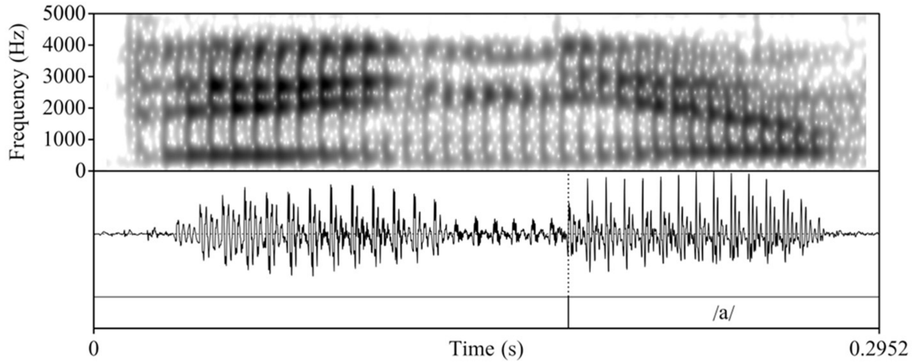


FIGURE 3 | Waveform and spectrogram of an L1 Spanish participant's production of the Spanish nonce item /dena/ "dena."

as /Cj/∼/C/. Diehm (1998) and Lukyanchenko and Gor (2011) examined this question as it relates to perception of the /C^j/ ∼ /C^{j}j/ contrast (akin to the Spanish /ɲ/ ∼ English /nj/ distinction), with Diehm also reporting production data of within-language and between-language contrasts.

Regarding perception of the /C^j/ ∼ /C^{j}j/ contrast, Lukyanchenko and Gor's data from an ABX task reflected above-chance accuracy in low-proficiency (∼73%)[8] and high-proficiency (80%) L2 Russian learners. In an identification task, Diehm found that high-experience learners perceived /C^j/ and /C^{j}j/ more accurately than low-experience learners when followed by a low vowel (23.5 and 10.2% more accurately, respectively) and that the increase in /C^j/ accuracy corresponded with a decrease in inaccurate identification of /C^j/ as /Cj/. While the learners in these two studies are neither at ceiling nor in line with L1 Russian accuracy, the data suggest that learners can develop perceptual acuity that at least partially circumvents the L1 phonological constraints that yield /Cj/ perception. Production data from Diehm, the only study to our knowledge to report a within-subjects comparison of L1 English and L2 Russian production data, points to partial acquisition in the oral modality. She found that even advanced L2 Russian learners' productions of /C^{j}V/ and /C^{j}jV/ syllables did not differ in F2 trajectory or duration from the point of consonantal release to the offset of the palatal element. This result lends support to the observation that L1 English learners decompose L2 /C^j/ into an L1-like /Cj/ sequence. Interestingly, a crosslinguistic comparison of a subgroup ($n = 4$) of advanced learners' L1 English and L2 Russian productions revealed that, while one learner produced L1 /Cj/ and L2 /C^j/ as (L1-like) /Cj/, the other three produced a distinct sequence in Russian that fell between their L1 English /Cj/ and the L1 Russian /C^j/ comparison. That is, although the L2 Russian /C^{j}V/ F2 trajectories did not approximate those of the L1 Russian comparison group, they were shorter and had a more negative slope (i.e., a larger degree of gestural overlap) than their English /CjV/. Diehm posited that this intermediate representation was indicative of partial L2 acquisition.

Taking the L2 Russian data as a point of departure, it seems that (at least partially) overcoming the relevant L1 English constraints is possible. However, we recognize that the learning scenario is different in L2 Russian vs. L2 Spanish. Specifically, the functional load (i.e., the importance in marking contrasts in a language) of the Russian /C^{j}V/ ∼ /CV/ contrast is higher than the functional load of the Spanish /ɲ/∼/n/ contrast. Russian, which has 42 consonantal phonemes, has 15 pairs of consonants that are phonemically distinguished by palatalization; palatalized consonant phonemes range in frequency ranking from 14 to 42 (Smirnova and Chistikov, 2011). Recall that Spanish /ɲ/, on the other hand, is the least frequent phoneme in the General Latin American Spanish inventory (Melgar de González, 1976), which contains 17 consonantal phonemes. Moreover, the phoneme is the only palatal consonant with dual articulation in Spanish and only contrasts in its palatalization with /n/. The phoneme thus has low functional load because it is infrequent but occurs in minimal pairs, with a low predictability of distribution. Functional load has been posited as a predictor of L2 phonological acquisition outcomes, whereby the probability of the acquisition of a contrast correlates with the functional load of that contrast (e.g., Best and Tyler, 2007). Relatedly, Archibald (Archibald, 2007, 2009) posits that, for learners to acquire a novel L2 contrast, there need to be sufficiently robust cues in the input to drive a revision to the representation that would allow for accurate perception of a single segment. Thus, it is wholly possible that L2 Spanish learners will not evidence the same success as L2 Russian learners. While the present study was not designed to explicitly test the effect of functional load and cue robustness, we will return to their potential role in the discussion.

Research Questions and Predictions

There are two research questions that drive the current study. The first regards whether learners acquire the relevant contrast in the L2, independent of how their /ɲ/ productions compare with the L1 Spanish /ɲ/.

(1a) Do L1 English learners of L2 Spanish produce acoustically distinct Spanish /n/ and /ɲ/, as measured by duration and formant trajectories of the following vocalic portion?

[8] Accuracy rates were presented graphically in a bar chart without numeric labels; exact percentages are not available.

(1b) Does the distinction of /n/ and /ɲ/ vary by proficiency?

Following the acoustic description of the nasal segments in section Nasal consonants in English and Spanish, distinct segments are predicted to take the form of a longer vocalic portion following /ɲ/ than following /n/; /ɲ/ is expected to present a higher F2 and a lower F1 than /n/, with an overall flatter shape for /n/ vs. /ɲ/.

In the case that learners distinguish /n/ and /ɲ/, the second question concerns the acoustic quality of /ɲ/.

(2a) If learners distinguish /n/ and /ɲ/, (i) do they rely on an L1 representation to produce /ɲ/ or (ii) have they overcome L1 constraints to establish a novel L2 representation? In this latter case, does the acoustic quality of the L2 representation pattern with that of an L1 Spanish representation?

(2b) Does the acoustic quality of the L2 Spanish /ɲ/ vary as a function of proficiency?

There are two logical outcomes. The first is that we will encounter evidence that learners have mapped Spanish /ɲ/ in the input onto their representation of English /nj/. In the Speech Learning Model (SLM, Flege, 1995, 2002; Flege and Bohn, 2020), this process of "equivalence classification" is predicted to eventually yield a representation (in SLM terms, a "phonetic category") that subsumes English /nj/ and Spanish /ɲ/ and has shifted to accommodate properties of both sounds. If one or both groups' /nj/ and /ɲ/ pattern together, we can compare their /nj/ and /ɲ/ productions with an L1 English baseline and an L1 Spanish baseline to determine where the learners might be in the equivalence classification process. The second possible outcome is that the learners have overcome the relevant L1 constraints and acquired a novel representation of Spanish /ɲ/ that is distinct from their L1 English /nj/. In this case, the difference is predicted to take the form of (a) shorter duration of Spanish /ɲ/ than English /nj/ and/or (b) a formant contour wherein /nj/ has a lower F1 valley and higher F2 peak than /ɲ/.

Based on Diehm's (1998) L2 Russian production data discussed in section L2 acquisition of complex palatal(ized) consonants, we can make tentative predictions with the caveat that the L2 Spanish developmental trajectory may diverge from that of the L2 Russian trajectory due to the status of the palatalized segments in Spanish vs. Russian. We predict that learners will distinguish /n/ from /ɲ/ in production even at beginner proficiency (RQ 1). They will approximate /ɲ/ in earlier stages of acquisition via L1-like /nj/; in later stages the duration and formant trajectory of /nj/ in L2 Spanish will shift toward the L2 target but will not fall within the acoustic parameters of the L1 comparison group's productions (RQ 2), resulting in an intermediate representation.

METHODS AND MATERIALS

Participants

Twenty-seven Spanish/English bilinguals participated in this study. Participants were all undergraduate or graduate students at a Midwest University at the time of testing ranging in age from 19 to 42 ($M = 25.80$, $SD = 5.24$). The Spanish/English bilinguals were divided into three groups based on order of acquisition and level of proficiency: (1) L2 Beginner ($n = 9$), 2) L2 Advanced ($n = 8$) and (3) L1 Spanish ($n = 10$). The L2 Beginner and L2 Advanced groups are comprised of L1 English speakers who learned Spanish as an L2. The L1 Spanish baseline comparison group mirror the L2 groups and are L1 Spanish speakers who learned English as an L2. The use of a Spanish baseline from a mirror-image bilingual group avoids the problematic comparison of bilinguals to monolinguals and acknowledges that a bilingual's systems do not act in isolation (see e.g., Grosjean, 2010). Further, as noted by an anonymous reviewer, the use of this baseline group is appropriate in the context of L2 learners in the United States, as it is often the case that learners' interactions are largely with bilingual Spanish speakers including, but not limited to, their instructors.

As measures of L2 language proficiency, L2 Spanish participants completed a 50-item multiple-choice test consisting of portions of the Diploma of Spanish as a Foreign Language (DELE) and Modern Language Association (MLA) that was first used in Slabakova and Montrul (Slabakova and Montrul, 2003) and has been widely used in L2 Spanish research; L1 Spanish participants completed a 50-item English proficiency cloze test adapted from the Oxford Placement Test. The L1 Spanish participants' English proficiency mirrors that of the L2 Advanced group's Spanish proficiency. Further, participants also completed the Bilingual Language Profile, BLP (Birdsong et al., 2012) as a proxy for language dominance. The BLP is a biolinguistic questionnaire that asks questions about bilinguals' language use, language acquisition, etc. and calculates a score for language dominance on a scale of−218 (Spanish dominant) to 218 (English dominant) with "0" indicating "balance" between the two languages. Further, as a part of the BLP, participants rate their Spanish and English proficiency with respect to reading, writing, speaking and understanding. **Table 1** illustrates the participant demographics by group.

Materials

The experiment consisted of Delayed Repetitions Tasks (e.g., Trofimovich and Baker, 2006) in English and Spanish. There were 40 trials (10 critical, 10 control, 20 distractor) in each task. A trial consisted of a target disyllabic nonce word with penultimate stress embedded within a carrier phrase, i.e., *Digo X para ti* in Spanish and its equivalent "I'm saying X to you" in English. A 1,000 ms pause followed the carrier phrase, after which participants were prompted to repeat the original sentence with the question *¿'Qué me dices?* In Spanish and its equivalent "What are you saying to me?" in English. All items were phonotactically licit in the target language: Critical and control items in Spanish consisted of (C)CV1.ɲV2 and (C)CV1.nV2 structures, respectively; critical and control items in English consisted of (C)CV1n.jV2 and (C)CV1.nV2 structures, respectively. Across conditions, V1 was a mid or low vowel (/ε/ or /ɲ/ in English; /e/ or /o/ in Spanish) and V2 was /a/ in Spanish and /ɲ/ in English. Distractors followed the same general (C)CV.CV structure as the control and critical stimuli. English and Spanish stimuli were recorded by phonetically trained female native speakers of Midwest American English and Northern Peninsular

TABLE 1 | Participant information.

	L2 Spanish Beginner		L2 Spanish Advanced		L1 Spanish	
	M	*SD*	*M*	*SD*	*M*	*SD*
Age of Spanish acquisition	14.00	3.20	14.50	4.21	Since birth	
Age of English acquisition	Since birth		Since birth		9.67	4.21
BLP dominance score	145.21	33.67	91.73	15.91	−75.99	25.59
Spanish proficiency score (out of 50)	20.33	5.09	45.38	2.92	n/a	
English proficiency score (out of 50)	n/a		n/a		43.50	2.87
Spanish self-rated proficiency						
Reading	2.44	1.33	5.50	0.53	6	0
Understanding	2.33	1.00	5.25	0.70	6	0
Speaking	1.56	0.88	5.13	0.83	6	0
Writing	1.89	1.36	5.25	0.46	6	0
English self-rated proficiency						
Reading	5.89	0.33	6	0	5.56	0.52
Understanding	5.89	0.33	6	0	5.33	0.50
Speaking	5.89	0.33	6	0	4.67	0.50
Writing	5.78	0.67	6	0	4.56	0.53

TABLE 2 | Spanish and English stimuli.

	n	English	Example		Spanish	Example	
Critical	10	(C)CVn.ja	/dɛnja/ [ˈdɛn.jə]	"denya"	(C)CV.ɲa	/deɲa/ [ˈde.ɲa]	*deña*
Control	10	(C)CV.na	/dɛna/ [ˈdɛ.nə]	"denna"	(C)CV.na	/dena/ [ˈde.na]	*dena*
Distractor	20	(C)CV.CV	/lɛka/ [ˈlɛ.kə]	"lecka"	(C)CV.CV	/meba/ [ˈme.β̞a]	*meba*

Spanish, respectively. **Table 2** illustrates the item composition for the English and Spanish tasks.

Trials were presented using E-prime (Psychology Tools, Inc.); audio stimuli were presented over Sennheiser HD-280 PRO headphones through a MOTU Ultralite mk3 interface. Recordings were made in a sound-attenuated booth using a head-mounted Shure SM 10A dynamic microphone and a Marantz PMD 661 solid-state recorder at a 44.1 kHz sampling rate.

Procedure

The experiment was conducted in a single session divided into Spanish mode and English mode segments, with the language mode order counterbalanced across participants. After providing informed consent, participants started the first segment with a 10-min interview in order to establish the first language mode, followed by the delayed repetition task in that language. The English mode segment ended with completion of the BLP and the Spanish mode segment ended with the Spanish written proficiency assessment. The L1 Spanish comparison group only participated in the Spanish mode segment and completed the interview, repetition task, and English written proficiency assessment, in that order.

Analysis

Duration Analysis

Acoustic Analysis

Following the research questions presented in section Research questions and predictions, this study examines the duration and formant trajectories of the following vocalic portion as acoustic indices to differentiate between nasal segments. To that end, we used Praat [6.1.16] (Boersma and Weenink, 2020) to segment and analyze the sound files. The theoretical ceiling of tokens was 710 or 30 per L2 learner (10 Spanish critical, 10 English critical, 10 Spanish control) + 20 per L1 control (10 Spanish critical, 10 Spanish control). Eighteen tokens were removed from data analysis due to non-target productions (participants repeating, skipping or producing different segments), creaky voice, or background noise for a final total of 692 tokens.

During segmentation, we used the following cues to determine the onset and offset of the vocalic portion: 1) the visual presence of an abrupt change in formant structure and frequencies (onset) and 2) a breaking up of the formant structure and a loss of energy and periodicity in the waveform (offset). Following Bongiovanni (Bongiovanni, 2019), boundaries between formant transitions or between the glide and the vowel /a/ were not marked. **Figures 4**, **5** illustrate two L2 productions of Spanish /ɲ/ and their segmentation: **Figure 4** aligns with the L1 /ɲ/ in **Figure 2**, in which /ɲ/ is represented by a steeper transition (i.e., slope) from the offset of the nasal into the following vowel /a/ (when compared with that of /nj/). **Figure 5**, on the other hand, aligns more closely with the L1 English /nj/ in **Figure 1**, in which the formant transition between the nasal segment and following vowel is marked by a raise in F2 frequency and a decrease in F1. After segmentation, we analyzed the sound files by using Praat scripts to extract the measurements (Hirst, 2012, for automatic duration measurements; McCloy and McGrath

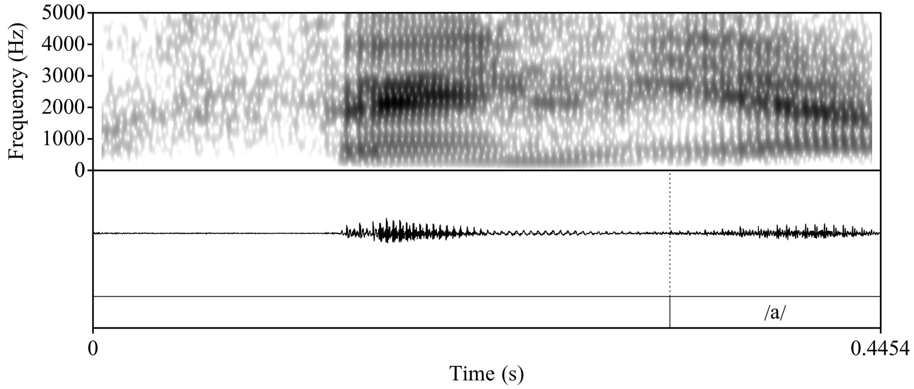


FIGURE 4 | Waveform and spectrogram of a L2 Spanish participant's target-like production of /ñ/ the Spanish nonce item /feña/ "feña."

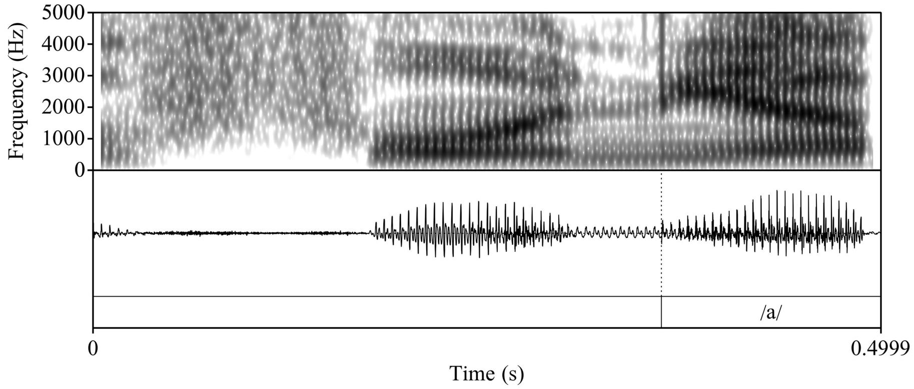


FIGURE 5 | Waveform and spectrogram of an L2 Spanish participant's L1 English-like production of /ñ/ in the Spanish nonce item /feña/ "feña."

(2012), for semi-automatic formant measurements). Formant measurements were taken at 20 points within the vocalic portion (i.e., every 5%).

Statistical Analysis

For duration of the vocalic portion, in order to normalize for potential between-participants differences in speech rate, we transformed raw duration to z-scores for each participant, with separate transformations for the L2 participants in English and Spanish. While English z-scores were transformed on /nj/ items and /n/ items, only /nj/ items were included in the analysis since the English /n/ data are not relevant to our research questions. In consideration of the sample size, rather than fitting the data to linear mixed-effects models, we follow Plonsky (Plonsky, 2015, p. 30) and instead rely on a combination of 95% confidence intervals (CIs) and effect sizes (here, Hedges' g, which corrects for bias from small sample size) to evaluate between-subjects and within-subjects differences. When using CIs for between-participants comparisons, a difference in means is significant when one group's mean does not fall within the comparison group's CI (Plonsky, 2015, p. 40). For within-groups comparisons, two means are considered significantly different if the CI of the mean of the two differences does not cross zero (Cumming and Finch, 2005). Small, medium, and large effect size thresholds are based on Plonsky and Oswald (Plonsky and Oswald, 2014), whereby between-participants thresholds are.40,0.60, and 1.00, and within-groups thresholds are 0.60, 1.00, and 1.40, respectively.

For the formant trajectories of the vocalic portion, we followed the analysis carried out in Bongiovanni (Bongiovanni, 2019). We transformed the formant values to Bark units and a Smoothing Spline ANOVA (SSANOVA) was fit to the data (time points and corresponding Bark units at each time point) in R, version 4.0.2 (R Core Team, 2020) with the gss package. As part of this analysis, a smoothing spline fits a smooth curve to the data and the SSANOVA determines whether the curves in question are statistically different from one another. Statistical significance is measured by non-overlapping confidence intervals around the splines. Following previous research (e.g., Simonet et al., 2008; Nance, 2014; Kirkham, 2017; Bongiovanni, 2019), we report only the graphical representations of the SSANOVA.

TABLE 3 | RQ 1: within-groups durational difference between /n/ and /ɲ/ for each group.

	Beg				Adv				L1			
	M	SD	CI[a]	g	M	SD	CI[a]	g	M	SD	CI[a]	g
/n/	−0.26	0.99	[−1.40, 0.37]	0.52	−0.46	0.87	[−1.73, −0.11][c]	1.02[b]	−0.64	0.72	[−1.98, −0.59][c]	1.67[b]
/ɲ/	0.25	0.89			0.46	0.86			0.65	0.75		

g, Hedges' g.

[a] *CI of the difference between the two means.*

[b] *Hedges' g >0.6.*

[c] *CI of the difference does not cross zero.*

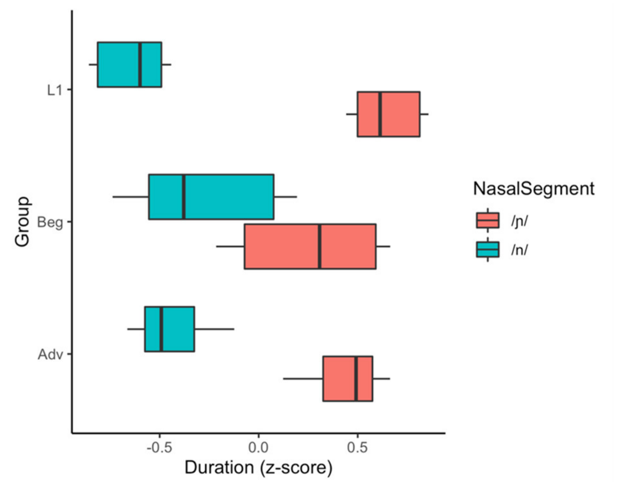


FIGURE 6 | Z-score transformed duration of Spanish /n/ and /ɲ/ produced by the beginner and advanced L2 groups and L1 Spanish group.

RESULTS

Research Question 1

Duration

To determine whether learners produce distinct /n/ and /ɲ/ as measured by duration and whether durational differences are moderated by proficiency, the analysis included within-subjects comparisons of the learners' English and Spanish productions as well as between-groups comparisons of the beginner vs. advanced learners. As shown in **Table 3** and **Figure 6**, the durational difference for /n/ and /ɲ/—whereby the vocalic portion following /ɲ/ would be predicted to be longer than that following /n/—was not significant for the beginner group and the effect size was negligible, with a large CI indicative of substantial variation within the group. However, the difference was significant for the advanced group with a medium effect size. All between-group comparisons were significant: The advanced group falls between the beginner group and L1 Spanish group, who make an even larger durational distinction between /n/ and /ɲ/.

Formant Trajectories

For the differences in formant structure, recall that with SSANOVA, statistical significance is indicated by non-overlapping confidence intervals plotted around the data-generated formant curves. With that in mind, **Figures 7**, **8** present the results of the SSANOVA for the beginner (**Figure 7**) and advanced (**Figure 8**) L2 groups' productions of English /nj/, Spanish /ɲ/, and Spanish /n/.

For the beginner L2 group, the formant trajectories are marked by non-overlapping confidence intervals. There is zero overlap comparing F1 in /n/ and /ɲ/ and no overlap between 0 and 85% of the F2 curve, with maximum differences of 0.94 and 2.41 Bark units, respectively. Similar results present for the advanced L2 group: There is zero overlap in the confidence intervals for both F1 and F2 when comparing /n/ and /ɲ/, with a maximum difference of 1.24 Bark units and 2.52 Bark units for F1 and F2, respectively.

To summarize the results for RQ 1: While the beginner group showed no significant difference between /n/ and /ɲ/ in terms of duration, there was a significant difference in the formant trajectories. The advanced group showed a significant difference between /n/ and /ɲ/ for both duration and formant trajectories.

Research Question 2

Duration

Because the beginner group does not produce durationally distinct /n/ and /ɲ/ and the advanced group does, we limit our duration analysis as it relates to RQ 2 to the advanced data. We examined whether the acoustic quality of the advanced group's duration of /ɲ/ reflects (i) perceptual mapping of /ɲ/ to English /nj/ or (ii) development of a novel L2 representation. We first compared the advanced learners' /ɲ/ to their English /nj/ and to the L1 English baseline (i.e., the beginner group's English /nj/. **Figure 9** provides a visual indication of the proximity of the advanced learners' /ɲ/ to their /nj/ as well as the proximity of the advanced /nj/ to the L1 English baseline (beginner) /nj/. The visual trends are supported by the data in **Table 4**; the advanced group did not make a significant durational distinction and the effect size did not reach the minimum threshold for a small effect. Moreover, a between-group comparison of the beginner and advanced /nj/ (**Table 5**) shows no difference.

Interestingly, **Figure 6** also shows the advanced learners' /ɲ/ trending with the L1 /ɲ/, and while the data in **Table 6** show that the advanced mean falls outside the L1 CI, the L1 CI falls on the edge of the advanced CI and the effect size does not approach the minimum threshold for a small effect. We also see in the same table that the L1 /ɲ/ is not different than the advanced /nj/.

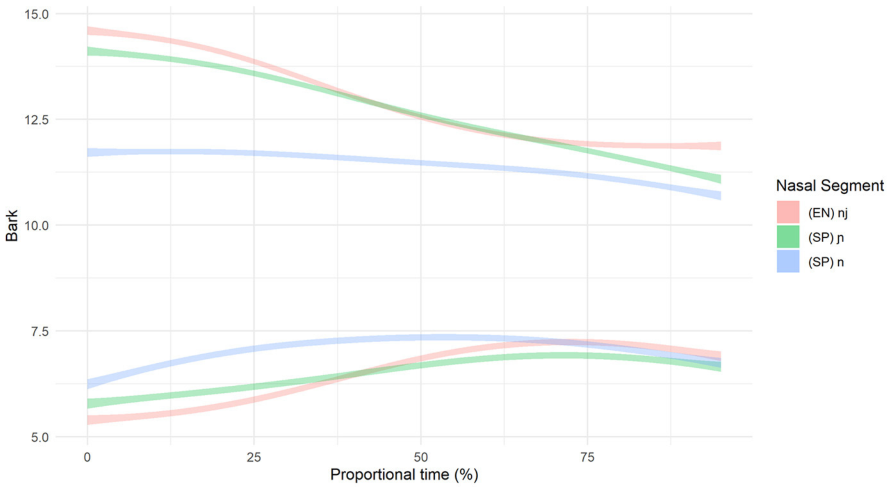


FIGURE 7 | Smoothing Spline ANOVA of formant trajectories for the L2 beginner group.

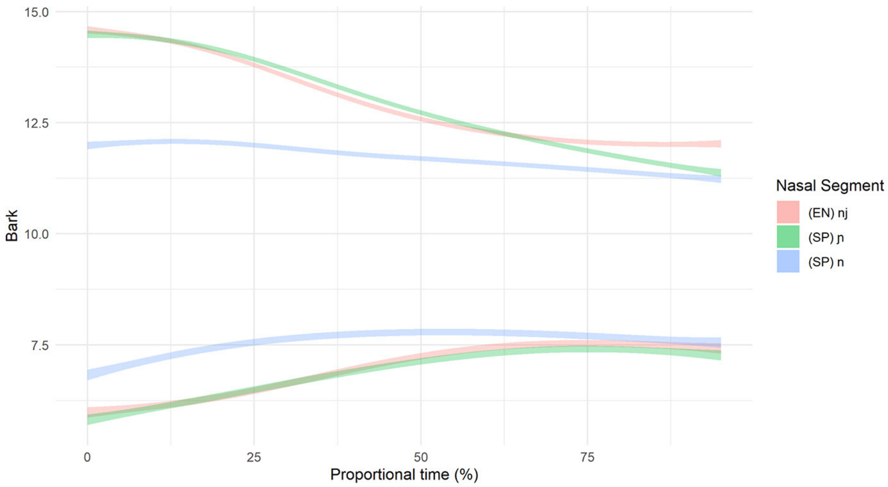


FIGURE 8 | Smoothing Spline ANOVA of formant trajectories for the L2 advanced group.

Formant Trajectories

To inform the nature of the L2 groups' production of Spanish /ɲ/, the formant structures were subject to a within-groups comparison (L2 Spanish /ɲ/ vs. L1 English /nj/) and a between-groups comparison (L2 Spanish /ɲ/ vs. L1 Spanish /ɲ/). Recall that, in addition to non-overlapping confidence intervals, differences between English /nj/ and Spanish /ɲ/ are expected to present in the form of a higher F2 peak and lower F1 valley for English /nj/ vs. Spanish /ɲ/. Keeping that in mind, **Figures 7**, **8** illustrate the within-groups comparisons for Beginner and Advanced L2 groups, respectively, and **Figure 10** illustrates the between-groups comparison.

For the Beginner L2 group, the within-group comparison (Beginner /ɲ/ vs. Beginner /nj/) revealed no overlap in confidence intervals for F1 (maximum difference =0.47 Bark), with the exception of the point where the two curves cross at 35–45%. The same pattern presents for F2, with no overlap in confidence intervals between 0–35 and 70–100% (maximum difference =0.51 Bark) and an overlap between 35 and 70% where the formant curves cross. The formant trajectories reflect the expected differences between /nj/ and /ɲ/, i.e., a lower F1 valley and a higher F2 peak for /nj/ vs. /ɲ/. The between-groups comparison (Beginner /ɲ/ vs. L1 Spanish /ɲ/) revealed no overlap in F1 confidence intervals between 0–55% and 85–100% (maximum difference =0.32 Bark), with an overlap between 55 and 85% where the curves cross. Further, there was no overlap in F2 confidence intervals between 0 and 65%, with a maximum difference of 0.51 Bark.

For the advanced L2 group, the within-groups comparison (Advanced /ɲ/ vs. Advanced /nj/) revealed that the F1 confidence intervals overlap between 0 and 50% and then run adjacent to one another from 50 to 100%. The F2 confidence intervals overlap at the beginning of the formant trajectory (0–20%) and then again when the curves cross around 65%. There is no overlap between 20 and 55% nor between 70 and 100% (maximum difference =0.21 Bark). Visually, the formant trajectories illustrate a higher F2 peak for English /nj/ vs. Spanish /ɲ/ but a lower F1 valley for Spanish /ɲ/ vs. English /nj/. However, the confidence intervals are overlapping at both of these points, thus rendering this distinction non-significant. The between-groups comparison (Advanced /ɲ/ vs. L1 Spanish /ɲ/) revealed zero overlap in the F1 confidence intervals with a difference of 0.64 Bark units at their most different. The F2 confidence intervals do not overlap at the beginning and the end of the trajectory (between 0 and 20% and between 75 and 100%) with a maximum difference of 0.22 Bark units). Visually, the Advanced /ɲ/ demonstrates a higher F2 peak (but not a lower F1 valley) when compared with the L1 control /ɲ/.

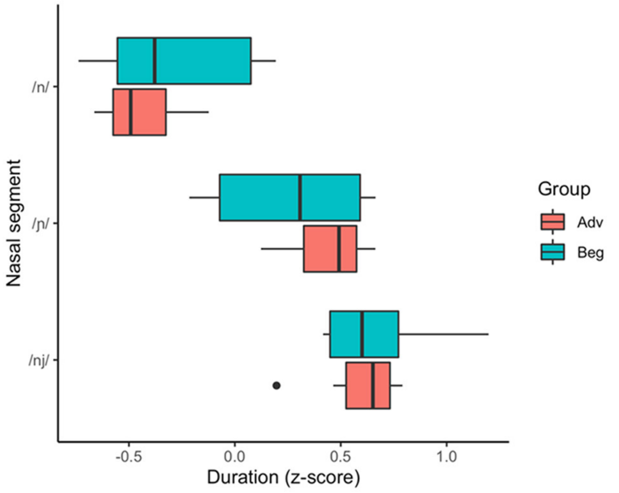


FIGURE 9 | Z-score transformed duration of Spanish /n/, Spanish /ɲ/, and English /nj/ produced by the beginner and advanced L2 groups.

TABLE 4 | RQ 2a: advanced learners' within-groups comparison of English /nj/ and /ɲ/.

	Adv			
	M	**SD**	**CI**[a]	**g**
/ɲ/	0.46	0.86	[-0.94,0.58]	0.21
/nj/	0.64	0.76		

g, Hedges' g.
[a]*CI of the difference between the two means.*

TABLE 5 | RQ 2b: beginner (L1 English baseline) and advanced learners' between-groups comparison of English /nj/.

	/nj/			
	M	**SD**	**CI**	**g**
Beg	0.61	0.86	[0.43, 0.79]	0.04
Adv	0.64	0.76	[0.47, 0.81]	

g, Hedges' g.

To summarize the results for RQ 2: As with RQ1, there were no differences in duration within learner groups or between learner groups and the L1 Spanish group. In terms of the beginners' formant trajectories, there was a significant three-way distinction between the beginners' Spanish /ɲ/, their English /nj/, and the L1 Spanish /ɲ/. In contrast, the advanced data's considerable overlap between /nj/ and /ɲ/ formant contours suggests a lack of difference. However, comparably limited overlap between the advanced /ɲ/ and L1 /ɲ/ contours indicate a significant difference in formant trajectories (see section structural equation modeling for details).

DISCUSSION

This study examined the speech production of beginner and advanced groups of L1 English/L2 Spanish learners and a comparison group of L1 Spanish/L2 English speakers to determine whether their production evidences a phonemic distinction between Spanish /n/ and /ɲ/, and, if so, whether learners establish the contrast via L1 English /nj/ or creation of a novel L2 representation. Between-segment patterns were established via two acoustic indices: Z-score transformed durations of the vocalic portion that follows the nasal segment and formant trajectories of the same vocalic portion. Durational differences were predicted to take the form of longer duration for /nj/ than /ɲ/ and for /ɲ/ than /n/. The formant trajectory for /nj/ was expected to consist of an F1 with a lower valley and an F2 with a higher peak compared with /ɲ/; /n/was predicted to evidence an earlier and higher F1 peak and an overall lower F2 contour.

TABLE 6 | Comparison of L1 group /ɲ/ with advanced group /ɲ/ and /nj/.

	Adv /ɲ/				Adv /nj/			
	M	**SD**	**CI**	**g**	**M**	**SD**	**CI**	**g**
Adv	0.46[b]	0.86	[0.27,0.65]	0.22	0.64	0.76	[0.47,0.81]	0.00
L1 /ɲ/	0.65	0.75	[0.50,0.80]		0.65	0.75	[0.50,0.80]	

g, Hedges' g.
[b]*Mean does not fall within comparison group's CI.*

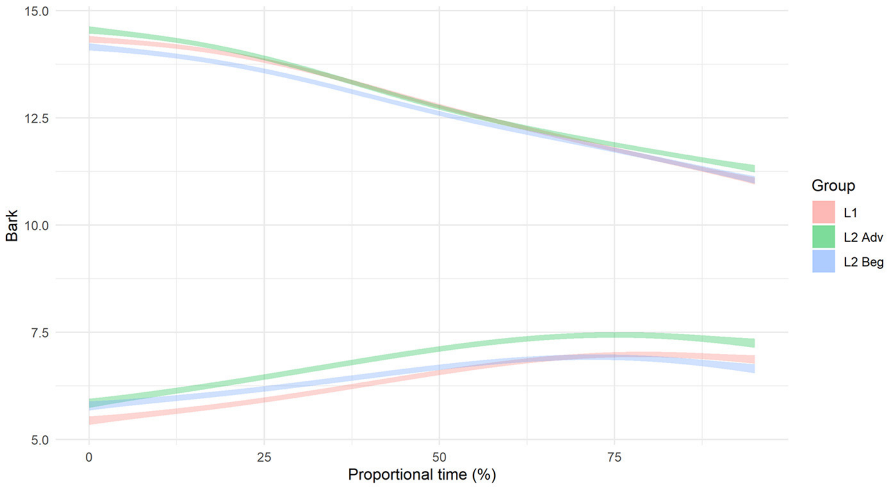


FIGURE 10 | Smoothing Spline ANOVA formant trajectories of Spanish /ɲ/ by group.

Before we turn to the discussion of the results, there are two notes regarding our analysis: First, because this is the first study to measure these sounds in L1 English/L2 Spanish bilinguals and there are no data to inform relative cue strength for these contrasts, we avoid the arbitrary assignment of relative weights to the indices of duration, vowel height (F1), and vowel frontedness (F2). Instead, we treat them as three separate strategies that speakers may use to distinguish between these nasal segments. Second, we eschew an arbitrary quantification of how little overlap in the confidence intervals of the spline curves constitutes a meaningful difference between the nasal segments and focus our qualitative interpretation on the first half of the formant trajectories (0–50%). In doing so, we home in on the nature of the transitions in /nj/ vs. /ɲ/ rather than differences in the vocalic portion, which is expected to differ due to cross-linguistic differences in the following vowel ([ə] in English vs. [a] in Spanish).

Research Question 1

Our first research question asked whether L1 English learners' L2 Spanish production reflects a distinction between /n/ and /ɲ/ in Spanish and if that distinction is subject to differences in proficiency. Beginning with the duration data, our beginner group did not evidence a significant difference in duration between /n/ and /ɲ/ but the advanced group did, producing longer vocalic portions following /ɲ/ than /n/ with a medium effect size. This difference between the learner groups suggests that the durational difference increases as a function of L2 proficiency. In comparison to the advanced group, however, the L1 Spanish group distinguishes via duration to a greater degree. Thus, while the advanced L2 Spanish and L1 Spanish both distinguish the segments via duration, the degree to which the L2 group does so does not approximate the L1 Spanish comparison.

If we were to use duration as the only acoustic index of the /n/∼/ɲ/ distinction, we would conclude that beginner learners have not acquired the /ɲ/ phoneme in Spanish since there is no durational difference between Spanish /n/ and /ɲ/. However, the formant trajectory data indicate that the beginners and advanced learners alike utilize height (F1) and frontedness (F2) of the vocalic portion to distinguish /n/ and /ɲ/. **Figures 7**, **8** illustrate majority non-overlap between the F1 and F2 formant contours for both beginners and advanced learners. In other words, in response to RQ1, the data indicate that yes, both groups of learners produce acoustically distinct Spanish /n/ and /ɲ/ segments. Further, the distinction varies by proficiency, with the beginners relying largely on F1 and F2 structure and the advanced learners utilizing both F1 and F2 structure and duration.

Given that the Spanish /n/∼/ɲ/ contrast varies by proficiency, the question that follows is: Why do beginner learners use vowel height and frontedness to make a distinction, but not duration? There are two points to consider. First, it could be the case that duration is not the primary cue that learners attend to in the input to distinguish the /n/∼/ɲ/ contrast, but rather that it is a later-acquired cue that learners have available to them at advanced proficiencies (see e.g., Kong and Lee, for discussion of the effects of proficiency on L2 cue-weighting strategies). Second, we remind the reader of the large standard deviation values for the duration results (**Table 3**), which indicate substantial variation within each proficiency group. For a more nuanced understanding of the relationship between proficiency and the use of duration, we plotted each learner's durational difference by their Spanish proficiency score (**Figure 11**).

While the group data indicated that the use of duration to differentiate /n/ and /ɲ/ was restricted to advanced proficiency, the individual data plotted by proficiency score indicate a range of durational difference across scores without a discernable pattern. That is, we do not see a clear relationship between an increase in proficiency score and an increase in duration difference to maximize the distance between /n/ and /ɲ/. This visualization is bolstered by a very weak positive correlation [$r_{(17)} = 0.08$, $p = 0.760$]. Thus, the individual data are suggestive of individual differences in cue weighting, which have been documented in L2

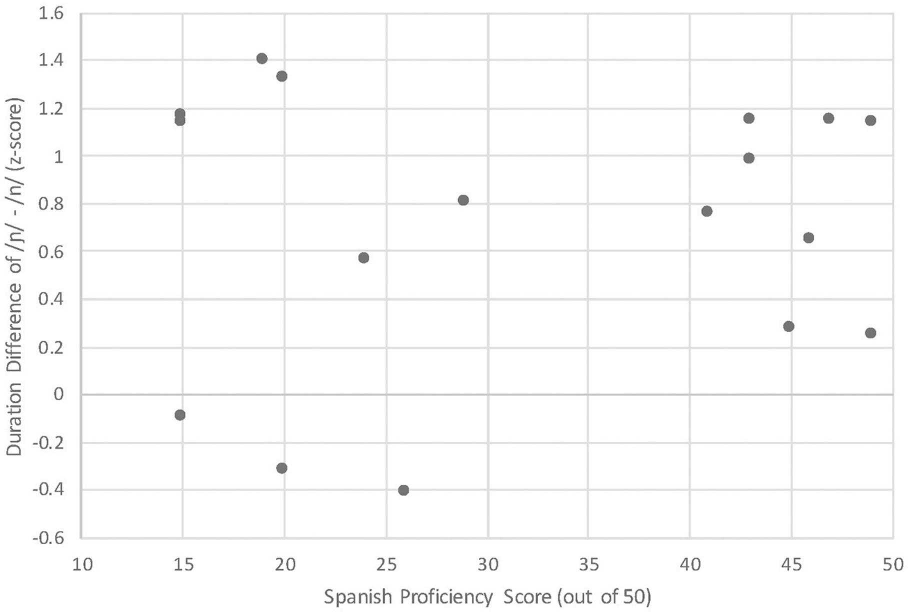


FIGURE 11 | Learners' duration difference in the following vocalic portion of /ɲ/ – /n/ by Spanish proficiency score.

acquisition (e.g., Chandrasekaran et al., 2010; Clayards, 2018), and specifically in the use of duration vs. formants in vowel discrimination (Kim et al., 2018). To confirm this hypothesis, we will need data from the perception of stimuli that isolate the acoustic indices and their possible combinations. Perception data from L1 and L2 Spanish speakers will inform the relative cue strength used by early vs. late learners of Spanish and longitudinal examination will inform whether L2 cue-weighting strategies change as a function of proficiency, as reported in Kong and Lee (2018). In addition, L1 Spanish perception of the L2 Spanish production data will be necessary to confirm that the quantitative differences in duration and formant contours are meaningful (in this case, perceivable).

Research Question 2

Our second research question concerned the quality of the learners' Spanish /ɲ/. That is, (a) do they rely on their L1 English /nj/ to approximate the novel Spanish contrast, or (b) have they overcome L1 constraints and established a single segment? We begin with the duration results, which are limited here to the advanced group since the beginner group did not use duration to contrast Spanish /n/ vs. /ɲ/. Since the learners' Spanish /ɲ/ was not different from their English /nj/, we posit that they do not use duration to differentiate them. Solely based on this outcome, we might conclude that the learners rely on English /nj/ to approximate the L2 Spanish /ɲ/ target. However, neither of these was different from the L1 Spanish /ɲ/. What might explain a scenario in which a learner's L1 and L2 sounds do not differ from each other and also do not differ from the L2 target? One possibility is that the learners' L2 Spanish /ɲ/ has affected their L1 English /nj/. L2 influence on the L1 aligns with a scenario of equivalence classification in which /ɲ/ is initially mapped onto /nj/ and, over time, the representation shifts in the direction of the L2 sound. Nevertheless, the advanced learners' English /nj/ did not differ from the English baseline (i.e., the beginner English /nj/ data). These inconclusive findings cast doubt on the reliability of duration as an acoustic correlate in this case, at least at the group level. A look at the individual-level duration difference between the advanced learners' English /nj/ and Spanish /ɲ/ (**Figure 12**, proficiency scores 40–50) supports the group data, with all but one advanced participant's differences clustered around zero. While there was a weak negative correlation between proficiency score and duration difference $[r_{(17)} = -0.30, p = 0.237]$, the weakness is likely due to the variation in the beginners' duration differences (proficiency scores < 30).

Turning to the formant data, we first compare the L2 learners' English /nj/ and Spanish /ɲ/, followed by the L1 Spanish and L2 Spanish /ɲ/. The beginners use vocalic quality to distinguish between English /nj/ and Spanish /ɲ/: They differentiate via vowel height (F1) and frontedness (F2) as illustrated by non-overlapping formant contours in the first half of the vocalic portion that follows the nasal segment (see **Figures 7**, **8**). The advanced learners, however, do not differentiate via F1, and the F2 contours overlap at the critical onset. Comparison of Spanish /ɲ/ across the three groups (Beginner, Advanced, L1 Spanish) shows a clear difference between the Beginner /ɲ/ and L1 Spanish /ɲ/ via F1 and F2, with no overlap in the critical regions. The Advanced /ɲ/ and L1 Spanish /ɲ/ comparison, however, is less straightforward. For F1, there is no overlap, although the shape of the formant contour is similar; for F2, there is no overlap at the onset.

Based on these comparisons of duration and formant contours, our tentative response to RQ 2a is that (i) the advanced

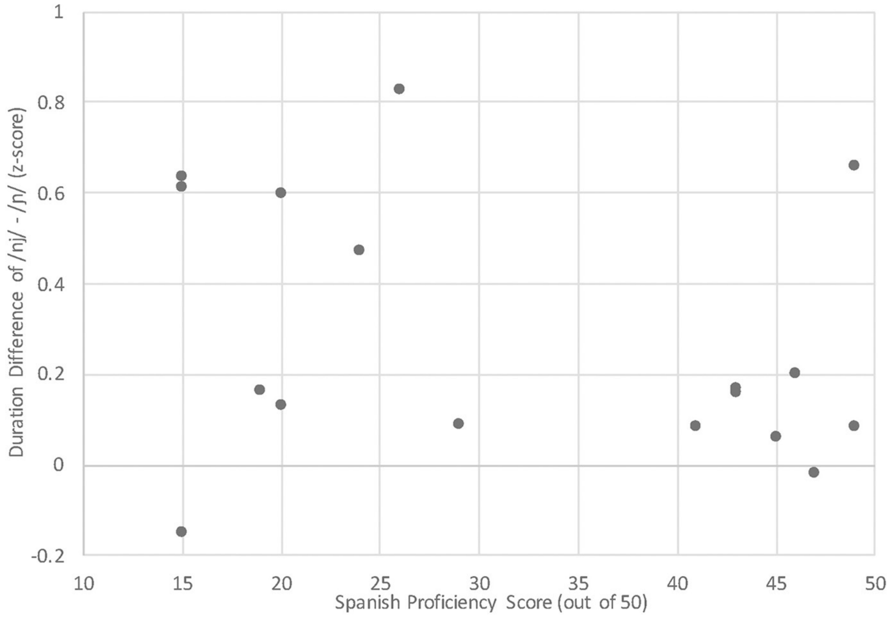


FIGURE 12 | FV Duration Difference [nj] - [ɲ] by Spanish proficiency score.

group relies on their L1 /nj/ representation when producing Spanish /ɲ/ while the beginner group does not, and (ii) neither group approximates the L2 Spanish target as measured by an L1 Spanish baseline[9]. In the case of the beginner group, they appear to have established an intermediate representation, although, as we note in our discussion of RQ 1, we will need perception data to determine whether the attested quantitative differences are perceivable. Considering that all of the maximum differences fell below the JND threshold of 1 Bark unit, these data are particularly warranted. Regarding RQ 2b, the acoustic realization of Spanish /ɲ/ varies as a function of proficiency: The beginners and advanced realizations differ from each other according to height and frontedness. Both groups differ from the L1 Spanish baseline along both parameters, although the advanced group approximates the L1 more closely than the beginner group.

The finding that beginner and advanced L2 learners differ from one another as well as from the L1 Spanish baseline is not unexpected; intermediate representations have been commonly documented in L2 production research (Zampini, 2008 for a review; see e.g., Broselow and Kang, 2013). In fact, recall that this is what Diehm (Diehm, 1998) found when comparing advanced L2 Russian learners' productions of palatalized consonants (section L2 acquisition of complex palatal(ized) consonants). The unexpected result, however, is that the advanced learners' productions (and not the beginners') show a persistent L1 effect. A common L2 developmental trajectory consists of initial pervasive L1 influence on the L2. Over time, these effects are thought to lessen as the L2 grammar develops, eventually yielding an L2 representation that (often partially) converges on the L2 target. This attested pattern has been formalized in models such as Major's (2001) Ontogeny Phylogeny Model (OPM), which explains the relationship between transfer, universals, and similarity. Of particular relevance to the present case is the OPM's Similarity Corollary, which posits that L1 transfer effects are persistent in later stages of development when the L1 and L2 phenomena are similar. These effects are thought to limit the role of universals, access to which is necessary to overcome L1 constraints, slowing down the L2 acquisition process. While the advanced learners produce the relevant L1 and L2 sounds similarly and thus align with this pattern, consideration of the beginner and advanced data in tandem suggest a case of U-shaped learning that can be likened to phonological regression attested in child phonological development (see e.g., Tessier, 2019). That is, it is possible that learners initially establish a novel (albeit intermediate) representation. Later, they recognize that an established L1 representation can be redeployed in the L2 (and potentially without compromised intelligibility, see discussion in section Results), which triggers the mechanism of equivalence classification. As noted in section L2 acquisition of complex palatal(ized) consonants, the shared representation is predicted to eventually shift to accommodate properties of both sounds. As we did with duration above, we can gauge the potential shift of /nj/ by comparing the Advanced /nj/ with the L1 English baseline (here, Beginner /nj/) (**Figure 13**).

What we find is that the F1 contour is indeed different, but in the opposite direction of what would be predicted (i.e., a higher – rather than lower–F1 valley), and there is no difference in F2. Thus, there is no evidence of a shift toward Spanish.

[9] Recalling that our L1 baseline are bilingual Spanish/English speakers, it is possible that a comparison of our baseline group to Spanish monolinguals could reveal differences in /ɲ/ production. However, we are limited in the current study to Spanish data from the baseline and future research will need to include their English data to examine the potential effect of L2 English /nj/ on L1 Spanish /ɲ/.

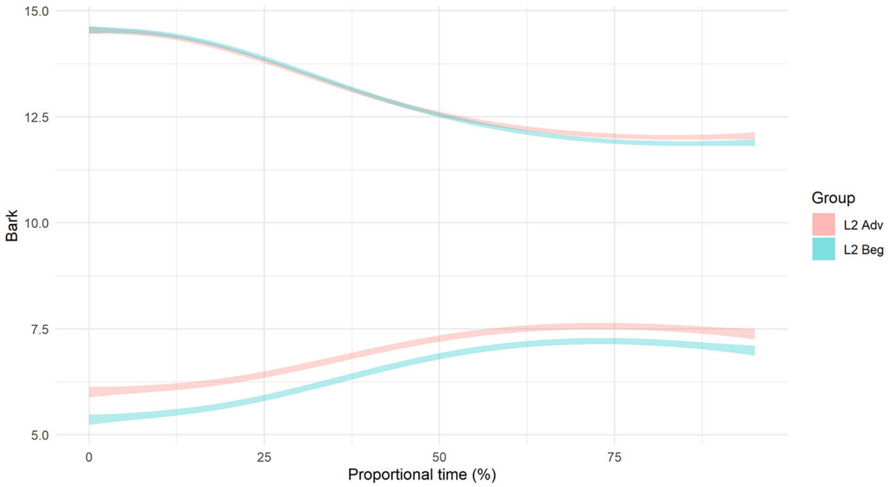


FIGURE 13 | Smoothing Spline ANOVA of English /nj/ formant trajectories.

Going forward, longitudinal data will best inform the following outstanding questions regarding the relationship between /ɲ/ and /nj/ via within-group developmental observations: (1) Do learners who first establish a three-way crosslinguistic distinction (/n/∼/ɲ/∼/nj/) eventually develop a single representation that is used for perception of English /nj/ and Spanish /ɲ/? (2) In the case of a shared representation, does the representation trend in the direction of the L1 or L2? And finally, as a complement to the variables we have examined here: (3) Is the representation that learners use to produce /ɲ/ tautosyllabic or heterosyllabic? That is, is there evidence that learners can overcome the L1 English constraint that militates against /nj/ in onset position? To address this question, future analyses will (a) examine production data for syllabic breaks via identification of glottalization (Scarpace and Mirza, 2014; González and Weissglass, 2017; Scarpace, 2017) and/or pauses (González and Weissglass) and measurement of preceding vowel duration (Scarpace and Mirza, 2014) and (b) test the perception of /nj/ in onset position vs. heterosyllabic position.

GENERAL DISCUSSION AND CONCLUSION

In this study, we have seen that learners are able to produce a novel L2 contrast even at a lower level of proficiency, but that the quality of the L2 representation does not approximate the L2 target. The logical question that follows then, is why not, particularly in the case of the advanced learners? In the remainder of the discussion, we consider two factors first introduced in section L2 acquisition of complex palatal(ized) consonants that have been shown to influence outcomes in L2 phonology—the functional load of a contrast (e.g., Best and Tyler, 2007) and the robustness of acoustic cues (e.g., Archibald, 2007, 2009)—and consider the practical implications of the attested outcomes.

First, recall Diehm's (1998) finding that highly proficient L1 English/L2 Russian speakers established a novel /C^j/ representation that differed from their English /Cj/ representation, thus outperforming the advanced L2 learners in our study. We could reasonably hypothesize that the difference between the L2 Russian and L2 Spanish learners can be attributed to the comparatively higher functional load of the contrast in Russian. However, the L2 Russian learners' new representation did not converge on the L2 target, which suggests that high functional load might be necessary but not sufficient for L2 convergence.

Instead, a lack of L2 convergence might be at least partially attributed to cue robustness. Archibald (2007, 2009) presented the hypothesis that acoustic cues must be sufficiently robust to trigger changes in the grammar that would yield accurate perception. The perceptual strength of the relevant cues in this case might be insufficient for an L2 learner; in fact, they might be insufficient even for L1 Spanish speakers: Bongiovanni (2015) reported that L1 Spanish speakers in Buenos Aires did not distinguish /ɲ/ and /nj/[10] in perception, even though a recent study shows that this same speaker population did so in production (Bongiovanni, 2019). If an L1 speaker cannot reliably perceive a difference, it is reasonable to predict that an L2 speaker cannot either, which could (at least partially) explain fossilization of a compromise or intermediate representation. That is, a shift in representation, while not "native-like," might never be triggered since intelligibility, or "the extent to which a listener understands a speaker's message" (Munro and Derwing, 2006), is not compromised. Another possibility, however, is that the learners have in fact established a target-like phonological representation that is simply not reflected in production. Following the approach of direct mapping from acoustics to phonology (DMAP, Darcy et al., 2012), a learner may establish an L2 phonological contrast prior to acquisition of a target phonetic representation. This is because the formation of lexical contrasts "do[es] not

[10] /ɲ/and /nj/ are contrastive in Spanish, although they form very few minimal pairs (e.g., *huraño* /ujaɲo/ 'unsociable' and *uranio* /uranjo/ 'uranium.'

require attunement to target-like category boundaries"; rather, construction of the relevant feature matrices "requires only the detection of acoustic correlates of phonological features in the raw percepts" (p. 16)[11]. Triangulation of our production data with perception data that reflect both categorization and discrimination will provide further insight into the learners' developmental trajectories.

Returning to the question of intelligibility, if an L2 speaker's message is not at risk of being lost, what are the practical implications of these findings? Pronunciation pedagogy objectives have shifted away from adherence to native-speaker norms and toward intelligibility (see e.g., Levis, 2018, for discussion). With this shift in mind, if intelligibility is not compromised, we posit that the limited instructional time that teachers have to dedicate to pronunciation does not need to be spent on this contrast. In fact, if /ɲ/ is addressed in Spanish pedagogical materials, it typically uses the English heterosyllabic /nj/ as a teaching tool, with statements such as "In speech, this letter [<ñ>] sounds like the middle sound in "canyon" and, in fact, the Spanish word for "canyon" is *cañon*" (Diversity Style Guide, 2020). This type of information could actually reinforce an intermediate /nj/ representation via conversion of explicit knowledge to implicit knowledge (see e.g., Ellis, 2015 for discussion of this relationship). As a complement the longitudinal observation of perception and production we have proposed, debriefing data on learners' experience with explicit instruction will help elucidate the effects of formal instruction on fossilization.

[11] See Baker (2004) for an overview of feature geometry analyses of Spanish palatals.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Illinois Office for Protection of Research Subjects. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SS and JC contributed equally toward conceptualization, methodology, and validation of the project, as well as preparation of the manuscript. SS was responsible for data collection and data analysis. JC was responsible for the theoretical framework. All authors contributed to the article and approved the submitted version.

REFERENCES

Antonova, D. N. (1988). *Fonetika i intonacija [Phonetics and intonation]*. Moscow: Russian Language.

Archibald, J. (2007). *The Role of Acoustic Prominence in L2 Phonology: Against the Deficit Model*. Paper presented at the Generative Approaches to Second Language Acquisition (GASLA 9), Iowa City, IA.

Archibald, J. (2009). Phonological feature re-assembly and the importance of phonetic cues. *Second Language Res.* 25, 231–233. doi: 10.1177/0267658308100284

Baker, G. K. (2004). *Palatal Phenomena in Spanish Phonology [Unpublished doctoral dissertation]*. University of Florida.

Best, C. T., and Tyler, M. (2007). "Nonnative and second-language speech perception," in *Language Experience in Second Language Speech Learning: In Honour of JAMES Emil Flege*, eds O.-S. Bohn and M. 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*. COERLL, University of Texas at Austin.

Boersma, P., and Weenink, D. (2020). *Praat: Doing Phonetics by Computer [Computer Program]*. Version 6.1.16. URL http://www.praat.org/

Bongiovanni, S. (2015). Neutralización del contraste entre /ɲ/ y /nj/ en el español de Buenos Aires: un estudio de percepción. *Signo y Seña* 11–46.

Bongiovanni, S. (2019). An acoustical analysis of the merger of /ɲ/and /nj/ in Buenos Aires Spanish. *J. Int. Phonetic Associat.* 1–25. doi: 10.1017/S0025100318000440

Broselow, E., and Kang, Y. (2013). Second language phonology and speech. In J. Herschensohn and M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 529-554). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139051729.031

Chandrasekaran, B., Sampath, P. D., and Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America,* 128, 456–465. doi: 10.1121/1.3445785

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *The Journal of the Acoustical Society of America*, *144*, EL172-EL177. doi: 10.1121/1.5052025

Colina, S. (2009). *Spanish phonology: A syllabic perspective*. Washington: Georgetown University Press.

Cumming, G., and Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist,* 60, 170. doi: 10.1037/0003-066X.60.2.170

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., Mcguire, M., and Scott, J. H. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English– L2 French acquisition. *Second Language Research,* 28, 5–40. doi: 10.1177/0267658311423455

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition,* 26, 249–273. doi: 10.1017/S0272263104262052

Diehm, E. E. (1998). *Gestures and Linguistic Function in Learning Russian: Production and Perception Studies of Russian Palatalized Consonants* (Unpublished doctoral dissertation). The Ohio State University, Columbus, Ohio, United States.

Diversity Style Guide (2020). *Eñe, ñ [Glossary entry]*. Available online at: https://www.diversitystyleguide.com/glossary/ene/

Ellis, N. C. (2015). "Implicit and explicit language learning: their dynamic interface and complexity," in *Implicit and Explicit Learning of Languages*, ed P. Rebuschat (Amsterdam: John Benjamins), 1–24.

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 (Baltimore: York Press), 233–277.

Flege, J. E. (2002). "Interactions between the native and second-language phonetic systems," in *An Integrated View of Language Development: Papers in Honor of Henning Wode*, eds P. Burmeister, T. Piske, and A. Rohde (Trier: Wissenschaftlicker Trier), 217–244.

Flege, J. E., and Bohn, O.-S. (2020). "The Revised Speech Learning Model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress*, ed R. Wayland (Cambridge: Cambridge University Press), 3–83.

Fujimura, O. (1962). Analysis of nasal consonants. *J. Acoust. Soc. Am.* 34, 1865–1875.

González, C., and Weissglass, C. (2017). *Hiatus Resolution in L1 and L2 Spanish.* Paper presented at the Romance Languages and Linguistic Theory 12: Selected papers from the 45th Linguistic Symposium on Romance Languages (LSRL), Campinas, Brazil.

Grosjean, J. (2010). *Bilingual. Life and Reality*. Cambridge, MA: Harvard University Press.

Hacking, J. (2011). "The production of palatalized and unpalatalized consonants in Russian by American learners," in *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2011*, eds M. Wrembel, M. Kul, K. Dziubalska-Kołaczyk (Frankfurt am Main: Peter Lang), 93–101.

Hacking, J. F., Smith, B. L., Nissen, S. L., and Allen, H. (2016). Russian palatalized and unpalatalized coda consonants: an electropalatographic and acoustic analysis of native speaker and l2 learner productions. *J. Phonet.* 54, 98–108. doi: 10.1016/j.wocn.2015.09.007

Hirst, D. (2012). *Analyse_tier.praat*. Retrieved from: http://uk.groups.yahoo.com/group/praat-users/files/Daniel_Hirst/analyse_tier.praat (accessed October 13, 2015).

Kim, D., Clayards, M., and Goad, H. (2018). A longitudinal study of individual differences in the acquisition of new vowel contrasts. *J. Phonet.* 67, 1–20. doi: 10.1016/j.wocn.2017.11.003

Kirkham, S. (2017). Ethnicity and phonetic variation in Sheffield English liquids. *J. Int. Phonetic Associat.* 47, 17–35. doi: 10.1017/S0025100316000268

Kong, E. J., and Lee, H. (2018). Attentional modulation and individual differences in explaining the changing role of fundamental frequency in Korean laryngeal stop perception. *Lang. Speech* 61, 384–408.

Kulikov,V. (2011). "Features, cues, and syllable structure in the acquisition of Russian palatalization by L2 American learners," in *Achievements and Perspctives in SLA of Speech: New Sounds*, eds M. Wrembel, M. Kul, and K. Dziubalska-Kołaczyk (Frankfurt am Main: Peter Lang Verlag), 193–204.

Larson-Hall, J. (2004). Predicting perceptual success with segments: a test of Japanese speakers of Russian. *Second Language Res.* 20, 33–76. doi: 10.1191/0267658304sr230oa

Levis, J. M. (2018). *Intelligibility, Oral Communication, and The Teaching Of Pronunciation*. Cambridge: Cambridge University Press.

Lukyanchenko, A., and Gor, K. (2011). "Perceptual correlates of phonological representations in heritage speakers and L2 learners," in *Proceedings of the 35th Annual Boston University Conference on Language Development,* eds N. Danis, K. Mesh, and H. Sung (Somerville, MA: Cascadilla Press), 414–426.

Major, R. C. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. New York, NY: Routledge. doi: 10.4324/9781410604293

Martínez Celdrán, E., and Fernández Planas, A. M. (2007). *Manual de fonética española. Articulaciones y sonidos del español*. Barcelona: Editorial Ariel.

McCloy, D., and McGrath, A. (2012). *SemiAutoFormantExtractor.Praat*. Retrieved from: https://github.com/drammock/praat-semiauto/blob/master/SemiAutoFormantExtractor.praat (accessed February 10, 2018).

Melgar de González, M. (1976). *Cómo detectar al niño con problemas del habla*. Mexico City: Trillas.

Munro, M. J., and Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: an exploratory study. *System* 34, 520–531. doi: 10.1016/j.system.2006.09.004

Nance, C. (2014). Phonetic variation in Scottish Gaelic laterals. *J. Phonet.* 47, 1–17. doi: 10.1016/j.wocn.2014.07.005

Plonsky, L. (2015). "Statistical power, *p* values, descriptive statistics, and effect sizes: A "back-to-basics" approach to advancing quantitative methods in L2 research," in *Advancing Quantitative Methods in Second Language Research*, ed L. Plonsky (New York, NY: Routledge), 23–45.

Plonsky, L., and Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learn.* 64, 878–912. doi: 10.1111/lang.12079

R Core Team (2020). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. Version 4.0.2. Available online at: http://www.R-project.org/

Recasens, D., and Romero, J. (1997). An EMMA study of segmental complexity in alveolopalatals and palatalized alveolars. *Phonetica* 54, 43–58. doi: 10.1159/000262209

Scarpace, D., and Mirza, M. (2014). "Learning to resyllabify across words: a training study," in *Paper Presented at Current Approaches to Spanish and Portuguese Second Language Phonology (CASPSLaP)* (Washington, DC: Georgetown University).

Scarpace, D. L. (2017). *The Acquisition of Resyllabification in Spanish by English Speakers.* Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign, Urbana, IL.

Simonet, M., Rohena-Madrazo, M., and Paz, M. (2008). *Preliminary evidence for incomplete neutralization of coda liquids in Puerto Rican Spanish.* Paper presented at the Selected proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology.

Slabakova, R., and Montrul, S. (2003). Genericity and aspect in L2 acquisition. *Language Acquisition* 11, 165–196. doi: 10.1207/s15327817la1103_2

Smirnova, N., and Chistikov, P. (2011). "Statistics of Russian monophones and diphones," in *Proceedings of Specom* 2011, 218–223.

Tessier, A. (2019). U-shaped development in error-driven child phonology. *Wires Cognitive Sci.* 10:e1505. doi: 10.1002/wcs.1505

Trofimovich, P., and Baker, W. (2006). Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *Stud. Second Language Acquisit.* 28, 1–30. doi: 10.1017/S0272263106060013

Zampini, M. (2008). "L2 speech production research: findings, issues and advances," in *Phonology and Second Language Acquisition*, eds J. G. Hansen Edwards and M. Zampini (Amsterdam: John Benjamins), 219–249. doi: 10.1075/sibil.36.11zam

Conflict of Interest: 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.

frontiers in Communication

MINI REVIEW
published: 16 February 2021
doi: 10.3389/fcomm.2020.626985

Ease and Difficulty in L2 Pronunciation Teaching: A Mini-Review

Mary Grantham O'Brien *

School of Languages, Linguistics, Literatures and Cultures, University of Calgary, Calgary, AB, Canada

Both L2 learners and their teachers are concerned about pronunciation. While an unspoken classroom goal is often native-accented speech (i.e., a spoken variety of the mother tongue that it not geographically confined to a place within a particular country), pronunciation researchers tend to agree that comprehensible speech (i.e., speech that can be easily understood by an interlocutor) is a more realistic goal. A host of studies have demonstrated that certain types of training can result in more comprehensible L2 speech. This contribution considers research on training the perception and production of both segmental (i.e., speech sounds) and suprasegmental features (i.e., stress, rhythm, tone, intonation). Before we can determine whether a given pronunciation feature is easy or difficult to teach and—more importantly—to learn, we must focus on: 1) setting classroom priorities that place comprehensibility of L2 speech at the forefront; and 2) relying upon insights gained through research into L2 pronunciation training. The goal of the mini-review is to help contextualize the papers presented in this collection.

Keywords: second language, pronunciation, training, priorities, effectiveness, comprehensibility

OPEN ACCESS

Edited by:
Antonio Benítez-Burraco,
Sevilla University, Spain

Reviewed by:
Murray J Munro,
Simon Fraser University, Canada
John M. Levis,
Iowa State University, United States

****Correspondence:***
Mary Grantham O'Brien
mgobrien@ucalgary.ca

Specialty section:
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: *07 November 2020*
Accepted: *24 December 2020*
Published: *16 February 2021*

Citation:
O'Brien MG (2021) Ease and Difficulty in L2 Pronunciation Teaching: A Mini-Review.
Front. Commun. 5:626985.
doi: 10.3389/fcomm.2020.626985

INTRODUCTION

Researchers and teachers alike agree that most adult second language (L2) learners will not sound like native speakers and that speaking with a nonnative accent is normal (Derwing and Munro, 2009). Nonetheless, both teachers and students express a desire for learners to achieve native-accented speech (Timmis, 2002; Sifakis and Sougari, 2005; Scales et al., 2006). Thus, the nativeness principle (i.e., a belief that nativelike pronunciation is both achievable and enviable (Levis, 2005; Levis, 2020)), serves as an implied objective in many language classrooms. In spite of this, recent studies demonstrate that teachers engage only intermittently in classroom pronunciation training, primarily because they lack training (Derwing and Munro, 2015) or confidence (Baker, 2011) or because they have relatively little knowledge about how to teach and assess pronunciation (Baker and Murphy, 2011; Baker, 2014; Couper, 2017). When they do teach pronunciation in their classrooms, teachers tend to focus on segmental production (Foote et al., 2016; Levis, 2016; Couper, 2017), most probably because materials—especially textbooks—tend to focus on segments (Derwing et al., 2012a; Foote et al., 2016).

It is not surprising that teachers might be reluctant to teach pronunciation if their ultimate objective is native-accented speech. However, a host of recent studies have demonstrated that being understood is a more realistic goal (Derwing and Munro, 2015). The intelligibility principle, with its acknowledgment that most foreign-accented speech is comprehensible[1], thus guides recent L2 pronunciation research (Levis, 2005; Levis, 2020). Researchers generally agree that both segments

[1] Intelligibility and comprehensibility are terms that are used in research to describe methods for testing listeners' understanding of speech. Levis's (2005, 2020) intelligibility principle incorporates both intelligibility and comprehensibility.

and suprasegmental features play an important role in being understood (Derwing and Munro, 2015) and that explicit pronunciation training can have a positive impact on the comprehensibility of L2 speech (Derwing et al., 1998; Isaacs, 2009; Lee et al., 2014; Thomson and Derwing, 2015).

Given the unspoken classroom goal of native-accented speech coupled with the sporadic attention paid to pronunciation on the one hand, and the research focus on comprehensible speech and a recommendation for regular pronunciation instruction on the other hand, there is clearly a disconnect between pedagogical practice and research findings. This contribution's focus on teaching pronunciation therefore considers the notions of ease and difficulty from two perspectives: 1) setting classroom priorities that place comprehensibility of L2 speech at the forefront; and 2) relying upon insights into research-informed L2 pronunciation training.

DEFINING EASE AND DIFFICULTY IN L2 PRONUNCIATION TEACHING[2]

Determining whether a given pronunciation feature—segmental or suprasegmental—is more or less difficult to learn depends on the extent to which improvement is shown after training. Given the variation in how pronunciation features are trained, how speech samples are elicited (e.g., reading individual words, sentences or paragraphs; repetition of a model speaker; semi-spontaneous or spontaneous utterances), and how improvement is measured (e.g., acoustic analyses, listener intelligibility tasks, listener ratings of comprehensibility and/or foreign accentedness), the field of L2 pronunciation research does not have an agreed-upon standard for determining whether a given type of training is successful. Nonetheless, the results of two recent meta-analyses have shown that pronunciation instruction almost always leads to improvement (Lee et al., 2014; Thomson and Derwing, 2015).

As a starting point in distinguishing between easy and difficult pronunciation features, it is important to consider the factors that may play a role in L2 pronunciation. First among these is language pairings: the combination of a learner's first language (L1) and their L2. Studies investigating similar groups of L1 learners of the same L2 often report conflicting results. For example, although the Japanese speakers in Haslam (2011) did not show improvement in English /l/ and /ɹ/ production even after training, other studies have shown improvement on these same segments among Japanese learners (e.g., Hardison, 2003; Hazan et al., 2005). The Mandarin native speakers who were trained in English vowel perception in Wang (2002) did not improve in their production of English vowels, but those in Thomson (2011) did. Given these inconsistent findings, it is clear that other factors must be at play in the ultimate success of pronunciation training. As such, L2 pronunciation researchers look beyond language pairings in their assessments of success of a given type of training. Additional factors may include participant's age of learning (Aoyama et al., 2008; Baker, 2010), quality of target language interactions (Derwing and Munro, 2015), motivational factors (Nagle, 2018), and learners' involvement in instructional decisions (Jenkins, 2004).

SETTING PRIORITIES

When it comes to determining which pronunciation features are easy and which are hard to learn, some research has shown that certain features are so easy to learn that they do not need to be trained. For example, the Mandarin- and Slavic-speaking learners of English in Derwing et al. (2012b) demonstrated an ability to accurately perceive sentence stress, intonation and the -teen/-ty distinction in the absence of instruction. While we should not deduce from such findings that accurate perception will result in accurate production, it makes little sense to train such features—in this case the perception thereof—in the classroom or to investigate their development. Moreover, individual variation is also quite common, and certain exceptional learners may not require training. For example, two Dutch-speaking learners of Slovak in Hanulíková et al. (2012) demonstrated nativelike perception and pronunciation of Slovak consonant clusters after only 15 min of exposure to the language. It is thus important to know which pronunciation features learners have mastered so that teachers do not waste time focusing on features that do not need to be trained.

In order to determine which pronunciation features learners have difficulty with and thus which should be the focus of classroom training, instructors are encouraged to develop a pronunciation needs assessment as described by Derwing and Munro (2015). Instructors should consider collecting both read and extemporaneous speech samples and assessing the samples both globally and analytically to determine learners' difficulties. The authors note that a perceptual task that requires learners to demonstrate their ability to perceive relevant segmental and suprasegmental distinctions can further guide the development of a pronunciation curriculum.

With the results of an assessment in hand, teachers are able to set priorities for their classrooms. Those pronunciation features that both cause difficulty and affect learners' comprehensibility—or those with the highest functional load (Catford, 1987)—should be the focus of training. At the segmental level, functional load can be determined, among other things, on the basis of the number of minimal pairs that are distinguished by two segments. For example, contrasting /l/ and /n/ distinguishes more English words than does producing a contrast between /d/ and /ð/ (Munro and Derwing, 2006). Although researchers have not established a functional load hierarchy for prosodic features of English, lexical (Zielinski, 2008; Isaacs and Trofimovich, 2012) and sentential stress assignment[3] (Hahn, 2004) both play an important role in being understood. While we have a good idea of which pronunciation features of

[2] An anonymous reviewer brought up the important point that it is possible to teach something well and for learners not to learn it. As such, the issue that we are most concerned with is that of learnability.

[3] Readers are reminded that L2 learners of English may not require training in the perception of sentential stress assignment as demonstrated by Derwing et al. (2012b).

English play a central role in understanding speech, that work is lacking for other target languages. Thus, when setting both segmental and suprasegmental pronunciation priorities in classes with target languages other than English, teachers are encouraged in their evaluation of their students' pronunciation needs assessments to consider the extent to which producing given distinctions plays a role in their ability to understand their students' speech.

EVALUATING THE EFFECTIVENESS OF TRAINING

Language learners—especially those in the early stages of language learning—tend to show improvement in their pronunciation over time. Thus, in order to determine whether a given type of training is effective, it is important when conducting research to include both a comparison group that receives a different type of training and a control group that receives no training. In addition, a delayed posttest allows researchers to determine whether the effects of training are long lasting (Thomson and Derwing, 2015).

Pronunciation improvement can be determined in two main ways: listener ratings and acoustic analyses. While listener ratings of understanding are considered the gold standard in pronunciation research (Derwing and Munro, 2009), some training studies also make use of acoustic analyses. Much of the research investigating the effectiveness of pronunciation training uses measures of understanding including comprehensibility ratings (e.g., Foote and McDonough, 2017; Martin, 2018) or intelligibility tasks (e.g., Derwing et al., 2014), often together with ratings of fluency and/or foreign accentedness. Acoustic analyses, completed by hand (e.g., Counselman, 2015) or automatically (e.g., Suemitsu et al., 2015; Tejedor-García et al., 2020) are also common and can be used to determine the extent to which certain pronunciation features change over time. Researchers note, however, that significant acoustic differences may not align with listener judgments (Derwing and Munro, 2015).

While few classroom teachers are able to carry out systematic analyses of their students' pronunciation development, they are encouraged to rely upon pronunciation training methods whose effectiveness has been demonstrated via research. Some of this work is outlined below.

RESEARCH-INFORMED PRONUNCIATION TRAINING

After setting priorities, the next step is to choose how to most effectively train pronunciation. While a teacher's status as a native or nonnative speaker of the target language does not play a role in learners' ultimate pronunciation (Levis et al., 2016), the results of research have generally demonstrated that explicit, form-focused instruction along with corrective feedback provides the greatest benefits to learners (Saito and Lyster, 2012; Saito, 2013). Derwing et al. (2014) describe an emergent training program designed to meet English language learners' (L1 = Vietnamese or Khmer) workplace needs. The classroom instruction, which targeted both perception and production, focused on those aspects of the participants' speech that affected their intelligibility (i.e., consonant clusters, rhythm and intonation). Participants' comprehensibility improved after only 17 h of classroom-based training.

A relatively large number of recent studies have investigated the effectiveness of ways to train pronunciation outside of the classroom. Researchers point to a number of benefits of computer-assisted pronunciation training (CAPT). These include unlimited practice time and flexibility as well as opportunities for varied input and immediate feedback (Engwall et al, 2004; Levis, 2007). Gao and Hanna (2016) indicate a further benefit: a computer's capacity for providing "infinite, patient modeling" (p. 214). An element of fun is also often added to CAPT. For example, Barcomb and Cardoso (2020) demonstrate the effectiveness of gamified pronunciation training (i.e., training that includes elements of a game but that is not actually a game). The Japanese junior high school learners of English in that study were rewarded with points and badges as they completed a series of metalinguistic tasks and perception and pronunciation activities focusing on English /l/ and /ɹ/. Learners in the study demonstrated both increased metalinguistic awareness and improved pronunciation accuracy over time. While a range of CAPT activity types exist, this contribution will focus on three that have been shown to play a positive role in improving learners' production: 1) listen and repeat; 2) perceptual training; and 3) visualization.

Although the effectiveness of traditional listen and repeat pronunciation tasks may be limited (O'Brien, 2019), a popular and effective way of training pronunciation by listening to a recording and then recording oneself is shadowing. The English learners in Foote and McDonough (2017) completed eight weeks of shadowing tasks in which they immediately repeated and recorded themselves while echoing dialogues from a sitcom as closely as possible. The task encouraged learners to focus on suprasegmental aspects of speech. Listeners rated pre-test, mid-training and post-test extemporaneous recordings for comprehensibility, accentedness and fluency. The authors found that learners had positive attitudes toward the activities and that learners' comprehensibility and fluency improved over time. A number of additional researchers have demonstrated the effectiveness of shadowing for the development of both segments (Zając and Rojczyk, 2014) and suprasegmental features (Lima, 2015).

Studies have investigated the efficacy of perceptual training for improving production (e.g., Counselman, 2015; Lee and Lyster, 2016; Sakai and Moorman, 2018). A popular and effective means of improving primarily segmental production through perceptual training is high variability phonetic training (HVPT), which trains listeners' perception with a relatively large quantity of speech samples that are produced by multiple speakers in a range of phonetic contexts (Thomson, 2018). The results of HVPT studies speak in its favor for the improvement of English vowels by native speakers of Greek (Lengeris, 2018), Mandarin (Thomson, 2011) and French (Iverson et al., 2011), as well as for the improvement of English consonants including English /l/ and /ɹ/ by Japanese speakers (Bradlow et al., 1997) and a number of English consonants by Korean learners (e.g., Huensch and Tremblay, 2015; Lee and Hwang, 2016). An additional type of perceptual training that has shown positive results is the use of speech synthesis systems (Mixdorff and Munro, 2013). For example, Liakin et al., (2017) found that L2 learners of French who made use of a simple text-to-speech (TTS) app on their mobile devices improved similarly to those learners who engaged in conversational practice with, and received feedback on their pronunciation from, their teachers in their in their production of French liaison. A highly innovative synthesis system that has demonstrated great promise generates a synthetic, native-accented version of a speaker's own voice (Ding et al., 2019). Participants in the study who made use of this so-called "golden speaker" version of their own voices showed improved comprehensibility and fluency.

Visualization techniques—including the use of acoustic displays (i.e., waveforms, spectrograms, and pitch tracks), ultrasound images that provide feedback on articulatory processes, and talking heads that provide learners with access to facial movements—allow learners to receive real-time visual feedback on productions. Tools used for visualization can include those designed for acoustic analyses such as Praat (Boersma and Weenink, 2020) and Audacity (Audacity Team, 2020) along with software that has been designed specifically to focus on L2 learners' pronunciation (e.g., Godfroid et al., 2017). At the segmental level, researchers have demonstrated that teaching learners how to interpret formant frequencies may enable them to improve their vowel productions, as demonstrated the native speakers of Japanese learning American English /æ/ in Suemitsu et al. (2015).[4] The English-Spanish L2 learners in Olson (2019), Offerman and Olson (2016), and Olson and Offerman (2020) who learned to interpret waveforms and spectrograms showing Spanish voice onset time also showed improvement after instruction. A number of researchers advocate for the use of waveforms and spectrograms for the teaching of suprasegmentals, especially duration and intonation (e.g., Levis, 1999; Hardison, 2004; Chun, 2013). For example, Levis and Pickering (2004) demonstrated the effectiveness of teaching contextualized discourse intonation to L2 learners of English by tracking intonation contours. The L2 Japanese learners in Okuno and Hardison (2016) received either audiovisual training consisting of audio files and waveform displays, audio-only training, or no training on vowel duration in Japanese. While participants in both experimental groups showed improvement and the ability to generalize what they learned to novel stimuli and new voices, participants in the audiovisual group improved their productions more than participants in the audio-only group. Similarly, Motohashi-Saigo and Hardison (2009) demonstrated the effectiveness of visualizations in learning vowel length and singleton/geminate distinctions. Chun et al. (2015) showed that L2 learners of Mandarin who compared the pitch contours of their own tone production with those of native speakers improved in their production of tones.

The type of feedback learners receive plays an important role in the extent of their improvement. Lee and Lyster (2016) investigated the effect of different types of corrective feedback on a series of perceptual tasks on the production accuracy of Korean-English L2 learners' vowels. Corrective feedback that took the form of either 1) rejection (i.e., indicating that the chosen answer was wrong) together with the target form; or 2) rejection together with the nontarget form was more effective than feedback that included either 3) a rejection along with both the target and nontarget forms; or 4) rejection only. The authors take this as evidence that providing learners with feedback indicating that their responses are incorrect is not sufficient for learning to occur.

It is important to consider that computer software designed to assess pronunciation "is not based on any particular theory or model of pronunciation which differentiates variation from (true) error" (Pennington, 1999; p. 431). As such, most CAPT promotes accuracy over intelligibility (Levis, 2007). Finally, although automatic speech recognition (ASR), which relies on a combination of acoustic analyses and artificial intelligence, has been touted as a promising way to evaluate and provide feedback on pronunciation (O'Brien et al., 2018), a number of researchers point to the relatively few studies that align ASR error detection and human judgments of speech (e.g., Chun, 2013; Chen and Li, 2016; Johnson and Kang, 2017; McCrocklin and Edalatishams, 2020).[5]

ADDITIONAL FACTORS

In addition to the type of pronunciation training and feedback learners receive, a number of other factors play a role in the success of training. Central among these is learner awareness. Although research has generally shown that learners have difficulty assessing their own pronunciation (e.g., Trofimovich et al., 2016), learners' awareness of pronunciation features may be positively related to listeners' comprehensibility ratings of their speech (Kennedy and Trofimovich, 2010). Explicit tasks that encourage awareness may be especially beneficial. For example, Añorga and Benander (2015) demonstrated the effectiveness of tasks that encourage learners to compare their own productions with models. Along similar lines, in addition to carrying out a range of production tasks, the German L2 learners in Martin (2018) completed tasks that required them to distinguish between foreign-accented and native speech. Their comprehensibility improved over time.

Additional factors that may play a role in the effectiveness of pronunciation training can include learners' proficiency levels, the length of training, and number of trained phonemes (Sakai and Moorman, 2018). Research has demonstrated that learners at lower levels of proficiency tend to make faster progress than more advanced learners (Sakai and Moorman, 2018), that there is an optimal length of pronunciation training (Lee et al., 2014; Olson and Offerman, 2020), and that the number of targeted phonemes should be constrained, possibly to as few as three (Sakai and Moorman, 2018).[6]

[4] Making use of spectrograms to interpret formant frequencies requires specialized knowledge, and this may be difficult for some teachers and learners (O'Brien et al., 2018).

[5] Garcia et al. (2020) demonstrated that the effectiveness of ASR training for the development of some L2 segments.

[6] Note, however, that Nishi and Kewley-Port (2007) report detrimental effects for training only a subset of vowels or consonants and advocate instead for training the entire set of vowels.

CONCLUSION

Accessing tools to train pronunciation has never been easier. Any language learner has easy access to a multitude of apps that promise to reduce accents quickly and easily. The focus of many of these tools, however, is often highly salient sounds that often do not play a role in comprehensibility and that may never improve after hours of training (Foote and Smith, 2013). This mini-review was written to provide readers of this collection with a background into the field of pronunciation training. Distinguishing between the notions of ease and difficulty in pronunciation teaching is overall much less important than distinguishing between effective and ineffective types of training. This is especially true if we consider the ultimate goal of pronunciation training to be comprehensible L2 speech.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Añorga, A., and Benander, R. (2015). Creating a pronunciation profile of first-year Spanish students. *Foreign Lang. Ann.* 48 (3), 434–446. doi:10.1111/flan.12151

Aoyama, K., Guion, S., Flege, J. E., Yamada, T., and Akahane-Yamada, R. (2008). The first years in an L2-speaking environment: a comparison of Japanese children and adults learning American English. *Int. Rev. Appl. Linguist.* 46 (1), 61–90. doi:10.1515/IRAL.2008.003

Audacity Team (2020). Audacity. Available at: https://www.audacityteam.org/.

Baker, A. (2011). Discourse prosody and teachers' stated beliefs and practices. *TESOL J.* 2 (3), 263–292. doi:10.5054/tj.2011.259955

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: teacher cognitions, observed classroom practices, and student perceptions. *Tesol Q.* 48 (1), 136–163. doi:10.1002/tesq.99

Baker, A., and Murphy, J. (2011). Knowledge base of pronunciation teaching: staking out the territory. *TESL Can. J.* 28 (2), 29–50. doi:10.18806/tesl.v28i2.1071

Baker, W. (2010). Effects of age and experience on the production of English word–final stops by Korean speakers. *Biling. Lang. Cognit.* 13 (3), 263–278. doi:10.1017/S136672890999006X

Barcomb, M., and Cardoso, W. (2020). Rock or lock? Gamifying an online course management system for pronunciation instruction: focus on English/r/and/l/. *CALICO J.* 37 (2), 127–147. doi:10.1558/cj.36996

Boersma, P., and Weenink, D. (2020). Praat: Doing phonetics by computer [Computer program] version 6.1.27. Available at: http://www.praat.org/ (Accessed October 25, 2020).

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997). Training Japanese listeners to identify English/r/and/l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am.* 101 (4), 2299–2310. doi:10.1121/1.418276

Catford, J. C. (1987). "Phonetics and the teaching of pronunciation," in *Current perspectives on pronunciation: practices anchored in theory*. Editor J. Morley (Alexandria, VA: Teachers of English to Speakers of Other Languages), 87–100.

Chen, N. F., and Li, H. (2016). "Computer-assisted pronunciation training: from pronunciation scoring towards spoken language learning," in Asia-Pacific signal and information processing association annual summit and conference (APSIPA), Jeju, South Korea, December 13–16, 2016 (Seoul, South Korea: IEEE), 1–7. doi:10.1109/APSIPA.2016.7820782

Chun, D. M. (2013). "Computer-assisted pronunciation teaching," in *Encyclopedia of applied linguistics*. Editor C. A. Chapelle (Oxford, United Kingdom: Wiley-Blackwell), 823–834. doi:10.1002/9781405198431.wbeal0172

Chun, D. M., Jiang, Y., Meyr, J., and Yang, R. (2015). Acquisition of L2 Mandarin Chinese tones with learner-created tone visualizations. *J. Sec. Lang. Pron.* 1 (1), 86–114. doi:10.1075/jslp.1.1.04chu

Counselman, D. (2015). Directing attention to pronunciation in the second language classroom. *Hispania* 98 (1), 31–46. doi:10.1353/hpn.2015.0006

Couper, G. (2017). Teacher cognition of pronunciation teaching: teachers' concerns and issues. *Tesol Q.* 51 (4), 820–843. doi:10.1002/tesq.354

Derwing, T. M., Diepenbroek, L. G., and Foote, J. A. (2012a). How well do general skills ESL textbooks address pronunciation? *TESL Can. J.* 30 (1), 22–44. doi:10.18806/tesl.v30i1.1124

Derwing, T. M., Thomson, R. I., Foote, J. A., and Munro, M. J. (2012b). A longitudinal study of listening perception in adult learners of English: implications for teachers. *Can. Mod. Lang. Rev.* 68 (3), 247–266. doi:10.3138/cmlr.1215

Derwing, T. M., and Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Lang. Teach.* 42 (4), 476–490. doi:10.1017/S026144480800551X

Derwing, T. M., and Munro, M. J. (2015). *Pronunciation fundamentals: evidence-based perspectives for L2 teaching and research*. Amsterdam, Netherlands: John Benjamins.

Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., and Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: a workplace training study. *Lang. Learn.* 64 (3), 526–548. doi:10.1111/lang.12053

Derwing, T. M., Munro, M. J., and Wiebe, G. E. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Lang. Learn.* 48 (3), 393–410. doi:10.1111/0023-8333.00047

Ding, S., Liberatore, C., Sonsaat, S., Lucič, I., Silpachai, A., Zhao, G., et al. (2019). Golden speaker builder: an interactive tool for pronunciation training. *Speech Commun.* 115, 51–66. doi:10.1016/j.specom.2019.10.005

Engwall, O., Wik, P., Beskow, J., and Granström, B. (2004). Design strategies for a virtual language tutor. 8th international conference on spoken, Jiju Island, South Korea, March 31, 2004 (Seoul, South Korea: ISCA), 1–4. Available at: http://www.speech.kth.se/ctt/publications/papers04/icslp2004_tutor.pdf.

Foote, J., and McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *J. Sec. Lang. Pron.* 3 (1), 34–56. doi:10.1075/jslp.3.1.02foo

Foote, J., and Smith, G. (2013). Is there an app for that? PhD thesis. Montreal, QC, Canada: Concordia University.

Foote, J. A., Trofimovich, P., Collins, L., and Soler Urzúa, F. (2016). Pronunciation teaching practices in communicative second language classes. *Lang. Learn. J.* 44 (2), 181–196. doi:10.1080/09571736.2013.784345

Gao, Y., and Hanna, B. E. (2016). Exploring optimal pronunciation teaching: Integrating instructional software into intermediate-level EFL classes in China. *CALICO J.* 33 (2), 201–230. doi:10.1558/cj.v33i2.26054

Garcia, C., Nickolai, D., and Jones, L. (2020). Traditional versus ASR-based pronunciation instruction: an empirical study. *CALICO J.* 37 (3), 213–232. doi:10.1558/cj.40379

Godfroid, A., Lin, C.-H., and Ryu, C. (2017). Hearing and seeing tone through color: an efficacy study of web–based, multimodal Chinese tone perception training. *Lang. Learn.* 67 (4), 819–857. doi:10.1111/lang.12246

Hahn, L. D. (2004). Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *Tesol Q.* 38 (2), 201–232. doi:10.2307/3588378

Hanulíková, A., Dediu, D., Fang, Z., Basnaková, J., and Huettig, F. (2012). Individual differences in the acquisition of a complex L2 phonology: a training study. *Lang. Learn.* 62 (2), 79–109. doi:10.1111/j.1467-9922.2012.00707.x

Hardison, D. M. (2003). Acquisition of second-language speech: effects of visual cues, context, and talker variability. *Appl. Psycholinguist.* 24 (4), 495–522. doi:10.1017/S0142716403000250

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: quantitative and qualitative findings. *Lang. Learn. Technol.* 8 (1), 34–52. doi:10.125/25228

Haslam, M. (2011). The effect of perceptual training including required lexical access and meaningful linguistic context on second language phonology. PhD dissertation. Salt Lake City, Utah: University of Utah.

Hazan, V., Sennema, A., Iba, M., and Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Commun.* 47 (3), 360–378. doi:10.1016/j.specom.2005.04.007

Huensch, A., and Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *J. Phonetics* 52, 105–120. doi:10.1016/j.wocn.2015.06.007

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Can. J.* 27 (1), 1–12. doi:10.18806/tesl.v27i1.1034

Isaacs, T., and Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Stud. Sec. Lang. Acquis.* 34 (3), 475–505. doi:10.1017/S0272263112000150

Iverson, P., Pinet, M., and Evans, B. G. (2011). Auditory training for experienced and inexperienced second language learners: native French speakers learning English vowels. *Appl. Psycholinguist.* 33 (1), 145–160. doi:10.1017/S0142716411000300

Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annu. Rev. Appl. Ling.* 24, 109–125. doi:10.1017/S0267190504000054

Johnson, D. O., and Kang, O. (2017). "Measures of intelligibility in different varieties of English: human vs. machine," in Proceedings of the 8th pronunciation in second language learning and teaching conference Editors M. O'Brien and J. Levis, Santa Barbara, CA, September, 2017. (Ames, IA: Iowa State University), 58–72. Available at: https://apling.engl.iastate.edu/alt-content/uploads/2017/05/PSLLT_2016_Proceedings_finalB.pdf.

Kennedy, S., and Trofimovich, P. (2010). Language awareness and second language pronunciation: a classroom study. *Lang. Aware.* 19 (3), 171–185. doi:10.1080/09658416.2010.48643

Lee, A. H., and Lyster, R. (2016). Can corrective feedback on second language speech perception errors affect production accuracy? *Appl. Psycholinguist.* 38 (2), 371–393. doi:10.1017/S0142716416000254

Lee, H. Y., and Hwang, H. (2016). Gradient of learnability in teaching English pronunciation to Korean learners. *J. Acoust. Soc. Am.* 139, 1859–1872. doi:10.1121/1.4945716

Lee, J., Jang, J., and Plonsky, L. (2014). The effectiveness of second language pronunciation instruction: a meta-analysis. *Appl. Ling.* 36 (3), 1–23. doi:10.1093/applin/amu040

Lengeris, A. (2018). Computer-based auditory training improves second-language vowel production in spontaneous speech. *J. Acoust. Soc. Am.* 144 (3), EL165–EL171. doi:10.1121/1.5052201

Levis, J. (1999). Intonation in theory and practice, revisited. *Tesol Q.* 33 (1), 37–63. doi:10.2307/3588190

Levis, J. (2007). Computer technology in teaching and researching. *Annu. Rev. Appl. Ling.* 27, 184–202. doi:10.1017/S0267190508070098

Levis, J. (2020). Revisiting the intelligibility and nativeness principles. *J. Sec. Lang. Pronunciation* 6 (3), 310–328. doi:10.1075/jslp.20050.lev

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *Tesol Q.* 39 (3), 369–377. doi:10.2307/3588485

Levis, J. M. (2016). Research into practice: how research appears in pronunciation teaching materials. *Lang. Teach.* 49 (3), 423–437. doi:10.1017/S0261444816000045

Levis, J. M., Sonsaat, S., Link, S., and Barriuso, T. A. (2016). Native and nonnative teachers of L2 pronunciation: effects on learner performance. *Tesol Q.* 50 (4), 894–951. doi:10.1002/tesq.272

Levis, J., and Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System* 32, 505–524. doi:10.1016/j.system.2004.09.009

Liakin, D., Cardoso, W., and Liakina, N. (2017). The pedagogical use of mobile speech synthesis (TTS): focus on French liaison. *Comput. Assist. Lang. Learn.* 30 (3–4), 348–365. doi:10.1080/09588221.2017.1312463

Lima, E. F. (2015). "Feel the rhythm! Fun and effective pronunciation practice using Audacity and sitcom scenes (teaching tip)," in Proceedings of the 6th pronunciation in second language learning and teaching conference Editors J. Levis, R. Mohammed, M. Qian, and Z. Zhou, Santa Barbara, CA, September 5–6, 2014 (Ames, IA: Iowa State University), 277–284. Available at: https://apling.engl.iastate.edu/alt-content/uploads/2015/05/PSLLT_6th_Proceedings_2014.pdf.

Martin, I. A. (2018). Bridging the gap between L2 pronunciation research and teaching: using iCPRs to improve German learners' pronunciation in distance and face-to-face classrooms. PhD dissertation. State College, PA: The Pennsylvania State University.

McCrocklin, S., and Edalatishams, I. (2020). Revisiting popular speech recognition software for ESL speech. *Tesol Q.* 54 (4), 1086–1097. doi:10.1002/tesq.3006

Mixdorff, H., and Munro, M. J. (2013). Quantifying and evaluating the impact of prosodic differences of foreign–accented English. Proceedings of the workshop on speech and language technology in education (SLaTE), Gernoble, France, August 30–September 1, 2013 (Valencia, Spain: ISCA), 147–152.

Motohashi-Saigo, M., and Hardison, D. M. (2009). Acquisition of L2 Japanese geminates: training with waveform displays. *Lang. Learn. Technol.* 13 (2), 29–47. doi:10.125/44179

Munro, M. J., and Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: an exploratory study. *System* 34 (4), 520–531. doi:10.1016/j.system.2006.09.004

Nagle, C. (2018). Motivation, comprehensibility, and accentedness in L2 Spanish: investigating motivation as a time-varying predictor of pronunciation development. *Mod. Lang. J.* 102 (1), 199–217. doi:10.1111/modl.12461

Nishi, K., and Kewley-Port, D. (2007). Second language vowel production training: effects of set size, training order and native language. Available at http://www.icphs2007.de/conference/Papers/1018/1018.pdf.

Offerman, H. M., and Olson, D. J. (2016). Visual feedback and second language segmental production: the generalizability of pronunciation gains. *System* 59, 45–60. doi:10.1016/j.system.2016.03.003

Okuno, T., and Hardison, D. M. (2016). Perception–production link in L2 Japanese vowel duration: training with technology. *Lang. Learn. Technol.* 20 (2), 61–80. doi:10.125/44461

Olson, D. J. (2019). Feature acquisition in second language phonetic development: evidence from phonetic training. *Lang. Learn.* 69 (2), 366–404. doi:10.1111/lang.12336

Olson, D. J., and Offerman, H. M. (2020). Maximizing the effect of visual feedback for pronunciation instruction: a comparative analysis of three approaches. *J. Sec. Lang. Pronunciation*. doi:10.1075/jslp.20005.ols

O'Brien, M. G. (2019). "Targeting pronunciation (and perception) with technology," in *Engaging language learners through CALL*. Editors N. Arnold and L. Ducate (Sheffield, United Kingdom: Equinox), 309–352.

O'Brien, M. G., Derwing, T. M., Cucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R., et al. (2018). Directions for the future of technology in pronunciation research and teaching. *J. Sec. Lang. Pronunciation* 4 (2), 182–207. doi:10.1075/jslp.17001.obr

Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: promise, limitations, directions. *Comput. Assist. Lang. Learn.* 12 (5), 427–440. doi:10.1076/call.12.5.427.5693

Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development. *Stud. Sec. Lang. Acquis.* 35 (1), 1–29. doi:10.1017/S0272263112000666

Saito, K., and Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of/ɹ/by Japanese learners of English. *Lang. Learn.* 62 (2), 595–633. doi:10.1111/j.1467-9922.2011.00639.x

Sakai, M., and Moorman, C. (2018). Can perception training improve the production of second-language phonemes? A meta-analytic review of 25 years of perception training research. *Appl. Psycholinguist.* 39 (1), 187–224. doi:10.1017/S0142716417000418

Scales, J., Wennerstrom, A., Richard, D., and Wu, S. H. (2006). Language learners' perceptions of accent. *Tesol Q.* 40, 715–738. doi:10.2307/40264305

Sifakis, N. C., and Sougari, A.-M. (2005). Pronunciation issues and EIL pedagogy in the periphery: a survey of Greek state school teachers' beliefs. *Tesol Q.*, 39, 467–488. doi:10.2307/3588490

Suemitsu, A., Dang, J., Ito, T., and Tiede, M. (2015). A real-time articulatory visual feedback approach with target presentation for second language pronunciation learning. *J. Acoust. Soc. Am.* 138 (4), EL382–EL387. doi:10.1121/1.4931827

Tejedor-García, C., Escudero-Mancebo, D., Cardeñoso-Payo, V., and González-Ferreras, C. (2020). Using challenges to enhance a learning game for pronunciation training of English as a second language. *IEEE Access* 8, 74250–74266. doi:10.1109/ACCESS.2020.2988406

Thomson, R. I. (2011). Computer-assisted pronunciation training: targeting second language vowel perception improves pronunciation. *CALICO J.* 28 (3), 744–765. doi:10.11139/cj.28.3.744-765

Thomson, R. I., and Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: a narrative review. *Appl. Ling.* 36 (3), 326–344. doi:10.1093/applin/amu076

Thomson, R. I. (2018). High variability [pronunciation] training. (HVPT). A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation* 4 (2), 208–231. doi:10.1075/jslp.17038.tho

Timmis, I. (2002). Native-speaker norms and international English: a classroom view. *ELT J.* 56 (3), 240–249. doi:10.1093/elt/56.3.240

Trofimovich, P., Iaacs, T., Kennedy, S., Saito, K., and Crowther, D. (2016). Flawed self-assessment: investigating self-and other-perception of second language speech. *Bilingualism* 19 (1), 122–140. doi:10.17/S1366728914000832

Wang, X. (2002). *Training Mandarin and Cantonese speakers to identify English vowel contrasts: long-term retention and effects on production.* Burnaby, Canada: Simon Fraser University.

Zając, M., and Rojczyk (2014). Imitation of English vowel duration upon exposure to native and non-native speech. *Poznań Stud. Contemp. Linguis.* 50 (4), 495–514. doi:10.1515/psicl-2014–0025

Zielinski, B. (2008). The listener: No longer the silent partner in reduced intelligibility. *System* 36 (1), 69–84. doi:10.1016/j.system.2007.11.004

Conflict of Interest: 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.

frontiers in Communication

MINI REVIEW
published: 03 March 2021
doi: 10.3389/fcomm.2021.626529

Ease and Difficulty in L2 Phonology: A Mini-Review

John Archibald *

Department of Linguistics, University of Victoria, Victoria, BC, Canada

A variety of phonological explanations have been proposed to account for why some sounds are harder to learn than others. In this mini-review, we review such theoretical constructs and models as markedness (including the markedness differential hypothesis) and frequency-based approaches (including Bayesian models). We also discuss experimental work designed to tease apart markedness versus frequency. Processing accounts are also given. In terms of phonological domains, we present examples of feature-based accounts of segmental phenomena which predict that the L1 features (not segments) will determine the ease and difficulty of acquisition. Models which look at the type of feature which needs to be acquired, and models which look at the functional load of a given feature are also presented. This leads to a presentation of the redeployment hypothesis which demonstrates how learners can take the building blocks available in the L1 and create new structures in the L2. A broader background is provided by discussing learnability approaches and the constructs of positive and negative evidence. This leads to the asymmetry hypothesis, and presentation of new work exploring the explanatory power of a contrastive hierarchy approach. The mini-review is designed to give readers a refresher course in phonological approaches to ease and difficulty in acquisition which will help to contextualize the papers presented in this collection.

Keywords: L2 phonology, L2 speech, second language acquisition (SLA), redeployment, learnability

OPEN ACCESS

Edited by:
Juhani Järvikivi,
University of Alberta, Canada

Reviewed by:
Anne-Michelle Tessier, University of British Columbia, Canada
John H. G. Scott,
University of Calgary, Canada

****Correspondence:***
John Archibald
johnarch@uvic.ca

Specialty section:
This article was submitted to Language Sciences,
a section of the journal
Frontiers in Communication

Received: *06 November 2020*
Accepted: *14 January 2021*
Published: *03 March 2021*

Citation:
Archibald J (2021) Ease and Difficulty in L2 Phonology: A Mini-Review.
Front. Commun. 6:626529.
doi: 10.3389/fcomm.2021.626529

INTRODUCTION

Why are some sounds harder to learn than others? A Japanese learner of English may have difficulty acquiring a novel L2 English [l]/[ɹ] contrast (Brown, 2000) but less difficulty acquiring a novel L2 Russian [l]/[r] contrast (Larson Hall, 2004). The same Japanese speaker may have no difficulty acquiring the novel L2 contrasts [b]/[v] or [s]/[θ] (Matthews, 2000). A Brazilian Portuguese learner of English may have difficulty acquiring consonant clusters such as [sl], [sn] or [st] which are absent from the L1 (Cardoso, 2007), while a Persian learner of English who also lacks L1 [sl], [sn] and [st] may find them quite easy to acquire (Archibald and Yousefi, 2018). A Spanish learner of English may find it easier to acquire the [i]/[ɪ] contrast (which is absent from the L1) when learning Scottish English than British English (Escudero, 2002). There are also examples of so-called directionality of difficulty effects (Eckman, 2004). For example, an English learner of German might find it easier to suppress a final voicing contrast than a German learner of English would find it to learn to make a new L2 final voicing contrast. These are the types of facts researchers need to explain (the *explanandum*). In this short paper, I will provide an overview of some of the proposed phonological accounts (the *explanans*) of such cases of ease or difficulty.

We begin by asking what it means to have *acquired* a sound. To probe such a question from a phonological perspective means that we must tackle the question of *contrast*. Phonemes are used to represent lexical contrasts. Such contrasts must also be implemented phonetically in both production and perception. Given that L2 production and perception may well be non-nativelike, this raises the interesting question for the L2 phonologist of determining whether the individual is 1) producing an inaccurate representation accurately, or 2) producing an accurate representation inaccurately. A case of 1) would be where an L2 learner might have the same representation for both /l/ and /r/ (i.e., not making a phonemic liquid contrast) and who also merged the production of [l] and [r]. A case of 2) would be where a learner might have a representational contrast for /b/ and /p/ (i.e., making a phonemic VOT contrast) but not implementing the contrast in a nativelike fashion. Methodologically, this reveals that researchers (and teachers) cannot rely on inaccurate production as a diagnostic of non-nativelike representation.

This leads us to a related question concerning production vs. perception. Much work in L2 speech proceeds on the assumption that accurate perception must (logically and developmentally) precede accurate production (Flege, 1995). Thus, much of the literature focusses on assessing whether the subjects can discriminate phonetic contrasts reliably, and represent phonological contrasts accurately. However, there are certain cases where learners may be accurate in either production (Goto, 1971) or lexical discrimination (Darcy et al., 2012) tasks and yet remain inaccurate on discrimination tasks. In both cases, it may be that metalinguistic knowledge plays an important role.

Ever since the Contrastive Analysis Hypothesis (Lado, 1957), linguists have tried to predict which aspects of L2 speech would be easy or difficult to learn. Since the 50s, both the representational models of phonology and the learning theories have become more sophisticated, and this has led to a consideration of multiple factors in exploring the construct of *difficulty*. Such approaches stand in marked contrast to the models of cross-language speech production (Flege, 1995) and cross-language speech perception (Best and Tyler, 2007) which primarily invoke acoustic and articulatory factors to explain difficulty in acquisition.

In the field of second language acquisition (SLA), there have been many factors explored to account for aspects of learner variation, including variation in nativelikeness of L2 speech. The following factors have been explored:

- L1 transfer (Trofimovich and Baker, 2006)
- amount of experience (Bohn and Flege, 1992)
- amount of L2 use (Guion et al., 2000)
- age of learning (Abrahamsson and Hyltenstam, 2009)
- orthography (Escudero and Wanrooi, 2010; Bassetti et al. 2015)
- frequency (Davidson, 2006)
- attention (Guion and Pederson, 2007)
- training (Wang et al., 2003)

It goes without saying that all of these factors *do* come in to play in accounting for learner behavior. What I will focus on in this mini-review are key representational issues which have informed phonological approaches to the construct of ease and difficulty.

REPRESENTATIONAL APPROACHES

This mini-review is focusing on representational models of phonology. There is a rich literature on output-based approaches (Tessier et al., 2013; Jesney, 2014) which tend to emphasize the computational system which generates the output form rather than emphasizing the form of the underlying (or *input*) representation.

Markedness

Some have looked to the notion of markedness (Parker, 2012) as an explanation by suggesting that unmarked structures are easier to acquire than marked ones (Carlisle, 1998). For example, it could be argued that 3-consonant onsets (e.g., [str]) were more difficult to acquire than 2-consonant onsets (e.g., [tr]) *because* they were more marked. Even within 2-consonant sequences work such as Broselow and Finer (1991), Eckman and Iverson (1993) demonstrate that principles such as Sonority Sequencing instantiate markedness with greater sonority distance between the adjacent segments being less marked (i.e., [pj] would be less marked than [fl]). Such machinery is designed to account for the observation that not all structures which are absent from the L1 are equally difficult to acquire in the L2. The developmental path would be from unmarked to marked structures.

Some have suggested that a markedness continuum was not enough but rather that markedness *differential* was the locus of explanation (Eckman, 1985). Under this approach, a structure which was absent from the L1 and more marked than the L1 structure would be difficult to acquire while one which was absent from the L1 but less marked than the L1 structure would be easier to acquire.

Often, however, the unmarked forms are the most frequent (e.g. 3-consonant clusters are more marked than 2-consonant clusters, and 3-consonant clusters are also less frequent than 2-consonant clusters) so it is difficult to tease these factors apart. If learners are more accurate on 2-consonant clusters is it because they are more frequent or less marked?

Frequency-Based Approaches

Usage-based (Wulff and Ellis, 2018) and Bayesian (Wilson and Davidson, 2009) approaches argue that targetlike production accuracy is correlated with input frequency. Thus, if there are two elements which are absent from the L1 and one is frequent in the L2 input while one is infrequent, then the frequent structure might be more easily acquired.

Frequency Versus Markedness

Cardoso (2007) documents a scenario in which the most frequent structure is the most marked so we can tell which construct is most explanatory. In looking at the acquisition of L2 English consonant clusters by L1 speakers of Brazilian Portuguese, he focused on [st], [sn] and [sl]. Without getting into the details of the markedness facts here, [st] is both the most frequent and the most marked of the clusters. When it came to learner production, the learners were least accurate on the most marked cluster ([st]) even though it was most frequent in their input. For production (though not perception), markedness seemed to be more explanatory than frequency.

The construct of markedness itself has its critics (Haspelmath, 2006; Zerbian, 2015). If the notion is ill-defined measure of complexity—difficulty or abnormality?—then how can it be a valid *explanans*? Responding to Archibald (1998) who suggested that positing markedness as an explanation (rather than a description) only bumped the explanation problem back a generation (because what explains markedness?), Eckman (2008; 105) counter-argues that, "to reject a hypothesis because it pushes the problem of explanation back one step misses the point that *all* hypotheses push the problem of explanation back one step–indeed, such 'pushing back' is necessary if one is to proceed to higher level explanations."

Processing Accounts

While more work has been done on the role of the processor in morpho-syntax in SLA (O'Grady, 1996; O'Grady, 2006; Truscott and Sharwood Smith, 2004), Carroll (2001) explores the role of the phonological parser in mapping the acoustic signal onto phonological representations. Carroll (2013) addresses these questions in initial-state L2 learners empirically. There has also been some work done on L2 phonological parsing at the level of the syllable (Archibald, 2003; Archibald, 2004; Archibald, 2017) which suggests that structures which can be parsed are easier to acquire than structures which the parser cannot yet handle.

Such models intersect with the perception literature insofar as the L2 acoustic input is filtered by the L1 phonological system (Pallier et al., 2001). In turn, such perceptual shoe-horning can lead to activation of phantom lexical competitors (Broersma and Cutler, 2007) which may slow lexical activation.

The notion that only some input can be processed at any given time, thus leading to the *intake* to the processor being a subset of the environment input, is well-studied in applied linguistics (Corder, 1967; Schmidt, 1990). What has proved more elusive is explaining when input becomes intake (and when it does not). Certainly one of the challenges is avoiding circularity of the following sort:

Q: why is x produced/perceived accurately before y?
A: Because it became intake
Q: How do you know it became intake?
A: Because it was produced/perceived accurately.

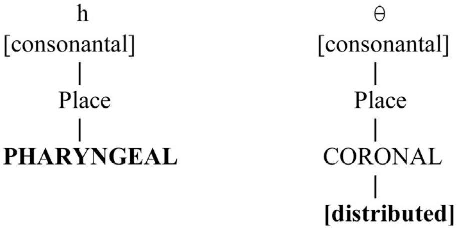
Processing accounts are not necessarily independent of abstract phonological studies as they have also been important in documenting the viability of abstract phonological features (Lahiri and Reetz, 2010; Schluter et al., 2017). Features can be explanatory when we note *classes* of sounds behaving in a similar fashion, for example, only nasals being allowed in syllable codas in a given L1. Thus difficulty may arise when these learners attempt to parse L2 stops into a coda. Note that the difficulty would affect, say, [p t k] as a class of voiceless stops.

Representational Accounts

Theories of phonological representation help us to model both synchronic and diachronic aspects of L2 phonological grammars. Özçelik (2016) addresses the general question of developmental path in L2 grammars (a fundamental concern of the field as we try to develop a transition theory). He proposes a cue-based model which clarifies which structural properties (i.e., parameters) are logical precursors to the acquisition of subsequent parameters. Özçelik and Sprouse (2016) demonstrate that interlanguage grammars are constrained by phonological universals (such as the behavior of feature spreading).

Feature-based models (Brown, 2000) can be contrasted with segment-based models (Flege, 1995). A segment-based model might say that a new segment will be difficult to acquire based on a comparison of the L1 and L2 phonetic categories. A feature-level account would argue that new L2 contrasts which were based on distinctive features that were absent from the L1 would be difficult while new contrasts based on L1 features would be easy. Brown (2000) showed that Korean learners of English could acquire new contrasts if the contrasts were based on an existing L1 feature (e.g., [continuant]) while L2 contrasts which were not based on L1 features (e.g., [distributed]) were more difficult to acquire.

LaCharité and Prévost (1999) suggest that this was too strong an approach and that some features which were absent (i.e., terminal nodes) would be acquirable while others (i.e., articulator nodes) would not, as shown in (1).

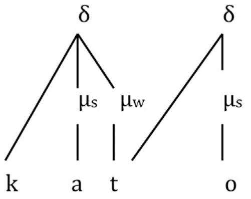


(1)

The features in boldface are the ones which are absent from the L1 French inventory. They predict that the acquisition of L2 English [h] will be more difficult than the acquisition of [θ] because [h] requires the learner to trigger a new articulator node. On a discrimination task, the learners were significantly less accurate identifying [h] than identifying [θ], however, on a word identification task (involving lexical access) there was no significant difference between the performance on [h] vs. [θ]. Özçelik and Sprouse (2016), however, show that L2 learners are able to acquire the features of secondary articulations (e.g., palatalized consonants). Hancin-Bhatt (1994) proposed that the functional load of a particular feature in implementing a contrast in a language would determine its weighting (with features with high functional load predicted to have greater cross-linguistic influence than those with low functional load).

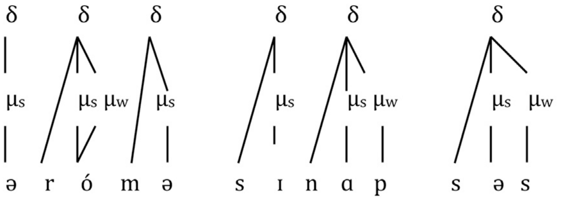
Archibald (2005) proposed the Redeployment Hypothesis in which it would be easier to acquire new L2 structures which could be built from existing L1 building blocks (e.g., features, or moras) than to acquire new building blocks. In some ways, this approach presages Lardiere's (2009) Feature Reassembly Hypothesis which looks to account for the difficulty that L2 learners have acquiring L2 morphology.

One example of redeployment is evidenced in the L2 acquisition of Japanese geminate consonants by L1 English speakers. Japanese geminate consonants have the moraic structure shown in (2).



(2)

English does not have geminate consonants, but does have a weight-sensitive stress system, shown in (3) where coda consonants project moras which attracts stress to heavy syllables.



(3)

Thus, the English quantity-sensitive system can be redeployed to acquire L2 geminates. The corollary to this would be that L2 structures which could not be built from L1 components would be more difficult to acquire.

Cabrelli et al. (2019), looking at Brazilian Portuguese learners of English coda consonants, also demonstrate that L2 learners can restructure their phonological grammars insofar as the L2 learners are licensing coda consonants which are not found in the L1. Carlson (2018) found similar effects in L1 Spanish.

Garcia (2020) describes an interesting case where a property of the L2 (stress placement) which could be acquired on the basis of transferring an L1 property of weight-sensitivity is, in fact, difficult to acquire because another property of the L1 is able to account for the L2 data, and this property (positional bias) is more robust in the L2 input.

Production, Perception and Representation

Darcy et al. (2012) present data which show, contra Flege (1995), that some learners who were able to lexically represent a contrast were unable to accurately discriminate it. The model is known as DMAP which stands for *direct mapping of acoustics to phonology.* The basic empirical finding which they report on is a profile where L2 learners of French (with L1 English which lacks/y/) can distinguish lexical items which rely on a /y/ - /u/ distinction while simultaneously being unreliable in discriminating [y] from [u] in an ABX task. Detection of acoustic properties can lead to phonological restructuring (according to general economy principles of phonological inventories) which will result in a lexical contrast but the phonetic categories may not yet be targetlike. The learners rely on their current interlanguage feature hierarchy to set up contrastive lexical representations even as phonetic category formation proceeds.

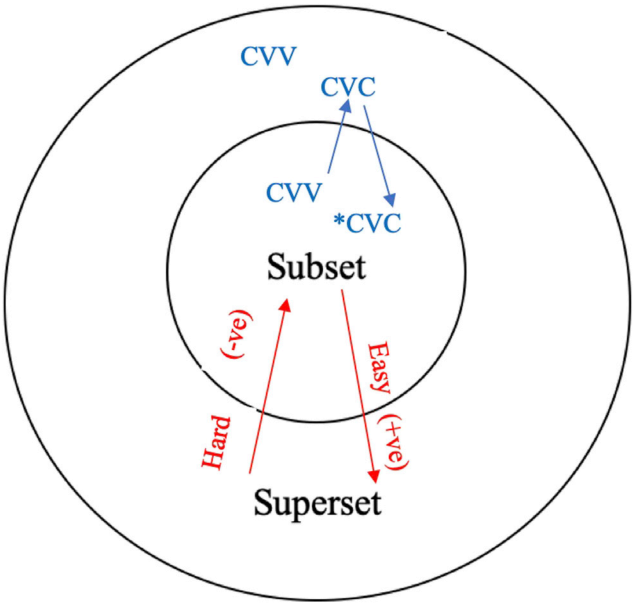


FIGURE 1 | The subset principle: positive (+ve) and negative (−ve) evidence and ease or difficulty of learning illustrated by the quantity-sensitivity of Hungarian and English stress assignment.

This is reminiscent of the Goto (1971) study where Japanese learners were able to produce an /l/-/r/ liquid contrast even while not being able to discriminate between them in a decontextualized task. It could be that the tactile feedback received in the production of these two sounds, and the orthographic distinction between "l" and "r" were able to cue the learners' production systems. This sort of metalinguistic knowledge can affect production.

Davidson and Wilson (2016) extend a body of research which documents L2 learners' sensitivity to non-contrastive phonetic properties (which might account for occurrences of prothesis vs. epenthesis in cluster repair) to look at learner behavior in the classroom. While subjects listening in a classroom (compared to a sound booth) showed some differences (e.g., less prothesis repair), by and large the performance was very similar. This suggests that laboratory research may well have quite direct implications for classroom learners.

Learnability and L2 Phonology

Learnability approaches (Wexler and Culicover, 1980; Pinker, 1989; White, 1991) argued that learning would be faster when there was positive evidence that the L1 grammar had to change, while change that was cued only by negative evidence would be acquired more slowly. Positive evidence is evidence in the linguistic environment of well-formed structures. Negative evidence is evidence given to the learner that a particular string is ungrammatical. It would be easier to move from an L1 which was a subset of the L2 (because there is positive evidence to indicate that the current grammar is incorrect) than it would be to move from an L1 which was a superset of the L2 (as this would require negative evidence).

Consider the example of L1 English and L2 Hungarian as shown in **Figure 1**. Hungarian secondary stress (Kerek, 1971) is quantity-sensitive to the Nucleus (meaning that only branching nucleii (i.e., long vowels (CVV)) are treated as Heavy but not branching Rhymes (i.e., closed syllables (CVC)). English stress is quantity sensitive to branching nuclei *and* branching rhymes.

If your L1treated long (i.e., bimoraic) vowels (CVV) *and* closed syllables (CVC) as heavy (as English does) but the L2 only treated long vowels as heavy then it might take a while for the learner to hypothesize "wait, I've never heard a secondary stress on a closed syllable!". But L1 Hungarian to L2 English would have clear positive evidence when the learner hears stress placed on a closed syllable (as in *agénda)*. An English learner of Hungarian would have to notice that Hungarian *never* stressed closed syllables. Dresher and Kaye (1991) argued that when the data reveal that closed syllables and branching nuclei behave the same with respect to stress assignment this is the universal cue for the system to be quantity sensitive to the rhyme. See Archibald (1991) for further discussion and empirical investigation.

Young-Scholten's (1994, 2004) Asymmetry Hypothesis predicts that if an L2 phonological rule applies in a prosodic domain that is a superset of the L1 phonological domain then the positive evidence will make it easier to acquire. However, when the target domain is smaller than the L1 domain then the lack of positive evidence will make acquisition more difficult. In English, the rule of flapping applies within a phonological utterance (e.g., *Don't sit on the mat* [ɾ], *it's dirty*.). German has a rule of final devoicing which applies within a phonological word (e.g., *Ich ha* [b]*e* ~ *Ich hab*[p]). So, English learners of German are predicted to have difficulty acquiring phonological patterns which are licensed only in smaller phonological domains.

In addition to positive evidence or direct negative (i.e., correction) evidence, however, Schwartz and Goad (2017) have demonstrated that indirect positive evidence can play a role in second language learning where the L2 is a subset of the L1. In this case, L2-accented English was shown to be a source of evidence for some subjects as to the phonotactics of Brazilian Portuguese.

There is one area which is just starting to be explored in L2 phonology and that is Dresher (2009) contrastive hierarchy as an explanatory tool for ease and difficulty. Dresher's model suggests that L2 features which are *active* (i.e., involved in many phonological processes in the language) will be easier to learn than L2 features which are inactive due to the type of evidence they present to the learner. Active features provide robust cues to the learner that a given feature must be highly ranked in a contrastive hierarchy, and is, therefore, evidence to restructure the L1 hierarchy. Archibald (2020) has explored this model in an analysis of L3 phonological systems. Such a mechanism is reminiscent of Hancin-Bhatt (1994) notion of how functional load defines featural prominence.

CONCLUSION

What I have attempted to show in this mini-review is that there is a rich history in addressing the question of ease vs. difficulty in L2 phonology. I hope that this overview will provide useful background to the readers of this collection. Unsurprisingly, there is no easy answer to the difficult question of *ease* vs. *difficulty*.

AUTHOR CONTRIBUTION

I am the sole author of this piece.

ACKNOWLEDGMENTS

I would like to thank the reviewers for this piece. Their keen eye for clarity and accuracy has greatly improved this mini-review. But I have to say that the friendly, supportive scholarly exchange was as enjoyable as it is rare.

REFERENCES

Abrahamsson, N., and Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: listener perception versus linguistic scrutiny. *Lang. Learn.* 59 (2), 249–306. doi:10.1111/j.1467-9922.2009.00507.x

Archibald, J. (1991). Language learnability and L2 phonology: the acquisition of metrical parameters. Dordrecht: Kluwer.

Archibald, J. (2004). Interfaces in the prosodic hierarchy: new structures and the phonological parser. *Int. J. Bilingualism* 8 (1), 29–50. doi:10.1177/13670069040080010301

Archibald, J. (2003). Learning to parse second language consonant clusters. *Can. J. Linguistics* 48 (3/4), 149–177. doi:10.1353/cjl.2004.0023

Archibald, J. (1998). *Second Language Phonology*. Amsterdam, NL: John Benjamins.

Archibald, J. (2005). Second language phonology as redeployment of L1 phonological knowledge. *Can. J. Linguistics* 50 (1/2/3/4), 285–314. doi:10.1353/cjl.2007.0000

Archibald, J. (2017). "Transfer, contrastive analysis and interlanguage phonology," in *The Routledge handbook of English pronunciation*. Editors R. Thompson, O. Kang, and J. Murphy (London, UK:Routledge), 9–24.

Archibald, J. (2020). in *Microvariation in multilingual situations: the importance of property-by-property acquisition*. Turtles all the way down: micro-cues and piecemeal transfer in L3 phonology. Commentary on Westergaard. *Second Language Research*. doi:10.1177/0267658320941036

Archibald, J., and Yousefi, M. (2018). "The redeployment of marked L1 Persian codas in the acquisition of marked L2 English onsets: redeployment as a transition theory," in Paper presented at conference on central Asian language and linguistics, Bloomington, IN, April 17–April 19, 2020 (Bloomington, IN: Indiana University).

Bassetti, B., Escudero, P., and Hayes-Harb, R. (2015). Second language phonology at the interface between acoustic and orthographic input. *Appl. Psycholinguistics* 36, 1–6. doi:10.1017/s0142716414000393

Best, C., 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*. Editors M. J. Munro and O.-S. Bohn (Amsterdam, NL: John Benjamins Publishing), 13–34.

Bohn, O. S., and Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Stud. Second Lang. Acquisition* 14 (2), 131–158. doi:10.1017/s0272263100010792

Broersma, M., and Cutler, A. (2007). Phantom word activation in L2. *System* 36, 22–34. doi:10.1016/j.system.2007.11.003

Broselow, E., and Finer, D. (1991). Parameter setting in second language phonology and syntax. *Second Lang. Res.* 7 (1), 35–59. doi:10.1177/026765839100700102

Brown, C. (2000). "The interrelation between speech perception and phonological acquisition from infant to adult," in *Second language acquisition and linguistic theory*. Editor J. Archibald (Hoboken, NJ: Wiley-Blackwell), 4–63.

Cabrelli, J., Luque, A., and Finestrat-Martinez, I. (2019). Influence of L2 English phonotactics in L1 Brazilian Portuguese illusory vowel perception. *J. Phonetics* 73, 55–69. doi:10.1016/j.wocn.2018.10.006

Cardoso, W. (2007). "The development of sC onset clusters in interlanguage: markedness vs. frequency effects," in Proceedings of the 9th generative approaches to second language acquisition conference (GASLA 2007), May 18–20, 2007. Editors R. Slabakova, J. Rothman, P. Kempchinsky, and E. Gavruseva (Cascadilla Press), 15–29.

Carlisle, R. S. (1998). The acquisition of onsets in a markedness relationship: a longitudinal study. *Studies in second language acquisition* 20, 254–260. doi:10.1017/S027226319800206X

Carlson, M. T. (2018). Making room for second language phonotactics: effects of L2 learning and environment on first language speech perception. *Lang. Speech* 61, 598–614. doi:10.1177/0023830918767208

Carroll, S. (2001). *Input and evidence: the raw material of Second Language acquisition*. Amsterdam, NL: John Benjamins Publishing

Carroll, S. (2013). Introduction to the special issue: aspects of word learning on first exposure to a second language. *Second Lang. Res.* 29 (2), 131–144. doi:10.1177/0267658312463375

Corder, S. P. (1967). The significance of learners' errors. *Int. Rev. Appl. Linguistics* 5, 160–170. doi:10.1515/iral.1967.5.1-4.161

Darcy, I., Dekydtspotter, L., Sprouse, R., Glover, J., Kaden, C., McGuire, M., and Scott, J. (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 (1), 5–40. doi:10.1177/0267658311423455

Davidson, L. (2006). Phonology, phonetics, or frequency: influences on the production of non-native sequences. *J. Phonetics* 34, 104–137. doi:10.1016/j.wocn.2005.03.004

Davidson, L., and Wilson, C. (2016). Processing nonnative consonant clusters in the classroom: perception and production of phonetic detail. *Second Lang. Res.* 32 (4), 471–501. doi:10.1177/0267658316637899

Dresher, B. E., and Kaye, J. (1991). A computational learning model for metrical phonology. *Cognition* 34, 137–195.

Dresher, E. B. (2009). *The contrastive hierarchy in phonology*. Cambridge, UK: Cambridge University Press.

Eckman, F. (2004). From phonemic differences to constraint rankings. *SSLA* 26 (4), 513–549. doi:10.1017/s027226310404001x

Eckman, F., and Iverson, G. (1993). Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Lang. Res.* 9, 234–252. doi:10.1177/026765839300900302

Eckman, F. (1985). Some theoretical and pedagogical implications of the markedness differential hypothesis. *Stud. Second Lang. Acquisition* 7, 289–307. doi:10.1017/S0272263100005544

Eckman, F. (2008). "Typological markedness and second language phonology," in *Phonology and second language acquisition*. Editors J. Hansen and M. Zampini (Amsterdam, NL: John Benjamins Publishing), 95–115.

Escudero, P. (2002). "The perception of English vowel contrasts: acoustic cue reliance in the development of new contrasts," in Proceedings of the 4th International Symposium on the Acquisition of Second-Language Speech, New Sounds, September, 2000.

Escudero, P., and Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Lang. Speech* 53 (3), 343–365. doi:10.1177/0023830910371447

Flege, J. E. (1995). "Second language speech learning: theory and findings and problems," in *Speech perception and linguistic experience: issues in cross-linguistic research*. Editor W. Strange (York, UK: York Press), 233–277.

Garcia, G. D. (2020). Language transfer and positional bias in English stress. *Second Lang. Res.* 36 (4), 445–474. doi:10.1177/0267658319882457

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia* 9, 317–323. doi:10.1016/0028-3932(71)90027-3

Guion, S., Flege, J. E., Liu, S., and Yeni-Komshian, G. (2000). Age of learning effects on the duration of sentences produced in a second language. *Appl. Psycholinguistics* 21, 205–228. doi:10.1017/s0142716400002034

Guion, S., and Pederson, E. (2007). "Investigating the role of attention in phonetic learning," in Second language speech learning: the role of language experience in speech perception and *production*. Editors O.-S. Bohn and M. Munro (Amsterdam, NL: John Benjamins Publishing), 99–116.

Hancin-Bhatt, B. (1994). Segment transfer: a consequence of a dynamic system. *Second Lang. Res.* 10 (3), 241–269. doi:10.1177/026765839401000304

Haspelmath, M. (2006). Against markedness (and what to replace it with). *J. Linguistics* 42, 25–70. doi:10.1017/s0022226705003683

Jesney, K. (2014). "A learning-based account of L1 vs. L2 cluster repair differences," in Selected Proceedings of the 5th Conference on generative Approaches to language acquisition north America (GALANA 2012), University of Kansas, October 11–13, 2012. Editors C. Y. Chu, C. E. Coughlin, B. L. Prego, U. Minai, and A. Tremblay (Somerville, MA: Cascadilla Proceedings Project), 10–21.

Kerek, A. (1971). *Hungarian metrics*. Bloomington, IN: Indiana University.

LaCharité, D., and Prévost, P. (1999). "The role of L1 and of teaching in the acquisition of English sounds by Francophones," in *Proceedings of BUCLD 23*. Editors A. Greenhill, H. Littlefield, and C. Tano (Somerville, MA: Cascadilla Press), 373–385.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor, MI: Michigan University Press.

Lahiri, A., and Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *J. Phonetics* 38, 44–59. doi:10.1016/j.wocn.2010.01.002

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Lang. Res.* 25 (2), 173–227. doi:10.1177/0267658308100283

Larson-Hall, J. (2004). Predicting perceptual success with segments: a test of Japanese speakers of Russian. *Second Lang. Res.* 20 (1), 33–76. doi:10.1191/0267658304sr230oa

Matthews, J. (2000). "The influence of pronunciation training on the perception of second language contrasts," in *New sounds 1999*. Editors J. Leather and A. James (University of Klagenfurt Press), 223–229.

Özçelik, Ö., and Sprouse, R. (2016). Emergent knowledge of a universal phonological principle in the L2 acquisition of vowel harmony in Turkish: a "four"-fold poverty of the stimulus in L2 acquisition. *Second Lang. Res.* 33 (2), 179–206. doi:10.1177/0267658316679226

Özçelik, Ö. (2016). The prosodic acquisition path hypothesis: towards explaining variability in L2 acquisition of phonology. *Glossa* 1 (1), 1–48. doi:10.5334/gjgl.47

O'Grady, W. (1996). Language acquisition without Universal Grammar: a general nativist proposal for L2 learning. *Second Lang. Res.* 12 (4), 364–397. doi:10.1177/026765839601200403

O'Grady, W. (2006). "The syntax of quantification in SLA: an emergentist approach," in Proceedings of the 8th generative Approaches to second language acquisition conference (GASLA 2006): the Banff conference, Banff, Canada, April 27–30, 2006. Editors M. O'Brien, C. Shea, and J. Archibald (Somerville, MA: Cascadilla Press), 98–113.

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

Parker, S. (2012). "Sonority distance vs. sonority dispersion – a typological survey," in *The sonority controversy*. Editors S. Parker and A. Lahiri (Berlin, Germany: De Gruyter Mouton), 101–166.

Pinker, S. (1989). *Language learnability and cognition: the acquisition of argument structure*. Cambridge, MA: MIT Press.

Schluter, K., Politzer-Ahles, S., Al Kaabi, M., and Almeida, D. (2017). Laryngeal features are phonetically abstract: mismatch negativity evidence from Arabic, English, and Russian. *Front. Psychol.* 8, 746. doi:10.3389/fpsyg.2017.00746

Schmidt, R. (1990). The role of consciousness in second language learning. *Appl. Linguistics* 11, 129–158. doi:10.1093/applin/11.2.129

Schwartz, M., and Goad, H. (2017). Indirect positive evidence in the acquisition of a subset grammar. *Lang. Acquisition* 24 (3), 234–264. doi:10.1080/10489223.2016.1187616

Tessier, A. M., Sorenson Duncan, T., and Paradis, J. (2013). Developmental trends and L1 effects in early L2 learners onset cluster production. *Bilingualism: Lang. Cogn.* 16 (3). doi:10.1017/s136672891200048x

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), 1–30. doi:10.1017/s0272263106060013

Truscott, J., and Sharwood Smith, M. (2004). Acquisition by processing: a modular perspective on language development. *Bilingualism: Lang. Cogn.* 7 (1), 1–20. doi:10.1017/s1366728904001178

Wang, Y., Jongman, A., and Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *J. Acoust. Soc. Am.* 113, 1033. doi:10.1121/1.1531176

Wexler, K., and Culicover, P. (1980). *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

White, L. (1991). Adverb placement in second language acquisition: some effects of positive and negative evidence in the classroom. *Second Lang. Res.* 7, 133–161. doi:10.1177/026765839100700205

Wilson, C., and Davidson, L. (2009). Bayesian analysis of non-native cluster production. *Proc. NELS* 40.

Wulff, S., and Ellis, C. N. (2018). "Usage-based approaches to second language acquisition," in *Bilingual cognition and language: the state of the science across its subfields*. Editors D. T. Miller, F. Bayram, J. Rothman, and L. Serratrice (Amsterdam, NL: John Benjamins Publishing), 37–56.

Young-Scholten, M. (1994). On positive evidence and ultimate attainment in L2 phonology. *Second Lang. Res.* 10, 193–214. doi:10.1177/026765839401000302

Young-Scholten, M. (2004). Prosodic constraints on allophonic distribution in adult L2 acquisition. *Int. J. Bilingualism* 8 (1), 67–77. doi:10.1177/13670069040080010501

Zerbian, S. (2015). "Markedness considerations in L2 prosodic focus and givenness marking," in *Prosody and language in contact: L2 acquisition, attrition and languages in multilingual situations*. Editors E. Delais-Roussarie, M. Avanzi, and S. Herment (Berlin, Germany: Springer), 7–27.

Conflict of Interest: 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.

frontiers in Communication

ORIGINAL RESEARCH
published: 06 April 2021
doi: 10.3389/fcomm.2021.627316

Exploring the Complexity of the L2 Intonation System: An Acoustic and Eye-Tracking Study

Di Liu[1]* *and Marnie Reed*[2]

[1] *Teaching and Learning Department, College of Education and Human Development, Temple University, Philadelphia, PA, United States,* [2] *Wheelock College of Education and Human Development, Boston University, Boston, MA, United States*

Phonological research has demonstrated that English intonation, variably referred to as prosody, is a multidimensional and multilayered system situated at the interface of information structure, morphosyntactic structure, phonological phenomena, and pragmatic functions. The structural and functional complexity of the intonational system, however, is largely under-addressed in L2 pronunciation teaching, leading to a lack of spontaneous use of intonation despite successful imitation in classrooms. Focusing on contrastive and implicational sentence stress, this study explored the complexity of the English intonation system by investigating how L1 English and Mandarin-English L2 speakers use multiple acoustic features (i.e., pitch range, pitch level, duration, and intensity) in signaling contrastive and implicational information and how one acoustic feature (maximum pitch level) is affected by information structure (contrast), morphosyntactic structure (phrasal boundary), and a phonological phenomenon (declination) in L1 English and Mandarin-English L2 speakers' speech. Using eye-tracking technology, we also investigated (1) L1 English and Mandarin-English L2 speakers' real-time processing of lexical items that carry information structure (i.e., contrast) and typically receive stress in L1 speakers' speech; (2) the influence of visual enhancement (italics and bold) on L1 English and Mandarin-English L2 speakers' processing of contrastive information; and (3) L1 English and Mandarin-English L2 speakers' processing of pictures with contrastive information. Statistical analysis using linear mixed-effects models showed that L1 English speakers and Mandarin-English L2 speakers differed in their use of acoustic cues in signaling contrastive and implicational information. They also differed in the use of maximum pitch level in signaling sentence stress influenced by contrast, phrasal boundary, and declination. We did not find differences in L1 English and Mandarin-English L2 speakers' processing of contrastive and implicational information at the sentence level, but the two groups of participants differ in their processing of contrastive information in passages and pictures. These results suggest that processing limitations may be the reason why L2 speakers did not use English intonation spontaneously. The findings of this study also suggest that Complexity Theory (CT), which emphasizes the complex and dynamic nature of intonation, is a theoretical framework that has the potential of bridging the gap between L2 phonology and L2 pronunciation teaching.

Keywords: intonation, sentence stress, complexity theory, pronunciation teaching, applied linguistics, linear mixed effect model, L2 Phonology, eye-tracking

OPEN ACCESS

Edited by:
Mary Grantham O'Brien,
University of Calgary, Canada

Reviewed by:
Geoffrey Schwartz,
Adam Mickiewicz University, Poland
Joan Carles Mora,
University of Barcelona, Spain

****Correspondence:***
Di Liu
diliu@temple.edu

Specialty section:
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: *09 November 2020*
Accepted: *08 March 2021*
Published: *06 April 2021*

Citation:
Liu D and Reed M (2021) Exploring the Complexity of the L2 Intonation System: An Acoustic and Eye-Tracking Study. Front. Commun. 6:627316. doi: 10.3389/fcomm.2021.627316

INTRODUCTION

The past 30 years have witnessed evolutionary and far-reaching advances in the field of L2 pronunciation, including establishment of the intelligibility principle (Munro and Derwing, 1995; Murphy, 2014), development of a holistic approach acknowledging the importance of both segmental and suprasegmental features (Anderson-Hsieh et al., 1992; Derwing et al., 1998; Levis and Levis, 2018), and development of technologies such as speech visualization (Levis and Pickering, 2004), automated speech recognition (ASR) (Cucchiarini and Strik, 2018) and speech synthesis (Ding et al., 2019).

Despite significant advances, challenges remain. Levis (1999), for example, stated that "[p]resent intonational research is almost completely divorced from modern language teaching…" (p. 37). The lack of a guiding theory causes many issues. For example, teachers may only focus on the aspects that they are conscious of using, resulting in an overemphasis on the attitudinal and emotional aspects in the teaching of suprasegmental features (Levis, 1999). Further, while teachers may assume that successful imitation and reproduction of target intonation patterns in classrooms leads to spontaneous production, students "may walk out of the class without having accepted the system at all. Or they may think intonation is simply decorative" (Gilbert, 2014). Gilbert's observation echoes what Allen (1971) had noted half a century ago: "there is little carry-over into the students' own conversations outside the classroom and the listen and repeat approach has never yielded satisfactory long-term results" (p. 79).

One main issue that sets research and teaching apart is the lack of applicability of research theories and models. In the past few decades, the field of L2 phonology has seen a number of highly influential theories, approaches and models including Autosegmental-Metrical phonology (AM) (Pierrehumbert, 1980), the Systemic Functional Approach (Halliday, 2015), and Discourse Intonation (Brazil, 1980) as well as the PENTA model (Xu, 2004), the Kiel intonation model (KIM) (Kohler, 1995), and the Fujisaki model (Mixdorff, 2000). While these theories and models have gained wide popularity in research and software development, direct transfer of these theories and models into classroom teaching has faced tremendous difficulties. For example, in his attempt to apply Discourse Intonation in language teaching, Chapman (2007) found that both students and teachers encountered difficulties such as identifying rising and falling tone as well as locating tone-unit boundaries and prominence. The core of this issue is that L2 phonology and L2 pronunciation teaching, albeit sharing the same underlying subject of investigation, focus on different issues and have different goals. L2 phonology analyzes speech samples in an effort to develop a theory or a model that explains underlying schema of speech production. L2 pronunciation teaching, on the other hand, focuses on the development of learners' abilities in navigating the system of intonation in spontaneous speech.

Larsen-Freeman (2017) pointed out that the field of second language acquisition is dominated by an approach which "seeks to understand phenomena by taking them apart" (Larsen-Freeman, 2017, p. 22). However, it could be the case that "it is from the components and their relationships that the system we are trying to understand emerges. If we isolate components artificially, we lose the essence of the phenomena we are attempting to describe" (Larsen-Freeman, 2017, p. 29). To promote L2 speakers' spontaneous use of intonation, a systematic view of intonation is needed. Complexity Theory (CT), which views the relationship among phonological features and phenomena as an interrelated and dynamic system, thus is a more appropriate theoretical framework for L2 pronunciation teaching.

LITERATURE REVIEW

Intonation

Scholars have long acknowledged that English intonation relates to multiple acoustic features and perceptual phenomena. Palmer (1922), for example, stated that "all phenomena connected with this musical pitch or tone [such as word-prominence, word-group prominence, intensity, command, doubt, concession, reassurance, etc.] are designated by the term Intonation" (p. 7). Halliday (2015) argued that intonation is a system with three phonological and systemic variables: tonality, tonicity, and tone. Pierrehumbert (1980) views the intonation system as having three components—a grammar of phrasal tunes, a metrical representation of the text, and rules that line up tune with the text. Over the past century, numerous scholars defined intonation from different perspectives and with different focuses. Some scholars focused on the pragmatic meaning and phrasal structure (Gussenhoven, 2004; Levis and Wichmann, 2015), others emphasized pitch patterns or tones (Kingdon, 1958; O'Connnor and Arnold, 1973), still others highlighted the emotional and attitudinal aspects (Bolinger, 1989) (see **Table 1**).

Despite differences in approaches and focuses, scholars generally agree that the system of intonation includes multiple suprasegmental features and is closely related to information structure, morphosyntactic structure, and pragmatic functions (Gilbert, 2014; Levis and Wichmann, 2015). Gilbert (2014), for example, stated that "[i]n English, prosodic cues serve as navigation guides to help the listener follow the intentions of the speaker. These signals communicate emphasis and make clear the relationship between ideas (new and old information) so that listeners can readily identify these relationships and understand the speaker's meaning" (p. 123). Wennerstrom (1998) proposed that there is an intonation system in English that functions at the discourse level to signal relationships in information structure and to mark interdependencies among constituents; she proposes a model in which intonation functions as a grammar of cohesion. These studies pointed out the importance of intonation and the necessity of teaching English intonation to L2 English speakers.

Sentence Stress

One essential component of English intonation is sentence stress (Kingdon, 1958; Pierrehumbert, 1980; Pickering, 2018), which is also commonly referred to as *prominence* (Celce-Murcia et al., 2010), *pitch accent* (Pierrehumbert, 1980; Pierrehumbert and Hirschberg, 1990), *focus* (Grant and Yu, 2017), *nucleus* (Palmer, 1922; Cruttenden, 1990; Cruz-Ferreira, 1998; Wells, 2006), and

TABLE 1 | Scholars' definitions of intonation and the suprasegmental features included.

Scholars	Definition	Suprasegmental features	Related variables
Ladd (2008)	"The use of *suprasegmental* phonetic features to convey 'postlexical' or *sentence-level* pragmatic meanings in a *linguistically structured* way" (p. 4).	Suprasegmental phonetic features	Pragmatic meanings, linguistic structure
Pickering (2018)	"The term *intonation* is narrowly defined in English as the use of pitch structure over the length of a given utterance" (p. 2). "Intonation is the grammatical system that includes our use of pitch, pause, and prominence (or sentence stress)…" (p. 3).	Pitch, pause, prominence	
O'Connnor and Arnold (1973)	"When we talk about English intonation we mean the pitch patterns of spoken English, the speech tunes or melodies, the musical features of English" (p. 1).	Pitch, tunes/melody, musical features	
Kingdon (1958)	"The active elements of intonation are the Tones, which always occur in association with stresses" (p. 3).	Tones, stress	
Levis and Wichmann (2015)	"The use of pitch variations in the voice to communicate phrasing and discourse meaning in varied linguistic environments" (p. 139).	Pitch	Phrasing and discourse meaning
Gussenhoven (2004)	"Intonation is treated as the use of phonological tone for non-lexical purposes, or—to put it positively—for the expression of phrasal structure and discourse meaning" (p. 12).	Phonological tone	Phrasal structure and discourse meaning
Bolinger (1989)	"Intonation manages to do what it does by continuing to be what it is, primarily a symptom of how we feel about what we say, or how we feel when we say" (p. 1).		Emotions
Palmer (1922)	"All phenomena connected with this musical pitch or tone [such as word-prominence, word-group prominence, intensity, command, doubt, concession, reassurance, etc.] are designated by the term Intonation" (p. 7).	Pitch/tone, prominence, intensity	Attitudes

primary stress (Hahn, 2004; Celce-Murcia et al., 2010). Sentence stress denotes the relative emphasis or prominence a word receives primarily by the manipulation of fundamental frequency (F0), duration, and intensity of the stressed syllable as well as the modification of vowel quality.

Sentence stress plays a central role in English intonation because of its close connection with discourse meaning and information structure. It is frequently used to signal given vs. new (Pierrehumbert and Hirschberg, 1990) and contrast (Liu, 2020). It is also commonly used to make corrections (Ip and Cutler, 2016) or help listeners to anticipate what the speaker is going to say (Levis and Levis, 2018).

Sentence stress affects intelligibility, which is defined as "the extent to which a listener actually understands an utterance" (Derwing and Munro, 2005, p. 385). For example, investigating L1 English speakers' processing, comprehension, and evaluation of speech with correctly placed, misplaced, and missing sentence stress, Hahn (2004) found that the same speaker is more intelligible when sentence stress was used correctly. One of the multiple functions of intonation that impacts intelligibility is sentence stress used to show contrast. Levis and Levis (2018) argued that contrastive stress is a "high-value pronunciation feature" that should be given more attention in L2 pronunciation teaching.

L2 Sentence Stress

Prior studies investigating Mandarin-English L2 speakers' intonation showed that even advanced level speakers face difficulties in using English sentence stress effectively (Chun, 1982; Wennerstrom, 1998; Pickering, 2001, 2004). This issue relates to both sentence stress realization and placement. Pickering (2001), for example, found that the pitch structure of Chinese international teaching assistants' speech is relatively flat and monotone compared to native speaker teaching assistants. Chun (1982) found that "Chinese speakers sometimes failed to place sentence stress on the appropriate word or syllable" (p. 386), pointing out the issue of stress placement.

It was assumed by some scholars that Mandarin-English L2 speakers' ineffective use of English intonation is due to the lexical tone system in Mandarin (Clennell, 1997). The claim is that Mandarin uses pitch variation to signal lexical tone, prohibiting it from using the same acoustic feature to indicate sentence stress. However, recent studies investigating the Mandarin sentence stress system suggested that this may be a false assumption (Xu, 1999; Chen and Gussenhoven, 2008; Kabagema-Bilan et al., 2011). Ouyang and Kaiser (2015), for instance, found that all three suprasegmental features that English uses to indicate sentence stress (fundamental frequency (F0), duration, and intensity) are all used to encode sentence stress in Mandarin. They further concluded that Mandarin uses sentence stress to indicate discourse-level information and make contrasts. Analyzing five types of focus in Mandarin and English, Ip and Cutler (2016) found that Mandarin speakers showed greater increase in pitch range and pitch level for new-information focus.

Despite similarities between English and Mandarin sentence stress, there is evidence of lack of transfer of suprasegmental features from L1 Mandarin to L2 English. For example, comparing L1 English speakers' use of sentence stress to Mandarin-English L2 speakers' sentence stress in both English and Mandarin, Liu (2020) found that Mandarin-English L2 speakers did not use pitch to indicate English sentence stress even though they resemble L1 English speakers in the signaling of sentence stress when speaking in L1 Mandarin.

It is worth noticing that the lack of transfer of similar phonological features from L1 to L2 is not a language specific issue. For example, investigating the use of prosodic cues in German L2 English speakers' English and German speech, O'Brien et al. (2014) found that German-English L2 speakers do not transfer all prosodic uses from L1 to L2. The findings suggested that even speakers of Germanic languages that share similar morphosyntactic structure and uses of phonological cues with English do not transfer intonational cues directly from L1 to L2.

One factor that may account for the lack of transfer of phonological cues from L1 to L2 is the complex and dynamic nature of language. As Ortega-Llebaria and Colantoni (2014) stated, "…acquiring intonation in a L2 not only is an issue of learning to perceive and produce the target melody but, crucially, involves a new mapping between form and meaning that is affected by L1 transfer" (p. 351). The fact that intonation involves multiple acoustic cues and needs to be used with consideration of the information structure, morphosyntactic structure, and phonological phenomena may pose a significant challenge to L2 speakers. Complexity Theory (CT), which views the system of intonation as complex and dynamic, thus will be informative in L2 phonology research and L2 pronunciation teaching.

Complexity Theory

Language is a complex dynamic system that "emerges bottom-up from interactions of multiple agents in speech communities" (Larsen-Freeman, 2017, p. 49). Language is also a social endeavor constantly influenced and shaped by the interaction and accommodation among different individuals, speech communities, and communities of practice (CoP) (Gumperz, 1971; Labov, 1972; Giles et al., 1987; Lippi-Green, 1989; Holmes and Meyerhoff, 1999). In the system of intonation, two types of complexity have been identified: structural complexity and functional complexity.

Structural complexity, also referred to as inherent or absolute complexity (Housen et al., 2019) captures the intrinsic complexity of a system. Encompassing multiple interrelated and interacting variables, the system of English intonation involves structural complexity. For example, when a speaker stresses a constituent, multiple segmental and suprasegmental features (i.e., vowel quality, pitch, duration, and intensity) are manipulated. The changes in these features affect not only the use of prosodic features at the syllable and word level, but also the intonational contour of the entire intonational phrase.

Another level of complexity is derived from the dynamic relationship between intonation and other variables such as information structure, morphosyntactic structure, and phonological phenomena. We list six examples of this functional complexity. Intonation dynamically encodes information structure (e.g., new vs. old, contrast, etc.) (Pierrehumbert and Hirschberg, 1990; Hahn, 2004; Levis and Wichmann, 2015), indicates grammatical structure (Pickering, 2018), signals discourse level meanings and implications (Levis and Wichmann, 2015; Pickering, 2018), regulates speaker-listener interactions (Wennerstrom, 2001; Hellermann, 2003), directs listeners' attention (Chun, 1988; Gilbert, 2014), and expresses emotions and attitudes (Horley et al., 2010; Pell and Kotz, 2011). Functional complexity poses challenges to L2 speakers because not only can intonation be optionally used for these functions, it is also governed and dynamically shaped by all these aspects. In this sense, using L2 intonation is not simply manipulating a collection of acoustic cues, but navigating the entire linguistic repertoire while representing and balancing the influence of numerous linguistic and non-linguistic variables in real-time.

The Present Study

The present study explores the structural and functional complexity of the intonation system by comparing L1 English and Mandarin-English L2 speakers' use of acoustic cues in speech production and visual processing of contrastive and implicational information that typically receives stress in L1 English speakers' speech. The present study both serves as an example for research investigating intonation from a CT perspective and offers insights into Mandarin-English L2 speakers' use of English intonation and the dynamic mapping between L1 and L2.

In their discussion of a developing cognitive system, DeBot et al. (2007) asserted, "the system is in constant complex interaction with its environment and internal sources. Its multiple interacting components produce one or many self-organized equilibrium points, whose form and stability depend on the system's constraints" (p. 14). As applied to L2 phonological phenomena, sentence stress, situated at the nexus of information structure, morphosyntactic structure, and other linguistic and non-linguistic variables, is appropriate for investigation. Treating every speech sample as an equilibrium point, acoustic analysis reveals information about how phonological features are used. From a Complexity Theory (CT) perspective, the present study investigated L1 English and Mandarin-English L2 speakers' use of various acoustic features in signaling English sentence stress and the influence of information structure, morpho-syntactical structure, and phonological phenomena.

Eye-tracking technology has been used widely in the field of second language research as a means to investigate L1 and L2 speakers' parsing of temporarily ambiguous sentences (Papadopoulou and Clahsen, 2003; Dussias and Sagarra, 2007); processing of lexical and morphological cues (Kambe et al., 2001; Lew-Williams and Fernald, 2010), and processing of spoken language (Tanenhaus, 2007; Ito and Speer, 2008). Researchers found "systematic relations between fixation duration and the characteristic of the fixated words" (Dussias, 2010, p. 150), providing information about the incremental processing of sentence comprehension. Using eye-tracking technology, this study explored real-time processing of lexical items (1) that are contrastive or non-contrastive, (2) at different positions within written sentences, (3) that do or do not carry implications, and (4) that appear with or without visual enhancement. The guiding research questions are:

- How do L1 and Mandarin-English L2 speakers use intonational features (i.e., pitch range, pitch level, duration, and intensity) to signal contrast and implication in speech production?
- How do L1 and Mandarin-English L2 speakers use a phonological feature (maximum pitch level) influenced by information structure (contrast), morphosyntactic structure (sentence boundaries), and phonological phenomena (declination)?
- How do L1 and Mandarin-English L2 speakers orally produce and visually process lexical items and pictorial information that are contrastive or implicational and that typically receive stress in L1 speakers' speech?
- How do L1 and Mandarin-English L2 speakers orally produce and visually process contrastive information written with and without visual enhancement (italics and bold)?

METHODS

Participants

Ten subjects participated in the study. Five were L1 English speakers and five were Mandarin-English L2 speakers enrolled in degree programs in a university in the US. Participants' biographical information is summarized in **Tables 2**, **3**.

Procedure

The study was conducted in a Language Acquisition and Visual Attention lab at a large research university in the northeast. Participants were seated in front of a PC connected to an *Eyelink 1000 Plus* eye-tracker. All participants went through a calibration process. Then, the participants were presented a scenario with pictures contextualizing the first experiment.

In the first experiment, participants saw 18 sentences on the computer screen in random order. Participants saw one sentence at a time and each sentence was presented in a single line. Participants were asked to first read the sentence silently. Then, when they were ready, they read the sentence aloud. In the second experiment, the participants saw three sets of sentences in random order. Each set of sentences contains two sentences of the same wording, but different meanings or implications presented in parentheses. Participants were asked to first read silently and then read aloud the sentences to express the meanings or implication included in the parentheses. In the third experiment, participants read silently and then read aloud two short passages with contrastive information. In the fourth experiment, participants saw an eight-frame picture cartoon that describes a single story. They were asked to look at the pictures silently and prepare to tell the story. Then, they were asked to tell the story in their own words. The experiment materials and data analysis are further discussed within each experiment in the following sections.

While participants silently read the sentences and passages (in experiments I—III) and viewed the pictures (in experiment IV), the eye-tracker documented the fixation count, fixation percentage, dwell time, dwell percentage, run count, and regression of the Areas of Interest (AOI), which were the contrastive/implicational information or equivalent places in the distractors. The eye-tracker was recalibrated before each experiment. When the participants read aloud the information, they were audio-recorded using both an audio recording application software (*Audacity*) and a handheld recorder (*Zoom H4N*).

Data Elicitation and Analysis

Participants' fixation count (total number of fixations within the interest area), fixation percentage (percentage of total fixations in a trial falling within the current interest area), dwell time (total time (in milliseconds) spent on the current interest area), dwell percentage (percentage of trial dwell time spent on the current interest area), run count (number of times the interest area was entered and left), and regression on the focused areas were documented using the eye-tracker (definitions elicited from *EyeLink Data Viewer User's Manual*, Version 1.11.900, p. 39–40). Participants' speech data were analyzed using *Praat* version 6.0.37 (Boersma and Weenink, 2018). Maximum pitch level, minimum pitch level, and pitch range in Hertz and in semitones relative to 1 Hz, as well as the duration of words in seconds and the intensity of words in decibels (dBs) were elicited using a *Praat* script. Participants' pitch in Hertz and pitch in semitones were normalized based on each participant's average pitch level. We used *R* (R Core Team, 2017) and *lme4* (Bates et al., 2015) to analyze the speech and processing data. Linear mixed-effects models were constructed for the statistical analyses; *P*-values were obtained by likelihood ratio tests.

EXPERIMENT 1

In Experiment 1, we investigated L1 and Mandarin-English L2 speakers' production and processing of sentence stress using a contextualized sentence read aloud task. The scenario we used was a Christmas Tree decoration task adapted from Ito and Speer (2008). The participants were told that a friend of theirs, Martin, is decorating a Christmas tree using items from an ornament board. Martin is trying to select two of the ornaments at a time and the participants were tasked with telling him which ones to hang. In the contextualizing process, the participants were asked to tell Martin what to hang based on circled items on pictures of the ornament board. Then, after participants understood the scenario, they were asked to complete the task using 18 written sentence prompts.

The sentences belong to six sentence sets. Each set has three sentences: one sentence with contrastive information presented not at phrasal boundaries, one sentence with contrastive information presented at phrasal boundaries, and a distractor that does not include contrastive information (see **Appendix A** for a complete list of sentences).

- First, hang the blue drum, then hang the yellow drum ("blue" and "yellow" are contrastive and not at phrasal boundaries).
- First, hang the blue drum, then hang the blue ball ("drum" and "ball" are contrastive and at sentence phrasal boundaries).
- First, hang the blue drum, then hang the pink ball (distractor: no contrastive information in the sentence).

TABLE 2 | L1 English speakers' biographical information.

Gender	Age	L1	Other language(s)	Major	Degree
Female	21	English	Spanish (advanced); French (intermediate)	Neuroscience	Bachelor
Male	24	English	Amharic (bilingual)	Biomedical engineering	Bachelor
Female	21	English	ASL (intermediate)	Psychology	Bachelor
Female	20	English	Mandarin (intermediate)	International relations, linguistics	Bachelor
Female	21	English	Spanish (beginner)	Health science	Bachelor

TABLE 3 | Mandarin-English L2 speakers' biographical information[*].

Gender	Age	L1	Other language(s)/years of learning	Major	Degree	Standardized Test Scores	Speaking	Listening	Percentage of English use daily
Male	28	Mandarin	English (10 years)	Computer Science	Master	TOEFL 100	TOEFL 18–25	TOEFL 22–30	20–40%
Male	24	Mandarin	English (10 years)	Advertising	Bachelor	TOEFL 109	TOEFL 18–25	TOEFL 22–30	20–40%
Male	26	Mandarin	English (15 years)	System Science	Bachelor	IELTS 6.5	IELTS 5.5	IELTS 5.5	0–20%
Female	24	Mandarin	English (17 years)	Advertising	Master	TOEFL 111	18–25	22–30	20–40%
Female	20	Mandarin	English (15 years), Japanese (intermediate)	Biology	Bachelor	TOEFL 100	18–25	22–30	60–80%

[*]A total of 6 L1 English speakers and 11 Mandarin-English L2 speakers were recruited to participate in the study. 1 L1 English speaker and 6 Mandarin-English L2 speakers were excluded from the dataset because of technological/calibration failure primarily due to the reflection of participants' eyeglasses.

TABLE 4 | Sample sentences with contrastive and non-contrastive information at different locations.

			Position 1	Position 2				Position 3	Position 4
1. First,	hang	the	***Blue*** Contrastive + not at boundary	***Drum,*** Non-contrastive + boundary	then	hang	the	***Yellow*** Contrastive + not at boundary	***Drum*** Non-contrastive + boundary
2. First,	hang	the	***Blue*** Non-contrastive + not at boundary	***Drum,*** Contrastive + boundary	then	hang	the	***Blue*** Non-contrastive + not at boundary	***Ball*** Contrastive + boundary
3. First,	hand	the	***Blue*** Non-contrastive + not at boundary	***Drum,*** Non-contrastive + boundary	then	hang	the	***Pink*** Non-contrastive + not at boundary	***Ball*** Non-contrastive + boundary

All sentences were scrambled and presented to the participants in random order. For each sentence, the participants were asked to read the sentence silently to themselves first, and then produce the sentence as if they were providing Martin directions on which items to hang on the Christmas tree.

Our first question was:

(1) How do L1 and Mandarin-English L2 speakers use intonational features (i.e., pitch range, pitch level, duration, and intensity) to signal contrast and implication in speech production?

To answer this question, we analyzed speech data using linear mixed-effects models that predicted normalized pitch range, normalized maximum pitch level, duration, and intensity. We used different L1s, contrastive or non-contrastive, and the interaction between L1 and contrast as the fixed effects. Individual participants, sentences, and words were included in the model as the random effects. The results showed that in L1 English speakers' speech, there were statistically significant differences between the contrastive and non-contrastive information regarding the use of pitch range (estimate = 1.136, SE = 0.107, df = 445.824, t = 10.609, $^{***}p < 0.0001$), maximum pitch level (estimate = 0.77, SE = 0.111, df = 455.753, t = 6.932, $^{***}p < 0.0001$), and duration (estimate = 0.112, SE = 0.012, df = 460.980, t = 9.284, $^{***}p < 0.0001$). However, there was no significant difference in the intensity of contrastive and non-contrastive information (estimate = 0.243, SE = 0.382, df = 457.689, t = 0.635, $p = 0.526$).

In Mandarin-English L2 English speakers' speech, however, there were no significant differences between the contrastive and non-contrastive information in terms of pitch range (estimate = −0.137, SE = 0.107, df = 443.794, t = −1.272, $p = 0.204$), maximum pitch level (estimate = −0.0247, SE = 0.111, df = 453.505, t = −0.222, $p = 0.824$), or duration (estimate = −0.008, SE = 0.012, df = 461.264, t = −0.679, $p = 0.498$). However, there was a significant difference between the contrastive and non-contrastive information signaled by intensity in Mandarin-English L2 speakers' speech (estimate = 1.092, SE = 0.383, df = 458.084, t = 2.853, $^{*}p = 0.005$) (see **Figure 1**).

The results showed differences in the acoustic cues that L1 and Mandarin-English L2 speakers used to signal contrastive information. L1 speakers used pitch and duration to indicate contrastive stress. Mandarin-English L2 speakers, on the contrary, did not signal contrast using pitch or duration. Their use of greater intensity may have been intended to signal contrastive information.

Our second research question was:

(2) How do L1 and Mandarin-English L2 speakers use a phonological feature (maximum pitch level) influenced by information structure (contrast), morphosyntactic structure (sentence boundaries), and phonological phenomena (declination)?

To answer question (2), we analyzed participants' use of maximum pitch level of contrastive and non-contrastive information at different positions within sentences (see **Table 4**). Specifically, we explored how L1 and Mandarin-English L2 speakers' pitch level is affected by (1) information structure (contrastive vs. non-contrastive), (2) morphosyntactic structure (at sentence boundary vs. not at sentence boundary), and (3) phonological phenomena (declination). We used the position of the words, contrastive or non-contrastive and the interaction between position and contrast as the fixed effects to predict normalized maximum pitch level. Individual participants, sentences, and words were entered into the model as random effects.

In L1 English speakers' speech, there was a significant difference between the maximum pitch level of the contrastive and non-contrastive information (estimate = 0.632, SE = 0.255, df = 222.926, t = 2.476, $^{*}p = 0.014$). We also found differences among positions. Compared to the words in position 1, words in position 3 had significantly lower pitch level (estimate = −0.882,

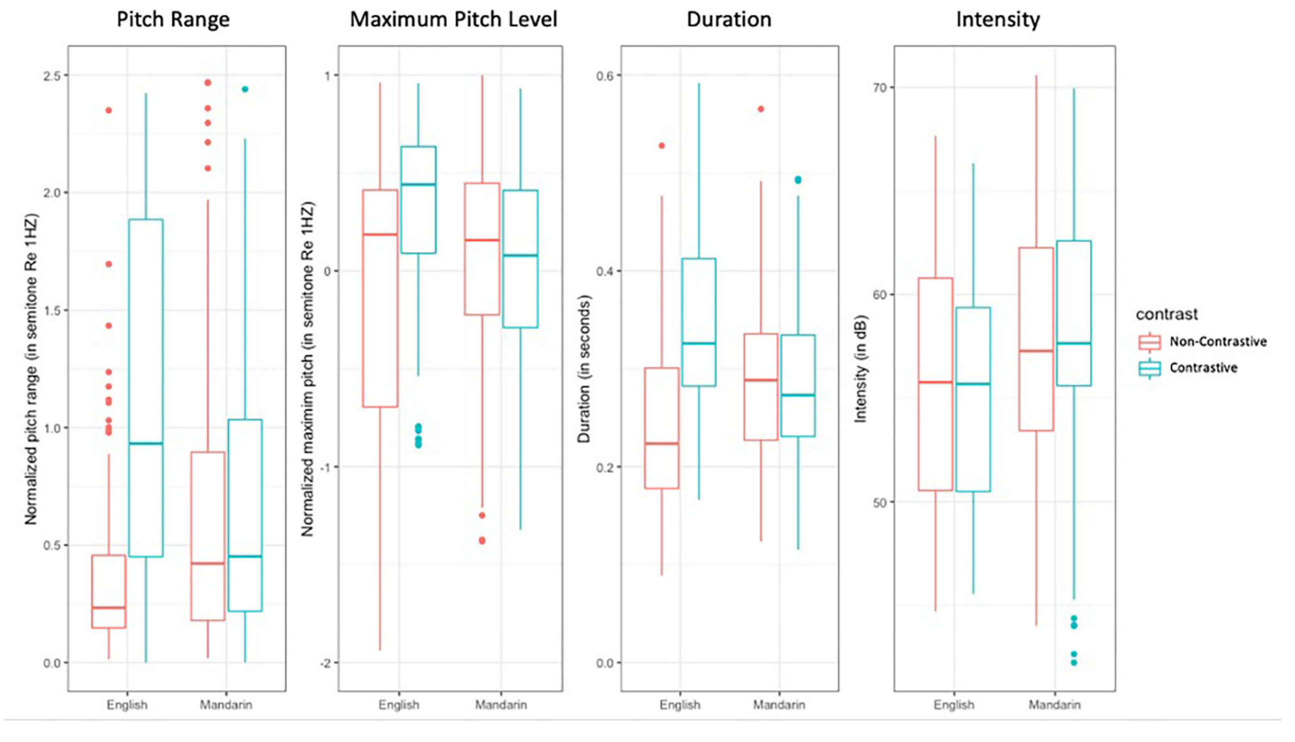


FIGURE 1 | Speech production of L1 English and Mandarin-English L2 Speakers contrastive and non-contrastive stress.

SE = 0.255, df = 222.926, t = 3.457, ***p = 0.0006). There was also a significant difference between words in position 1 and words in position 4 (estimate = −1.148, SE = 0.262, df = 223.083, t = −4.379, ***p < 0.0001). The position differences suggested the effect of declination. There were no significant differences between words in position 1 and words in position 2 (estimate = 0.256, SE = 0.255, df = −222.926, t = 1.005, p = 0.316), suggesting an influence of a H- (high) boundary tone at phonological boundary (Pierrehumbert and Hirschberg, 1990) used to signal non-finality.

In Mandarin-English L2 English speakers' speech, however, there was no significant difference between contrastive and non-contrastive information (estimate = −0.067, SE = 0.141, df = 53.646, t = −0.478, p = 0.635). There was a significant difference between words in position 1 and words in position 3 (estimate = −0.3646, SE = 0.128, df = 202.882, t = −2.842, **p = 0.00493), suggesting potential influence of declination. However, there was no significant difference between words in position 1 and words in position 4 (estimate = −0.251, SE = 0.153, df = 41.98, t = −1.632, p = 0.1101). There was also no significant difference between words in position 1 and words in position 2 (estimate = −0.0345, SE = 0.149, df = 38.088601, t = −0.232, p = 0.81805) (see **Figure 2**).

These results showed that in L1 English speakers' speech, one individual intonational cue—maximum pitch level—is affected by multiple variables including information structure (i.e., contrast), morphosyntactic structure (i.e., phrasal boundary), and phonological phenomena (i.e., declination). Although Mandarin-English L2 speakers' maximum pitch level reflected potential influence of declination, L2 speakers did not show schema that reflect functional complexity at a level comparable to L1 speakers.

The third and fourth research questions were:

(3) How do L1 and Mandarin-English L2 speakers orally produce and visually process lexical items and pictorial information that are contrastive or implicational and that typically receive stress in L1 speakers' speech?

(4) How do L1 and Mandarin-English L2 speakers orally produce and visually process contrastive information written with and without visual enhancement (italics and bold)?

To answer questions (3) and (4), we analyzed the subset of sentences with visual enhancement (i.e., bold and italics) signaling the contrastive information (two sets with italics and two sets with bold). We hypothesized that the orthographical conventions used in English may have different implications or functions in the Chinese logographic writing system (Mair, 1996, p. 200). These differences may affect the processing of written information and the production of contrastive information.

Sample sentence set with italicized words

- First, hang the *yellow* tree, then hang the *white* tree ("yellow" and "white" are contrastive, signaled by italics, and not at sentence boundaries).
- First, hang the yellow *tree*, then hang the yellow *star* ("tree" and "star" are contrastive, signaled by italics, and at sentence boundaries).
- First, hang the yellow tree, then hang the white bell (no contrastive information, no visual enhancement).

Sample sentences set with words in bold

- First, hang the **green** egg, then hang the **brown** egg ("green," and "brown" are contrastive, signaled by bold, and not at sentence boundaries).
- First, hang the green **egg**, then hang the green **sock** ("egg," and "sock" are contrastive, signaled by bold, and at sentence boundaries).
- First, hang the green egg, then hang the orange tree (no contrastive information, no visual enhancement).

Participants' pitch range was analyzed using L1, visual enhancement (no visual enhancement, italics, and bold), and contrast (contrastive vs. non-contrastive) as the fixed effect and individual participants, sentences and words as the random effects. The results show that L1 English speakers used a significantly greater pitch range to signal contrastive information regardless of the use of visual enhancement (estimate = 1.04, SE = 0.2, df = 216.14, t = 5.09, ***p < 0.0001). Mandarin-English L2 speakers, on the other hand, did not use pitch range to signal contrastive information even when the information was enhanced by bold or italics (estimate = −0.1, SE = 0.17, df = 223.56, t = −0.62, p = 0.54) (see **Figure 3**).

We then investigated whether there were processing differences indicated by the fixation percentage and dwell percentage of the AOI by using L1, visual enhancement, and the interaction between these two factors as the fixed effects, and individual participants, sentences, and words as the random effects in our models. The results suggest that there were no statistically significant differences between L1 and Mandarin-English L2 speakers in fixation percentage (estimate = 2.31, SE = 2.26, df = 20.19, t = 1.024, p = 0.32) or dwell percentage (estimate = 2.82, SE = 2.32, df = 25.75, t = 1.217, p = 0.235). Also, whether the contrastive information was visually enhanced by italics or bold did not lead to significant processing differences.

EXPERIMENT 2

The second experiment is a read aloud task. Participants were given three sets of sentences (adapted from *POSE*-test). For each set of sentences, two different implications were given in parentheses and could be signaled by altering the stressed constituents in the same sentence (for a complete list of sentences used in Experiment 2, refer to **Appendix B**).

- My brother is a doctor (not my sister).
- My brother is a doctor (not a teacher).

The participants were asked to read the sentences silently first, and then read the sentence out loud to express the implications in the parentheses. When they were reading the sentences silently, their fixation and dwell time and percentage were measured using an eye-tracker. When they were reading aloud, participants were asked not to read the information in the parentheses, and the

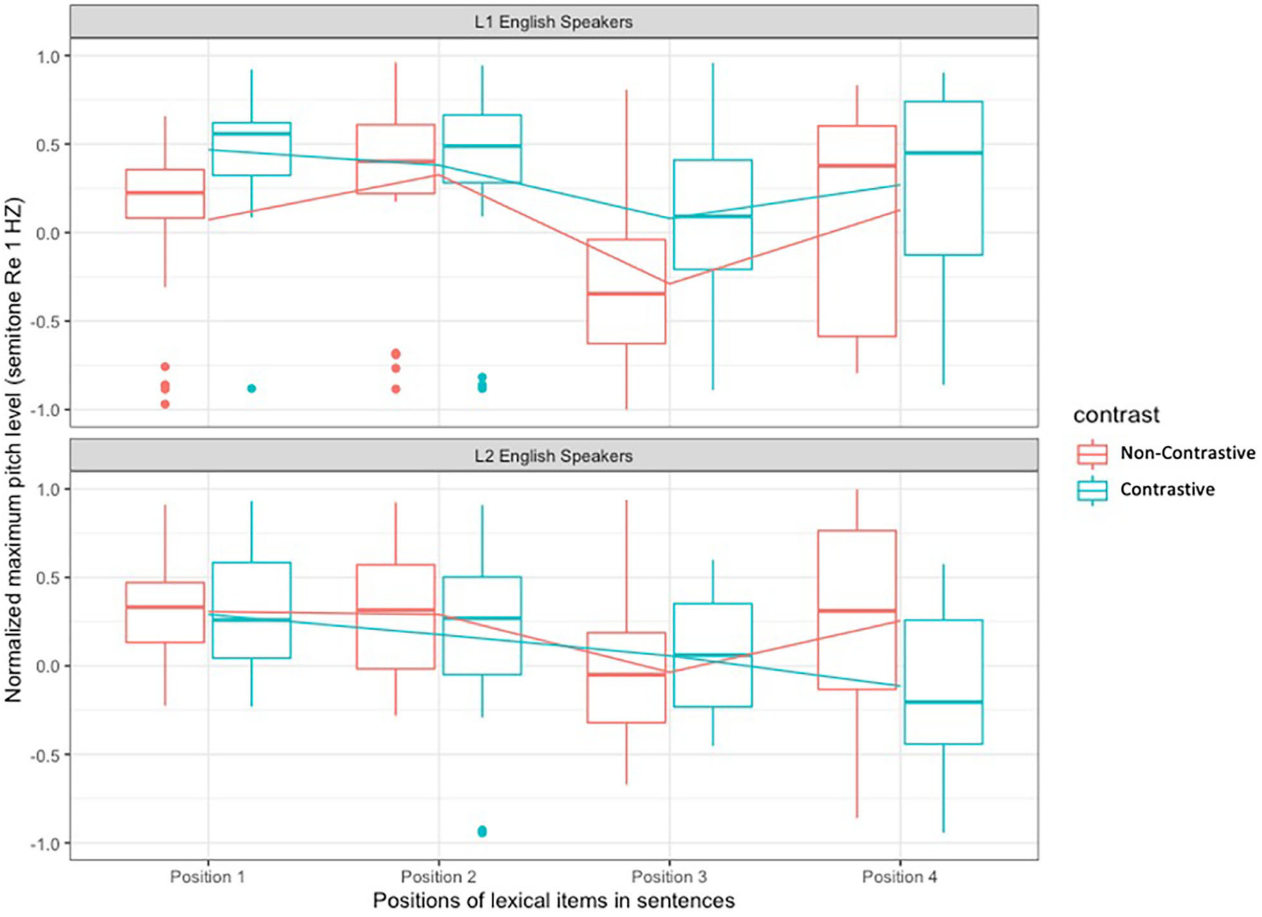


FIGURE 2 | Maximum pitch level of L1 English speakers' and Mandarin-English L2 speakers' contrastive and non-contrastive information at different positions.

sentences were recorded. The research questions we asked in experiment 2 were:

(1) How do L1 and Mandarin-English L2 speakers use intonational features (i.e., pitch range, pitch level, duration, and intensity) to signal contrast and implication in speech production?

(3) How do L1 and Mandarin-English L2 speakers orally produce and visually process lexical items and pictorial information that are contrastive or implicational and that typically receive stress in L1 speakers' speech?

Linear mixed-effects models were constructed with L1s, different implications, and the interaction between L1 and implication as the fixed effects, and individual participants, sentences, and words as the random effects. The result shows that L1 speakers use statistically significant greater pitch range (estimate = 1.2515, SE = 0.2687, df = 101.0587, t = 4.658, $^{****}p < 0.0001$), higher maximum pitch level (estimate = 0.61312, SE = 0.20432, df = 100.838, t = 3.001, $^{**}p$ = 0.003), and longer duration (estimate = 0.09075, SE = 0.02539, df = 102.385, t = 3.574, $^{***}p$ = 0.0005) to express the information related to the implication. There was no significant difference between the intensity that L1 speakers used to encode implicational or non-implicational information. For Mandarin-English L2 speakers, there were no significant differences between implicational vs. non-implicational conditions for all prosodic features we analyzed.

In terms of the processing of implicational information, we established linear fixed effects models with language as the fixed effect to predict both the fixation percentage and the dwell percentage. The results showed that there were no significant differences between L1 and Mandarin-English L2 speakers' processing of the information related to the implication as indicated by their fixation percentage (estimate = 1.880, SE = 3.998, df = 7.895, t = 0.470, p = 0.6508) and dwell percentage (estimate = 3.344, SE = 4.920, df = 7.918, t = 0.680, p = 0.516).

EXPERIMENT 3

We used two passage read-aloud tasks to further investigate L1 and Mandarin-English L2 speakers' production and processing of contrastive information with and without visual enhancement

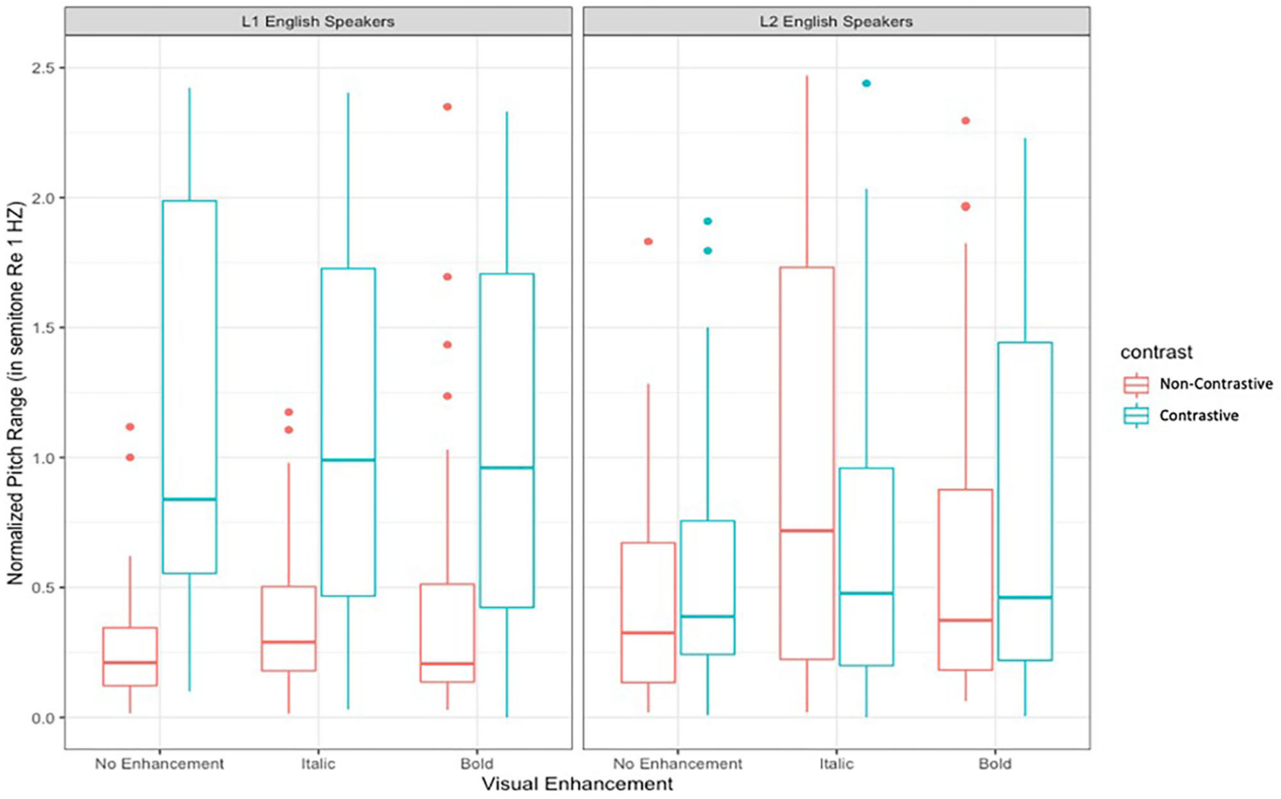


FIGURE 3 | L1 English and Mandarin-English L2 speakers' speech production by visual enhancements.

with less predictable text structures. The first paragraph was a passage adapted from a *New York Times* article. The contrastive information in this passage was signaled using italics. The lexical items analyzed were: "must," "may," "allows," and "requires."

> There are roughly 6,000 languages in the world. Are they mostly the same or are they different from each other? Fifty years ago, a famous linguist pointed out a crucial fact about differences among languages. He said, "Languages differ essentially in what they *must* convey and not in what they *may* convey." What this means is this: if different languages influence our minds in different ways, this is not because of what our language *allows* us to think but rather because of what it habitually *requires* us to think *about*.

The second passage we presented is adapted from Hahn (2004). This paragraph has contrastive information not signaled by any orthographic symbols. The lexical items analyzed were: "personal (1)," "group (1)," "group (2)," "personal (2)."

> I will start by defining the topic for today, which is individualism and collectivism. Individualism concerns the placing of personal goals ahead of group goals. And collectivism concerns placing group goals ahead of personal goals. So let's suppose you have a conflict at work about break time. Let's say your co-workers want longer breaks, but you want shorter breaks. If you're a collectivist, you'll give in to the group. But if you're an individualist, you'll go against the group.

Participants were directed to read each of these passages silently first, during which an eye-tracker was used to measure their processing of the contrastive lexical items where the areas of interest (AOI) were set. Then the participants were asked to read the two passages aloud in a natural way. The results answer question (3) and (4) at the passage level:

(4) How do L1 and Mandarin-English L2 speakers orally produce and visually process lexical items and pictorial information that are contrastive or implicational and that typically receive stress in L1 speakers' speech?
(5) How do L1 and Mandarin-English L2 speakers orally produce and visually process contrastive information written with and without visual enhancement (italics and bold)?

Experiment III differs from Experiments I and II in that all analyzed lexical items are contrastive. Thus, comparisons were made between L1 and Mandarin-English L2 speakers' uses of intonational cues in signaling the contrastive information instead of how L1 and Mandarin-English L2 speakers used intonational cues to signal contrastive information as opposed to non-contrastive information.

We used L1, visual enhancement and the interaction between these two factors as the fixed effects, individual participants, sentences, and words as the random effects to predict normalized pitch range, normalized maximum pitch level, duration, and intensity in the analyses of speech production. We found that L1 English speakers used a greater pitch range to signal the contrastive information compared to Mandarin-English L2 speakers when the contrastive information was visually enhanced (estimate = −1.423, SE = 0.305, df = 20.835, t = −4.668, $^{***}p$ = 0.0001) or not (estimate = 0.9547, SE = 0.3048, df = 20.8352, t = −3.132, $^{*}p$ = 0.005). L1 English speakers also used longer duration in indicating the contrastive information compared to Mandarin-English L2 speakers when italics were used (estimate = −0.131, SE = 0.036, df = 14.724, t = −3.658, $^{**}p$ = 0.0024). However, there was no significant difference in the two groups' use of duration when the contrastive information was not visually enhanced (estimate = −0.059, SE = 0.036, df = 14.724, t = −1.66, p = 0.118). There was no significant difference between L1 and Mandarin-English L2 speakers in the maximum pitch level or intensity with and without visual enhancement.

In terms of the processing of contrastive information in passages, when contrastive information was not visually enhanced (in Passage 2), Mandarin-English L2 speakers had a significantly lower fixation percentage (estimate = −0.882, SE = 0.399, df = 71.738, t = −2.213, $^{*}p$ = 0.03) and dwell percentage (estimate = −0.857, SE = 0.405, df = 72.147, t = −2.114, $^{*}p$ = 0.0379) compared to L1 English speakers. When italics were used to signal contrastive information (in Passage 1), there was no significant difference in the two groups' fixation percentage (estimate = −0.744, SE = 0.399, df = 71.738, t = −1.867, p = 0.066) and a slightly significant difference in the two groups' dwell percentage (estimate = −0.818, SE = 0.405, df = 72.147, t = −2.019, $^{*}p$ = 0.0472).

The production data support findings in Experiments I & II by showing differences in the use of pitch range and maximum pitch level by L1 English and Mandarin-English L2 speakers in speech production of contrastive information. Processing data suggested that L1 English and Mandarin-English L2 speakers differ in the processing of contrastive information at the passage level. Specifically, Mandarin-English L2 speakers did not fix on the contrastive information at a percentage comparable to the L1 English speakers when processing passages. when italics were used to signal contrastive information in passages, there was no difference between the two groups. These findings suggested that Mandarin-English L2 speakers may face challenges processing information with rich contextual information and less predictable text structure. Further, the use of visual enhancement may facilitate visual processing of contrastive information at the passage level.

EXPERIMENT 4

In experiment 4, we investigated L1 and Mandarin-English L2 speakers' processing and production of contrastive information in a picture narrative task. In this task, participants were given a set of eight pictures that describe a single story. Participants were asked to spend a couple of minutes looking at the pictures and then describe the story in English. This picture narrative task is adapted from Derwing et al. (2004) to elicit extemporaneous speech from speakers (see **Appendix C**). The research question that Experiment 4 answered was:

(3) How do L1 and Mandarin-English L2 speakers orally produce and visually process lexical items and pictorial information that are contrastive or implicational and that typically receive stress in L1 speakers' speech?

Two researchers listened to the recordings and selected the words being stressed with the assistance of speech visualization software *Praat*. The researchers found that L1 speakers used intonational cues to signal contrastive information. The following is a transcription of the story told by an L1 speaker with the stressed information in capitalized letters.

> "Two people are walking around in a big city. One man is leaving the building and one woman is ENTERING the building. But they bumped into each other at the entrance and they both dropped... their bags. Um... so they picked it up and walked away. And when the man gets home, he realizes that he has WOMAN's bag. And when the woman gets to work, she realizes that she has the man's bag." (L1 participant #1)

We found that when telling the story, some Mandarin-English L2 speakers did not specify as much contrastive information in their speech as the L1 speakers did. The transcript below illustrates this tendency.

> "In an apartment, um... um... a woman and a man run into each, each other. They took the package. They, they took the package. They make, make a mistake. So, when they return home, they open the package and found they make a mistake." (Mandarin-English L2 participant #3)

We also found that, in some cases, Mandarin-English L2 speakers included contrastive information in the story but did not use intonational cues to signal the contrastive information.

> "It's a big city, a woman and a man are walking on the street and carrying the same suitcases. And they crushed each other at the corner of the street, and the suitcases dropped on the ground. They stand up and pick up their suitcases and walked away. But when the man got home, he found a dress in the suitcase, so he actually got the woman's suitcase. And the woman is in... When the woman got home, she found a tie in her suitcase." (Mandarin-English L2 participant #4)

We used heatmaps to show the areas that the participants paid attention to, and noted differences. We found that L1 speakers paid attention to the details of the pictures in a more holistic way, especially when contrastive information was presented. For instance, in the top two frames in **Figure 4**, L1 speakers spent longer gazing at both the man and the woman as well as the items they found in the suitcases. The Mandarin-English L2 speakers, however, paid less attention to the details and not at the contrastive information. For example, in the bottom two frames in **Figure 4**, one Mandarin-English L2 speaker focused on the man and the item held by the woman and another Mandarin-English L2 speaker focused on the woman and the item held by the man (see **Figure 4**).

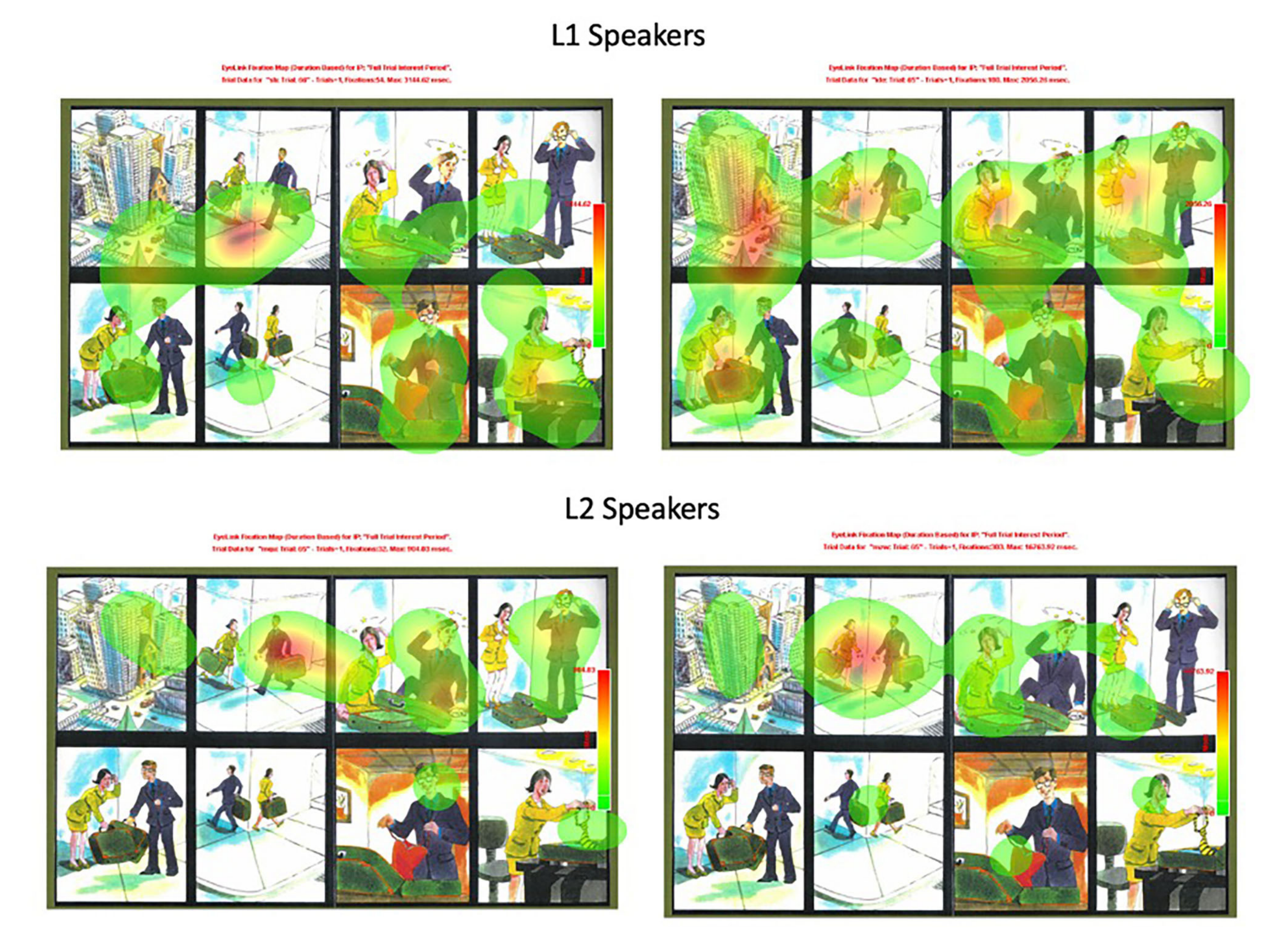


FIGURE 4 | L1 English and Mandarin-English L2 speakers' processing of pictorial information.

DISCUSSION

Analyzing L1 English and Mandarin-English L2 speakers' speech production and processing of written and pictorial information from a Complexity Theory (CT) perspective, the present study found that sentence stress is a multidimensional and multifaceted feature dynamically connected with a series of linguistic and non-linguistic variables. To promote spontaneous use of intonational features like sentence stress, a systematic view that takes into consideration both structure and functional complexity is needed.

In Experiment 1 and Experiment 2, L1 English speakers used a collection of intonational cues to signal contrastive and implicational information in sentences including pitch range, maximum pitch level, and duration. Mandarin-English L2 speakers did not show differences in the use of any of these acoustic features to encode information structure. However, Mandarin-English L2 speakers used greater intensity when producing contrastive information in Experiment 1. The findings in Experiment 3 also showed that, compared to Mandarin-English L2 speakers, L1 English speakers used greater pitch range and higher maximum pitch level when signaling contrastive information in passages. Altogether, these findings suggested that Mandarin-English L2 speakers have difficulties in the integrated use of multiple acoustic cues in signaling contrastive or implicational information. Mandarin-English L2 speakers may, as in Experiment 1, manipulate one acoustic cue (intensity) in an attempt to signal stress. However, intensity was an intonational feature that L1 speakers did not rely on when signaling sentence stress in the same context. These findings suggested the need to address the structural complexity of sentence stress.

To address the structural complexity of the system of intonation, the relationship between different acoustic features needs to be clarified. Intonation teaching will benefit from helping learners to understand intonation as a complex system encompassing multiple interacting features. In addition to chapters focusing on individual features, textbook chapters that adopt a systematic view of intonation and that summarize the relationship among different suprasegmental features will help teachers and learners to develop a systematic view of intonation. Teacher training and preparation programs would do well to also focus more on the systematic use of intonation features and the complexity within the system of intonation.

The results of Experiment 1 showed that L1 English speakers' use of a single intonational cue (i.e., maximum pitch level) is affected by multiple interrelated variables and phenomena including information structure (contrast), morphosyntactic structure (phrasal boundary) and a phonological phenomenon (declination). Mandarin-English L2 speakers' speech, while still showing a general declination trend, did not reflect the influence of information structure or morphosyntactic structure. The results of Experiment 2 further demonstrated that L1 English speakers use intonational features to highlight the lexical item associated with the meanings and implications of the sentences whereas Mandarin-English L2 speakers did not signal implicational information with acoustic cues. These results suggested that learners either were unaware of the connection among intonation, meaning or implication, and information structure, or were incapable of navigating the system of intonation while taking into consideration all influential variables. Processing limitations may be a crucial factor. As O'Brien and Féry (2015) commented about L2 speakers' use of information structure, "[a]n appeal to processing limitations might predict that L2 learners, regardless of their L1s, may have difficulty coordinating all of the potential cues at their disposal when producing structures at the interface… it may be that L2 learners rely on a particular default strategy (e.g., making use of a single syntactic or phonological structure or the same article, regardless of discourse status) as the result of their being unable to integrate all of the types of information in real time" (p. 405). Thus, promoting L2 speakers' awareness about the functions of intonation and their ability in coordinating different segmental and suprasegmental cues for pragmatic proposes may help address the functional complexity of intonation and facilitate more target-like spontaneous use of English intonation.

In Experiments 1 and 2, we did not find significant differences in L1 English and Mandarin-English L2 speakers' processing of contrastive or implicational information as indicated by their fixation percentage and dwell percentage of the focused lexical items. These results support Ip and Cutler's (2016) statement that "Information structure is a linguistic universal" (p. 330). These findings also suggested that when the structure of sentences is relatively stable and predictable, L1 English and Mandarin-English L2 speakers did not differed in their processing of written information. However, in Experiment 3, we found that Mandarin-English L2 speakers' fixation percentage of the contrastive information was significantly lower than that of the L1 speakers when no visual enhancement was used to indicate the information in a passage. In Experiment 4, we found processing differences in the contrastive information in the pictures: L1 speakers focused more holistically on the contrastive information while Mandarin-English L2 speakers gazed at the information that is not contrastive. These results suggested that L1 English and Mandarin-English L2 speakers differ in visual processing when the information structure and morphosyntactic structure were more complex and less predictable. These results further supported O'Brien and Féry's (2015) hypothesis that processing limitations may be the crucial factor in the processing and production of L2 intonation.

In Experiments 1 and 3, Mandarin-English L2 speakers did not use intonational cues in their speech to signal contrastive information even when the information is cued in text by visual enhancement such as italics and bold. The results suggested a lack of familiarity with the L2 orthographical conventions and the connection between visually enhanced information (italicized words) and speech production (sentence stress). Pronunciation textbooks use visual enhancements (e.g., italics) for pedagogical purposes (e.g., indicate the placement of sentence stress). However, the use of visual enhancement such as italics in authentic materials is often a conscious choice authors use to convey their intent (e.g., contrast, implication, etc.). When teaching pronunciation using pedagogical materials with visual enhancements, teachers need to explicitly point out the three-way connection among the constituents that are visually enhanced, the functions and rationale for the use of visual enhancement, and the role of intonation in conveying the meanings and functions in speech production.

CONCLUSION

This study found that the structure and functional complexity of the system of intonation poses challenges to Mandarin-English L2 speakers. Complexity Theory (CT), which emphasizes the connection and interaction of interwoven variables within intonation, is an appropriate framework for L2 pronunciation research and teaching. A systematic view that highlights the complex and dynamic nature of intonation is recommended. Future studies researching the processing limitations of L2 speakers are needed. Further research investigating the dynamic mapping between L1 and L2 intonation is also recommended.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Boston University Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DL and MR jointly identified the theoretical underpinning, designed, and carried out the experiments in the study. DL analyzed the data with the support from MR. DL and MR jointly wrote the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank Dr. Amy Liberman for allowing us to carry out the experiment in her lab. We would also like to thank Xinwen Hu for assisting us in the collection of data. We would like to thank Mary O'Brien and the two reviewers for their constructive feedback and comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.627316/full#supplementary-material

REFERENCES

Allen, V. F. (1971). Teaching intonation, from theory to practice. *TESOl Q.* 5, 73–81. doi: 10.2307/3586113

Anderson-Hsieh, J., Johnson, R., and Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Lang. Learn.* 42, 529–555. doi: 10.1111/j.1467-1770.1992.tb01043.x

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Boersma, P., and Weenink, D. (2018). *Praat: Doing Phonetics by Computer. Version 6.0.37.* Available online at: http://www.praat.org (accessed March 14, 2018).

Bolinger, D. (1989). *Intonation and Its Uses: Melody in Grammar and Discourse*. Stanford, CA: Stanford University Press.

Brazil, D. (1980). *Discourse Intonation and Language Teaching*. Harlow: Longman.

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., and Griner, B. (2010). *Teaching Pronunciation: A Course Book and Reference Guide, 2nd Edn.* Cambridge, UK: Cambridge University Press.

Chapman, M. (2007). Theory and practice of teaching discourse intonation. *ELT J*, 61, 3–11. doi: 10.1093/elt/ccl039

Chen, Y., and Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *J. Phonet*. 36, 724–746.

Chun, D. M. (1982). *A contrastive study of the suprasegmental pitch in Modern German, American English, and Mandarin Chinese* (Doctoral dissertation). University of California, Berkeley, CA, United States.

Chun, D. M. (1988). Teaching intonation as part of communicative competence: suggestions for the classroom. *Die Unterrichtspraxis Teach. German* 21, 81–88. doi: 10.2307/3530748

Clennell, C. (1997). Raising the pedagogic status of discourse intonation teaching. *ELT J.* 51, 117–125. doi: 10.1093/elt/51.2.117

Cruttenden, A. (1990). The origins of nucleus. *J. Int. Phon. Assoc*. 20, 1–9. doi: 10.1017/S0025100300003984

Cruz-Ferreira, M. (1998). "Intonation in European Portuguese," in *Intonation Systems: A Survey of Twenty Languages*, eds D. Hirst and A. Di Cristo (Cambridge, UK: Cambridge University Press), 167–178.

Cucchiarini, C., and Strik, H. (2018). "Automatic speech recognition for second language pronunciation assessment and training," in *The Routledge Handbook of English Pronunciation*, eds O. Kang, R. I. Thomson, and M. J. Murphy (London: Routledge), 556–569.

DeBot, K., Lowie, W., and Verspoor, M. H. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism Lang. Cogn.* 10, 7–21. doi: 10.1017/S1366728906002732

Derwing, T. M., and Munro, M. J. (2005). Second language accent and pronunciation teaching: a research-based approach. *TESOL Q*. 39, 379–397.

Derwing, T. M., Munro, M. J., and Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Lang. Learn.* 48, 393–410. doi: 10.1111/0023-8333.00047

Derwing, T. M., Rossiter, M. J., Munro, M. J., and Thomson, R. I. (2004). Second language fluency: judgments on different tasks. *Lang. Learn.* 54, 665–679. doi: 10.1111/j.1467-9922.2004.00282.x

Ding, S., Liberatore, C., Sonsaat, S., Lučić, I., Silpachai, A., Zhao, G., et al. (2019). Golden speaker builder–an interactive tool for pronunciation training. *Speech Commun*. 115, 51–66. doi: 10.1016/j.specom.2019.10.005

Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annu. Rev. Appl. Linguist*. 30:149. doi: 10.1017/S026719051000005X

Dussias, P. E., and Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism* 10:101. doi: 10.1017/S1366728906002847

Gilbert, J. (2014). "Myth 4: intonation is hard to teach," in *Pronunciation Myths: Applying Second Language Research to Classroom Teaching*, ed L. Grant (Ann Arbor, MI: University of Michigan Press), 107–136.

Giles, H., Mulac, A., Bradac, J., and Johnson, P. (1987). "Speech accommodation theory: the first decade and beyond," in *Communication Yearbook*, Vol. 10, ed M. McLaughlin (Beverly Hills, CA: Sage), 13–48.

Grant, L., and Yu, E. E. (2017). *Well Said Intro: Pronunciation for Clear Communication, 2nd Edition*. Boston, MA: Cengage Learning

Gumperz, J. J. (1971). *Language in Social Groups*. Stanford, CA: Stanford University Press.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation.* Cambridge, UK: Cambridge University Press.

Hahn, L. D. (2004). Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL Q.* 38, 201–223. doi: 10.2307/3588378

Halliday, M. A. K. (2015). *Intonation and Grammar in British English*, Vol. 48. The Hague: Walter de Gruyter GmbH & Co KG.

Hellermann, J. (2003). The interactive work of prosody in the IRF exchange: teacher repetition in feedback moves. *Lang. Soc.* 32, 79–104. doi: 10.1017/S0047404503321049

Holmes, J., and Meyerhoff, M. (1999). The community of practice: theories and methodologies in language and gender research. *Lang. Soc.* 28, 173–183. doi: 10.1017/S004740459900202X

Horley, K., Reid, A., and Burnham, D. (2010). Emotional prosody perception and production in dementia of the Alzheimer's type. *J. Speech Lang. Hearing Res.* 53, 1132–1146. doi: 10.1044/1092-4388(2010/09-0030)

Housen, A., De Clercq, B., Kuiken, F., and Vedder, I. (2019). Multiple approaches to complexity in second language research. Second *Lang. Res.* 35, 3–21. doi: 10.1177/0267658318809765

Ip, M. H. K., and Cutler, A. (2016). "Cross-language data on five types of prosodic focus," in *Proceedings of Speech Prosody*, eds J. Barnes, A. Brugos, S. Shattuck-Hufnagel, and N. Veilleux (Boston, MA), 330–334. doi: 10.21437/SpeechProsody.2016-68

Ito, K., and Speer, S. R. (2008). Anticipatory effects of intonation: eye movements during instructed visual search. *J. Mem. Lang.* 58, 541–573. doi: 10.1016/j.jml.2007.06.013

Kabagema-Bilan, E., Lopez-Jimenez, B., and Truckenbrodt, H. (2011). Multiple focus in Mandarin Chinese. *Lingua* 121, 1890–1905. doi: 10.1016/j.lingua.2011.02.005

Kambe, G., Rayner, K., and Duffy, S. A. (2001). Global context effects on processing lexically ambiguous words: evidence from eye fixations. *Mem. Cognit*. 29, 363–372. doi: 10.3758/BF03194931

Kingdon, R. (1958). *The Groundwork of English Intonation*. New York, NY: Longmans.

Kohler, K. J. (1995). "The Kiel Intonation Model (KIM), its implementation in TTS synthesis and its application to the study of spontaneous speech," in *ATR Workshop on Computational Modelling of Prosody for Spontaneous Speech Processing* (Kyoto).

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge, UK: Cambridge University Press.

Larsen-Freeman, D. (2017). "Complexity theory: the lessons continue," in *Complexity Theory and Language Development: In Celebration of Diane Larsen-Freeman*, eds L. Ortega and Z. Han (Philadelphia, PA: John Benjamins), 11–50.

Levis, J., and Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System* 32, 505–524. doi: 10.1016/j.system.2004.09.009

Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Q.* 33, 37–63. doi: 10.2307/3588190

Levis, J. M., and Levis, G. M. (2018). Teaching high-value pronunciation features: contrastive stress for intermediate learners. *CATESOL J.* 30:139.

Levis, J. M., and Wichmann, A. (2015). "English intonation–form and meaning," in *The Handbook of English Pronunciation*, eds M. Reed and J. Levis (West Sussex UK: John Wiley & Sons, Inc), 139–155.

Lew-Williams, C., and Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *J. Mem. Lang.* 63, 447–464. doi: 10.1016/j.jml.2010.07.003

Lippi-Green, R. L. (1989). Social network integration and language change in progress in a rural alpine village. *Lang. Soc.* 18, 213–234. doi: 10.1017/S0047404500013476

Liu, D. (2020). Prosody transfer failure despite cross-language similarities: evidence in favor of a complex dynamic system approach in pronunciation teaching. *J. Second Lang. Pronunciation* 6, 1–24. doi: 10.1075/jslp.18047.liu

Mair, V. (1996). "Modern Chinese writing," in *The World's Writing Systems*, eds P. T. Daniels and W. Bright (New York, NY: Oxford University Press), 200–208.

Mixdorff, H. (2000). "A novel approach to the fully automatic extraction of Fujisaki model parameters," in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100)*, Vol. 3 (IEEE) (Istanbul). 1281–1284.

Munro, M. J., and Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* 45, 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x

Murphy, J. M. (2014). Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System* 42, 258–269. doi: 10.1016/j.system.2013.12.007

O'Brien, M. G., and Féry, C. (2015). Dynamic localization in second language English and German. *Bilingualism Lang. Cogn.* 18, 400–418. doi: 10.1017/S1366728914000182

O'Brien, M. G., Jackson, C., and Gardner, C. E. (2014). Cross-linguistic differences in prosodic cues to syntactic disambiguation in German and English. *Appl. Psycholinguist.* 35, 27–70. doi: 10.1017/S0142716412000252

O'Connnor, J. D., and Arnold, G. F. (1973). *Intonation of Colloquial English*. London: Longman.

Ortega-Llebaria, M., and Colantoni, L. (2014). L2 English intonation: relations between form-meaning associations, access to meaning, and L1 transfer. *Stud. Second Lang. Acquisition* 36, 331–353. doi: 10.1017/S0272263114000011

Ouyang, I. C., and Kaiser, E. (2015). Prosody and information structure in a tone language: an investigation of Mandarin Chinese. *Lang. Cogn. Neurosci.* 30, 57–72. doi: 10.1080/01690965.2013.805795

Palmer, H. E. (1922). *English Intonation, With Systematic Exercises*. Cambridge: W. Heffer & Sons Limited.

Papadopoulou, D., and Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: a study of relative clause attachment in Greek. *Stud. Second Lang. Acquisition* 25, 501–528. doi: 10.1017/S0272263103000214

Pell, M. D., and Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE* 6:e27256. doi: 10.1371/journal.pone.0027256

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Q.* 35, 233–255.

Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English Specific Purposes* 23, 19–43. doi: 10.1016/S0889-4906(03)00020-6

Pickering, L. (2018). *Discourse Intonation: A Discourse-Pragmatic Approach to Teaching the Pronunciation of English.* Ann Arbor, MI: University of Michigan Press. doi: 10.3998/mpub.6731

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (PhD thesis). Cambridge, MA: Indiana University Linguistics Club. Available online at: http://dspace.mit.edu/handle/1721.1/16065

Pierrehumbert, J., and Hirschberg, J. B. (1990). "The meaning of intonational contours in the interpretation of discourse," in *Intentions in Communication*, eds P. Cohen, J. Morgan, and M. Pollack (Cambridge, MA: MIT Press), 271–311.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Tanenhaus, M. K. (2007). "Spoken language comprehension: insights from eye movements," in *The Oxford Handbook of Psycholinguistics*, eds P. Levelt and A. Caramazza (New York, NY: Oxford University Press), 309–326.

Wells, J. C. (2006). *English Intonation: An Introduction*. Cambridge, UK: Cambridge University Press.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: a study of Chinese speakers of english. *Stud. Second Lang. Acquisition* 20, 1–25. doi: 10.1017/S0272263198001016

Wennerstrom, A. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis*. New York, NY: Oxford University Press.

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

Xu, Y. (2004). "Transmitting tone and intonation simultaneously-the parallel encoding and target approximation (PENTA) model," in *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages* (Beijing), 215–220.

Conflict of Interest: 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.

frontiers in Communication

ORIGINAL RESEARCH
published: 08 April 2021
doi: 10.3389/fcomm.2021.628780

Vowel Quality and Direction of Stress Shift in a Predictive Model Explaining the Varying Impact of Misplaced Word Stress: Evidence From English

Monica Ghosh[1] *and John M. Levis*[2]*

[1] *Institute for Transportation, Iowa State University, Ames, IA, United States,* [2] *English Department, Iowa State University, Ames, IA, United States*

The use of suprasegmental cues to word stress occurs across many languages. Nevertheless, L1 English listeners' pay little attention to suprasegmental word stress cues and evidence shows that segmental cues are more important to L1 English listeners in how words are identified in speech. L1 English listeners assume strong syllables with full vowels mark the beginning of a new word, attempting alternative resegmentations only when this heuristic fails to identify a viable word string. English word stress errors have been shown to severely disrupt processing for both L1 and L2 listeners, but not all word stress errors are equally damaging. Vowel quality and direction of stress shift are thought to be predictors of the intelligibility of non-standard stress pronunciations—but most research so far on this topic has been limited to two-syllable words. The current study uses auditory lexical decision and delayed word identification tasks to test a hypothesized English Word Stress Error Gravity Hierarchy for words of two to five syllables. Results indicate that English word stress errors affect intelligibility most when they introduce concomitant vowel errors, an effect that is somewhat mediated by the direction of stress shift. As a consequence, the relative intelligibility impact of any particular lexical stress error can be predicted by the Hierarchy for both L1 and L2 English listeners. These findings have implications for L1 and L2 English pronunciation research and teaching. For research, our results demonstrate that varied findings about loss of intelligibility are connected to vowel quality changes of word stress errors and that these factors must be accounted for in intelligibility research. For teaching, the results indicate that not all word stress errors are equally important, and that only word stress errors that affect vowel quality should be prioritized.

Keywords: word stress, intelligibility, comprehensibility, error gravity, L2 pronunciation, pronunciation teaching and learning

OPEN ACCESS

Edited by:
Mary Grantham O'Brien,
University of Calgary, Canada

Reviewed by:
Amy Neel,
Western New Mexico University,
United States
Vickie Yu,
California State University, Northridge,
United States

****Correspondence:***
John M. Levis
jlevis@iastate.edu

Specialty section:
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: *12 November 2020*
Accepted: *10 March 2021*
Published: *08 April 2021*

Citation:
Ghosh M and Levis JM (2021) Vowel Quality and Direction of Stress Shift in a Predictive Model Explaining the Varying Impact of Misplaced Word Stress: Evidence From English.
Front. Commun. 6:628780.
doi: 10.3389/fcomm.2021.628780

INTRODUCTION

Word stress, also called lexical stress, refers to a phonological feature of all multisyllabic words in a variety of languages, including English. Word stress is critical in how listeners identify words in the stream of speech, and misplaced stress can make words unintelligible; that is, listeners may misidentify the intended word or they may not identify it at all (Benrabah, 1997).

Stressed syllables have thus been called "islands of reliability" in word identification (Dechert, 1984, p. 227; see also Field, 2005). In other words, stress imposes formulaic phonological patterns that make speech processing easier for listeners. When these expected patterns are not followed, listeners must put forth more effort for understanding (that is, words become less comprehensible) or understanding becomes impossible (that is, words become unintelligible).

Not all languages use stress to mark word prosody. Some use tone (e.g., Chinese, Thai), some pitch accents (e.g., Japanese, Swedish), and some have no identifiable word prosody (e.g., Korean, French). Of the languages with word stress, some have fixed stress (e.g., Polish, Hungarian), in which the same syllable is stressed in all words. For example, Hungarian words have the main stress on the initial syllable and Polish words on the penultimate syllable. Other languages have variable or free word stress, which means that stress occurs initially for some words, finally for others, and on the penultimate or antepenultimate syllable for yet others (e.g., *PHOtograph, eLECtrical, ecoNOmic, questionNAIRE*). Besides English, other free stress languages include Dutch, German, Spanish, Italian, and Russian.

Word stress in English can be signaled by multiple prosodic cues, including syllable length (i.e., duration), pitch (i.e., fundamental frequency), and loudness (i.e., amplitude). Each of these cues can signal distinctions in word stress by itself (Zhang and Francis, 2010) or in conjunction with the other cues, but the default cue used by L1 English listeners to identify stressed syllables is not prosodic but rather segmental—vowel quality. In other words, L1 English listeners, in evaluating whether a syllable is stressed, pay attention first to its vowel quality (Cutler, 2015). If the vowel is full, listeners judge it as stressed. If the vowel is reduced to schwa, listeners do not judge it as stressed. This tendency to evaluate full vowels as stressed extends even to unstressed full vowels, such as the initial vowel in *audition* (Fear et al., 1995). In other languages with vowel quality as a cue to stress, such as Dutch, vowels are not as reliable a stress cue as in English (Cutler et al., 2007). Other variable stress languages like Spanish do not use vowel quality as a cue to stress (Soto-Faraco et al., 2001).

When L2 learners learn a language with word stress, they face a variety of challenges in signaling stress so that listeners find stress to be an island of reliability. If the L2 learner comes from a language with another word prosody, or if they come from a language that has no word prosody, they must learn an entirely new system. If they come from a language with word stress, they need to learn both to hear and produce a new stress system with a new set of cues.

Misplaced stress can result in reduced comprehensibility or unintelligibility. But misplaced stress does not always seriously damage understanding. Slowiaczek (1990) found that changes in stress placement without a change in vowel quality (e.g., *CONcenTRATE* → *CONcenTRATE*) resulted in somewhat slower processing, but the words were successfully understood. Cutler (1986) found that hearing one member of stress minimal pairs such as *INsight/inCITE* and *INsult/inSULT* activated both words for listeners, resulting in no loss of processing time. In other cases, misplaced stress results in L1 English listeners hearing different words altogether. Benrabah (1997) described British listener transcriptions of English words spoken by Indian, Nigerian, and Algerian speakers. Unexpected stress patterns caused listeners to hear completely different words: *UPset* (with initial stress) was transcribed as *absent* (also with initial stress), *riCHARD* as *the child*, and *seCONdary* was heard as *country*. In other words, stress remained an island of reliability for listeners—but they identified the wrong island. When word stress errors cause loss of understanding is thus an open question that has implications both for phonological research and for L2 language teaching and learning.

LITERATURE REVIEW

Stress in English

English is a free or variable stress language. Although in principle multisyllabic words can have the main stress on any syllable, each multisyllabic word has an expected stress pattern. English speakers employ several criteria when deciding how to stress words. Guion et al. (2003) used two-syllable non-sense words to determine that stress decisions are phonologically conditioned, affected by word class, and related by analogy with other visually or phonologically similar words. They found that heavy syllables (CVV, CVCC) were more likely to attract stress than light syllables (CV, CVC), that noun and verb frames (e.g., I'd like a ____ vs. I'd like to _____) affected stress decisions differently, and that unknown words are likely to be stressed similarly to familiar, look-alike words.

Stress in longer words in English is often morphologically conditioned. Words that are etymologically related may be stressed differently based on affixes (e.g., *eLECtric, elecTRIcity, electrifiCAtion*), especially suffixes (Chomsky and Halle, 1968; Dickerson, 1989). In almost all cases, these varied stress patterns become part of the cognitive representation of words, allowing listeners to efficiently access the vocabulary stored in their mental lexicon (Cooper et al., 2002).

Of the four acoustic correlates associated with English word stress—vowel quality, duration, pitch change, and variation in amplitude/intensity/volume, vowel quality [i.e., the distinction between clear (uncentralized/unreduced) and reduced (centralized) vowels] has repeatedly been found the most reliable cue to English word stress (Bond, 1981, 1999; Bond and Small, 1983; Cutler and Clifton, 1984; Cutler, 1986, 2015; Small et al., 1988; Fear et al., 1995; van Leyden and van Heuven, 1996; Cooper et al., 2002; Field, 2005; Cutler et al., 2007; Zhang and Francis, 2010). Reduced (centralized) vowels are *never* associated with stress.

L2 Speakers and Word Stress

L2 speakers of English and other free stress languages can find stress difficult to perceive and produce. This is true even when they speak another free stress language (Maczuga et al., 2017) although this background does facilitate L2 stress learning (Lee et al., 2019). L1 and the age at which L2 speakers learn English are important factors in stress acquisition. The intuitions of early and late bilingual L2 Korean and Spanish speakers were shown to differ in how stress was applied to unfamiliar two-syllable English words (Guion et al., 2004; Guion, 2005). The intuitions of late bilinguals were less like L1 speakers than those of early bilinguals.

L1 can be a dominant factor in how L2 speakers navigate word stress in free stress languages. In the case of French speakers learning Spanish (a free-stress language), Dupoux et al. (2008) asserted that the learners exhibited stress "deafness" in perception of Spanish stress (p. 700). The same difficulties have been reported for French L1 speakers in English stress production (Isaacs and Trofimovich, 2012). Even L1 speakers of free stress languages may not fully be able to use their stress identification abilities when learning other free-stress languages Ortega-Llebaria et al. (2013) found that English speakers were generally sensitive to differences in Spanish stress, but they still struggled to quickly identify differences in Spanish because of contextual stress deafness.

Effects of Misplaced Word Stress on Intelligibility and Comprehensibility

Stress is critical in how L1 English listeners identify words in the stream of speech. Because 90% of lexical (i.e., content) words in spoken English begin with an initially stressed syllable (Cutler and Carter, 1987), L1 listeners treat stressed (or "strong") syllables as marking the first syllable of a new word (Cutler and Norris, 1988; Cutler and Butterfield, 1992). It is no surprise, therefore, that lexical stress errors affect intelligibility and comprehensibility for L1 English listeners because listeners are trying to identify words without being able to identify the first syllable (Kenworthy, 1987; Brown, 1990; Anderson-Hsieh et al., 1992; Dalton and Seidlhofer, 1994; Jenkins, 2000; Field, 2005; Zielinski, 2008; Isaacs and Trofimovich, 2012).

By intelligibility and comprehensibility, we intend Munro and Derwing's (1995) definitions. Comprehensibility is the degree to which listeners can easily understand a speaker's message—that is, for comprehensibility, words, sentences or discourse can span the continuum of being highly comprehensible to being minimally comprehensible. Standard stress pronunciations are generally highly comprehensible—i.e., quickly and easily understood by listeners—and non-standard stress pronunciations with zero vowel errors can be expected to be more comprehensible than those with more errors (e.g., two vowel errors). Intelligibility, on the other hand, refers to the categorical distinction between intelligible and unintelligible pronunciations. Applied to words, listeners either understand a speaker's intended word or they do not.

Errors in English word stress placement interrupt how L1 listeners understand and process speech, thus affecting both intelligibility and comprehensibility. Zielinski (2008) identified word stress as critical for the intelligibility of L2 English speakers to L1 English listeners in both general and academic contexts. This was observed by having L1 English-speaking participants transcribe the utterances of three different L2 English speakers (L1: Chinese, Korean, Vietnamese). Each of the sentences was extracted from 2-h long interviews on the topic of education because listeners had difficulty transcribing it. Each sentence was phonetically transcribed to identify its phonetic deviations and to compare it with the words that were not transcribed correctly. Whenever there was a loss of intelligibility, Zielinski compared it to the non-standard features of the L2 speaker's pronunciation and concluded that L1 English listeners rely heavily on the lexical stress of L2 speakers to determine their intended meaning. Zielinski found that participant transcriptions maintained the L2 speakers' stress pattern 90% of the time.

Even though stress errors that do not change vowel quality are unlikely to prevent correct word identification, such stress errors can nevertheless force listeners to work harder (that is, cause deterioration of the words' comprehensibility). Slowiaczek (1990) examined the accuracy with which L1 English listeners identified mis-stressed words as real words as well as how quickly listeners repeated words. The study used words with two full vowels in which the stress pattern was switched (e.g., *ANgry* vs. *anGRY*) but vowel reduction was not involved. In the identification task, listeners were asked to type each word they heard in quiet and at three different Signal-to-noise (SNR) ratios. Results showed no difference in the accuracy of stressed and mis-stressed words, indicating that when vowel reduction was not involved, listeners successfully identified words despite mis-stressing. Second, listeners were less successful when SNR masking noise was involved, and increasing competition from this noise resulted in less accurate identifications. Third, the majority of the words listeners typed matched the intended word's stress pattern. A second experiment asked listeners to repeat the word they heard. Incorrectly stressed items were responded to more slowly than correctly stressed items, indicating that mis-stressed items interfered with processing.

In another study demonstrating the importance of word stress for comprehensibility, Isaacs and Trofimovich (2012) measured the correlation of 19 linguistic features to L1 English listeners' comprehensibility ratings. Using English picture narratives from 40 L1 French speakers, their goal was to identify the best features to use in an oral language assessment scale for teachers. Of this study's six phonological features, five were significantly correlated with listeners' comprehensibility ratings. Only one, however, word stress, was included in the recommended rating scale because of its high correlation and because teachers identified it as important.

The effect of stress on intelligibility and comprehensibility for L2 English listeners has been much more debated than for L1 English listeners. Largely on the basis of anecdotal evidence, some research has argued that word stress errors are unlikely to result in loss of intelligibility for L2 English listeners (Jenkins, 2000) while others have argued the opposite (Dauer, 2005; McCrocklin, 2012; Lewis and Deterding, 2018). These disagreements raise questions about whether stress errors affect L1 and L2 English listeners differently.

Empirical research suggests that word stress errors can cause loss of intelligibility and/or comprehensibility for both L1 and L2 listeners. Field (2005) developed a list of two-syllable words, half of which were stressed on the first syllable, the other half on the second syllable and recorded each word with standard stress and again with shifted stress. There was also a subset of words with a third condition: shifted stress plus a previously reduced vowel pronounced with full vowel quality. L1 and L2 high-school-aged listeners heard and transcribed the words. For L1 listeners, a shift in stress had a significant negative impact on intelligibility that was lessened if accompanied by full vowel quality. L2 listeners also appeared to follow this general pattern, but once full vowel quality was added, the decrease in intelligibility for L2 listeners was no longer significant. Field also found that stress shifted to the right had a stronger effect on intelligibility than stress shifted to the left.

HYPOTHESIZING AN ENGLISH WORD STRESS ERROR GRAVITY HIERARCHY

In English, L1 listeners attend primarily to vowel quality in evaluating word stress (Cutler and Clifton, 1984; Cutler, 1986, 2015; Cooper et al., 2002; Cutler et al., 2007). This appears to be due both to stressed vowels signaling the beginnings of words in speech and to the reliability of reduced vowels in eliminating possibilities from a listener's subconscious cohort of possible English words (Cutler, 2012). We thus predict that the success of L1 and L2 English listeners' processing of standard and non-standard English stress pronunciations can be predicted based on the number of vowel errors and direction of the stress shifts. A few studies (Cutler and Clifton, 1984; Field, 2005) have found that English word stress errors pushing stress rightward are more damaging than those pushing stress leftward—possibly because English regularly licenses leftward stress shift for the purpose of discourse-level contrastive stress (Field, 2005). Informed by this empirical evidence, we developed the English Word Stress Error Gravity Hierarchy (**Table 1**), leading to two research questions:

1. To what extent do L1 and L2 English listeners process English words (mis)pronounced in accord with the Hierarchy?
2. How do number of vowel errors and direction of stress shift help explain the relative intelligibility and comprehensibility of word stress errors for L1 and L2 English listeners?

TABLE 1 | A hypothesized English Word Stress Error Gravity Hierarchy.

Word stress error category	Vowel errors	Direction of stress shift (Cutler and Clifton, 1984; Field, 2005)	Example: intended word (impacted by) model word (leading to) incorrect stress[a]	Hypothesized error gravity impact
Standard	0	N/A	revision (stressed correctly)	N/A
0 Left	0	Leftward	altérnative ↓ álternate ↓ álternative [ˈɔltərnətɪv]	Low Error Gravity
0 Right	0	Rightward	cóncentrate ↓ (inversion of so-called primary/secondary stress) ↓ concentráte	
1 Left	1	Leftward	progréssive ↓ prógress ↓ prógressive [ˈpraˏgrɛsɪv]	
1 Right	1	Rightward	ínstrument ↓ instruméntal ↓ instrumént [ɪnstɹəˈmɛnt]	
2 Left	2	Leftward	análysis ↓ ánalyze ↓ ánalysis [ænələsɪs]	
2 Right	2	Rightward	célebrate ↓ celébrity ↓ celébrate [səˈlɛbɹet]	High error gravity

Syllables (1) whose vowel quality is changed in Intended Words and (2) which model these vowel quality changes in Model Words are underlined.

METHODS

Participants

Sixty-nine undergraduates with normal hearing volunteered to participate in this auditory lexical decision (LD) (Cutler, 2012) and word identification (WI) (Barca et al., 2002; Balota et al., 2007; Cutler, 2012; Kuperman et al., 2014) study to earn course credit for their introductory psychology class. Thirty-eight spoke English as an L1 (22 females; mean age = 19.34 years, range = 18–26). Thirty-one spoke English as an L2 (14 females; mean age = 21.42 years, range = 18–27). (In our pilot study, we had attempted to limit variability among L2 listeners by including only those whose L1 was either Chinese or Korean, but our U.S. Midwest university context did not include enough participants from these L2s who were taking introductory psychology to make this feasible. As a result, we opened our study to L2 English speakers more generally).

Materials

All participants heard the same (mostly academic) words but were randomly assigned to either Counterbalance Set A or Set B, whose difference lay in which of each word stress category's 16-word sublists was presented with standard vs. non-standard English word stress. Each 16-word sublist was matched as closely as possible for (1) word frequency (van Heuven et al., 2014), since word frequency has long been known to powerfully influence lexical processing; (2) phonological Levenshtein distance 20 (Balota et al., 2007), a phonological similarity (or edit distance) metric, since the more similar neighbors a spoken word has, the more competition words experience during processing, which leads to reaction time delays, etc.; (3) word frequency of phonological Levenshtein distance 20 neighbors (Balota et al., 2007); (4) number of syllables (Balota et al., 2007); (5) dominant word class (Brysbaert et al., 2012); (6) percentage of dominance for dominant word class (Brysbaert et al., 2012); (7) concreteness (Brysbaert et al., 2014); and (8) word stress pattern frequency as analyzed by this study's first author.

Except with the 0 Right category, derivationally related word family members were used to inform all of this study's non-standard pronunciations because American English has ∼14 stressed vowel sounds that are phonemic (Celce-Murcia et al., 2010). Thus, guidance regarding which particular stressed vowel to exchange with a given unstressed vowel (and vice versa) was needed. Because derivationally related words in English often do not have word stress on the same syllables, plausible stressed/unstressed vowel exchanges could be modeled by mapping the word stress pattern of a derivationally related word onto a given manipulated word. Thus, a mis-stressed word may have zero vowel errors (e.g., "altérnative" modeled on álternate to become "álternative"), one vowel error (e.g., "progréssive" modeled on "prógress" to become "prógressive"), two vowel errors (e.g., "económics" modeled on "ecónomy" to become "ecónomics"), etc. In the case of the 0 Right stress manipulation, each counterbalanced sublist manipulated only degree of stress for most words –i.e., exchanging primary vs. secondary stress. For all remaining words (Counterbalance A: 6/16 words; Counterbalance B: 5/16 words), the 0 Right stress manipulation rendered an ordinarily stressed syllable unstressed ("stress" being here defined only suprasegmentally) and an ordinarily unstressed syllable that nevertheless contained a clear (unreduced) vowel stressed (e.g., the word "therapy" pronounced as /ˌθɛrəˈpi/ instead of as its standard pronunciation /ˈθɛrəpi/).

Transcriptions based on the International Phonetic Alphabet for the General American English pronunciation of all stimuli and of all derivationally related word family members modeling stress manipulations were generally obtained from the Web app Lingorado (Jansz, n.d.). However, in the few cases where Lingorado failed to provide an American English IPA transcription or provided a transcription that violated the authors' American English intuitions, other online dictionaries were checked (Cambridge University Press, 2015; Merriam-Webster, 2015; Oxford University Press, 2015) and standard American English IPA transcriptions were developed or revised accordingly. This study's first author then used Ittiam Systems' free ClearRecord Lite iPhone app to record all stimuli in both their standard stress and manipulated stress forms within one of the following four neutral sentence carrier sentences:

1. The word ______________ is interesting.
2. The answer ______________ is reasonable.
3. The choice ______________ is appropriate.
4. The option ______________ is probable.

Recording stimuli in such neutral recording frames avoided effects from either discourse-level rising intonation (signaling the list of words being recorded was not yet finished) or falling intonation (signaling the last word in the list was now being spoken). Stimuli were recorded within their respective carrier sentences with a slight pause before and after each stimulus word, so it could be excised from the recording without contamination from the preceding or following context. Each pronunciation was then evaluated by this study's first author and, upon her initial approval, by the second author, based on their substantial background in phonetics and phonology. Each pronunciation was evaluated within the context of its particular standard or non-standard stress stimulus set for (1) whether it clearly instantiated the target word stress manipulation, (2) whether it included all segmentals appropriately and clearly pronounced and (3) whether it exhibited comparable suprasegmental markers of stress, speed of speaking, etc. Often, stimuli were recorded multiple times before they were deemed satisfactory.

Procedure

Participants were orally introduced to the experimental procedure approved by our university's Institutional Review Board and provided informed consent. Within a comfortable private cubicle, each was interviewed using an extensive Language Background Questionnaire addressing questions about their child and teenage language experience, about their English-language-learning experience and current daily English usage and proficiency, and about any L3 or L4 languages, etc. (see Richards, 2016, for the full questionnaire). Upon the interviewer initiating the experiment and the leaving the cubicle, the participant read: "In this experiment, you will hear a series of correctly and incorrectly pronounced English words. For each word you hear, you will be asked the question 'Was this a correctly pronounced English word?'… If the word was correctly pronounced, you should click '1' to indicate 'Yes, this was a CORRECTLY pronounced English word.' If the word was NOT correctly pronounced, you should click '2' to indicate 'No, this was NOT a correctly pronounced English word.'" Seven practice trials preceded the main experiment, after which participants were given a final review of the experiment's directions and encouraged to ask any questions. Each trial included the following steps.

1. Participants were directed to position their hands ready to click either "1" ("yes") or "2" ("no") as quickly and accurately as possible.
2. Participants pressed the number "1" when ready to continue and after 100 ms heard through their headset either a word spoken in isolation with standard stress or a word spoken in isolation with one of the Hierarchy's six stress manipulations described earlier. At the same time, he or she saw the prompt on the screen "Was this a correctly pronounced English word? Press the '1' key for yes and the '2' key for no."
3. Participants then clicked either "1" or "2" and E-Prime recorded both their LD accuracy and reaction time (RT).
4. Participants were then prompted: "Please type the English word you think the speaker was trying to say and then press 'enter.' (It's okay if you can't spell it correctly—just spell it as best you can☺). If the word was mispronounced and you have NO idea what word the speaker was trying to say, just press the 'enter' key directly." E-Prime recorded all characters typed by the participant.

The study's counterbalancing involved each L1 and L2 participant listening, in random order, to all of the Appendix 1's set A words spoken with standard stress intermixed with all set B words spoken with manipulated stress, or vice versa (Appendix 2 has the words with their phonetic transcriptions). Our word identification task used typed spellings rather than spoken accuracy as a proxy for word identification because of concerns that, particularly with standard pronunciations, it would otherwise have been impossible to identify whether participants' articulations were grounded in their having successfully identified the intended word or were instead the effect of priming leading to their (likely accidentally) simply repeating what they had heard (cf., Field, 2005).

Analysis

One common challenge faced in studies of L1 and L2 language users (Whelan, 2008) is that L1 participants generally perform relatively homogeneously, whereas L2 performance is characteristically much more variable. The current study was no exception though both groups included outliers. An additional source of variability was the wide-ranging difference in performance found across Hierarchy categories, with both L1 and L2 listeners performing for some Hierarchy categories at ceiling and for one category basically at floor. Although several transformations (i.e., logit, arcsine square root and folded square root transformation for the accuracy data and reciprocal and log-normal transformation for the RT data) were tried, none were particularly effective at addressing the failure of this study's accuracy and RT data to meet ANOVA's homogeneity of variance and normality assumptions. An additional issue with non-linear data transformation is that while it can address questions of rank order, it cannot resolve questions about relative degree of impact (Whelan, 2008; Lo and Andrews, 2015) since, for example, the square root of 25 is 5, of 16 is 4, and of 9 is 3 (i.e., non-linear transformation can render non-equidistant values equidistant). Details of all non-linear data transformations attempted are available from the dissertation of this study's first author (Richards, 2016). Because this study's research questions are not so much about how L1 listeners and L2 listeners perform *in relation to each other*, but rather about how each group's performance compares to the predictions of our hypothesized English Word Stress Error Gravity Hierarchy, the current paper reports ANOVA analysis of the untransformed L1 and L2 listener groups' data separately. In other words, although we could not justify inferential analysis of the two groups together in light of our L1 and L2 listeners' substantial difference in variance, we relied on ANOVA's noted robustness to normality violations in light of our L1 and L2 groups' respective sample size of >30—as licensed by the Central Limit Theorem that describes how, no matter a particular data distribution's shape (i.e., normal or not), the greater the sample size, the closer the sample means will approximate their respective population means.

RESULTS

Our results from testing the English Word Stress Error Gravity Hierarchy are presented in three parts. First, we report the results for Lexical Decision (LD) accuracy and reaction time in light of hierarchy predictions. These two variables, respectively, measure how accurate listeners were in determining whether words were correctly or incorrectly pronounced and how long it took them to decide. Next, we report the results of the Word Identification (WI) task, in which listeners typed out the word they heard. This task was our proxy measure for the intelligibility of (mis)pronounced words across the hierarchy. For each of this section's three parts, we present the L1 results, the L2 results, and then compare the L1 and L2 listeners. Finally, we look at how our study connects with the few others that have noted that listeners' word stress error processing appears to be predicted not only by the presence or absence of vowel errors, but also by direction of stress shift (Cutler and Clifton, 1984; Field, 2005).

Lexical Decision Accuracy and Reaction Time

The Hierarchy predicts that L1 and L2 English listeners' LD accuracy with the non-standard stress categories relatively close to standard stress will be poor but will progressively improve the further a non-standard stress pronunciation falls from the standard stress category of the Hierarchy. Specifically, it predicts that listeners' LD accuracy will be better for words at the

TABLE 2 | Percentage of mean difference between pairs of hierarchy categories in L1 listeners' LD accuracy.

	Standard	0 Leftward	0 Rightward	1 Leftward	1 Rightward	2 Leftward	2 Rightward
Standard	0	0.72*	0.41*	0.21*	0.13*	0.01	0.02
0 Leftward		0	0.31*	0.52*	0.59*	0.72*	0.71*
0 Rightward			0	0.20*	0.28*	0.40*	0.39*
1 Leftward				0	0.08*	0.20*	0.19*
1 Rightward					0	0.12*	0.11*
2 Leftward						0	0.01
2 Rightward							0

*Bonferroni-corrected post-hoc pairwise comparisons. *Shows the difference is significant at the 0.05 level.*

two ends of the hierarchy (i.e., pronounced with a standard pronunciation and those most clearly pronounced with a non-standard pronunciation). The Hierarchy conversely predicts that listeners will exhibit reduced LD accuracy and slower reaction times (RTs) for mis-stressed English words falling into categories in the middle section of the Hierarchy due to struggles in identifying whether these "almost-correctly-pronounced" words have in fact been correctly pronounced.

L1 English Listeners' LD Accuracy and LD RT

The L1 English listeners' LD accuracy data follow the expected pattern. A significant within-subjects ANOVA with a Greenhouse-Geisser correction, $F_{(2.33, 86.19)} = 136.98$, $p < 0.001$, and very large partial η^2 effect size show that 79% of the variance in L1 English listeners' LD accuracy can be attributed to Hierarchy category. Bonferroni-corrected pairwise comparisons of the percentage of mean difference between Hierarchy categories, as shown in **Table 2**, demonstrate that the L1 English listeners were nearly 100% accurate at identifying English words pronounced with standard stress as instantiating a standard pronunciation, and they were similarly nearly 100% accurate at recognizing basically all 2-vowel-error non-standard pronunciations as being non-standard. In contrast, their LD accuracy with the middle-of-the-Hierarchy non-standard stress categories was poor.

In terms of LD RT, seven of the L1 English listeners inaccurately rated *all* 0 Left non-standard pronunciations as instantiating "a correctly pronounced English word." As a result, they had *no* RT associated with an accurate LD for the 0 Left category. Seven other L1 English listeners rated only one 0 Left non-standard pronunciation as non-standard and therefore had only one RT associated with an accurate LD for the 0 Left category. Therefore, LD RT data across all categories of the Hierarchy was available for submission to statistical analysis for only 24 of our 38 L1 English listeners. For these 24 L1 listeners, Bonferroni-corrected pairwise comparisons make clear their significant within-subjects ANOVA with a Greenhouse-Geisser correction, $F_{(3.54, 120.27)} = 9.47$, $p < 0.001$, partial $\eta^2 = 0.218$, is merely the characteristic artifact of the LD task that is predicted by the Dual Route Cascaded model, namely that accurate "Yes" responses will be faster and less variable (i.e., responding to standard stress stimuli with a "Yes" LD) than accurate "No" responses (i.e., responding to non-standard stress stimuli with a "No" LD) (Coltheart et al., 2001; Cutler, 2012). Specifically, as **Figure 1** suggests, the RTs associated with L1 English listeners' increasingly greater number of accurate "No" LDs across the 0 Right−2 Right non-standard stress categories of the Hierarchy were statistically equivalent, indicating that across these non-standard stress categories, the mental cost of performing the LD task, as indexed by RT, was stable.

However, there is one telling exception to this overall RT trend. Although these 24 more sensitive L1 English listeners ultimately succeeded at identifying 0 Left non-standard stress pronunciations as non-standard, Bonferroni-corrected pairwise comparisons show this success came at a significant RT cost (median RT = 783 ms) relative to all Hierarchy categories except 2 Left. In other words, not only was the L1 English listeners' LD accuracy extremely low in recognizing 0 Left mis-stressings as non-standard, but on the rare occasions when they *did* succeed, the price tag was prolonged mental debate. L1 listeners' barely 50% LD accuracy and significantly slower median RT (=521 ms) when identifying the 0 Right mis-stressings as non-standard similarly contrasts with their nearly 100% accuracy and 365 ms median RT in recognizing each of the study's standard stress pronunciations as a "correctly pronounced English word."

These findings are not surprising since previous research has made it clear L1 English listeners have difficulty utilizing the suprasegmental word stress cues of duration, pitch and intensity that are often redundant with the more salient vowel quality cue (Cooper et al., 2002; Cutler et al., 2007). After all, this study's 0 Left and 0 Right Hierarchy-defined non-standard stress pronunciations offered only these suprasegmental word stress cues. It is also no surprise the L1 English listeners struggled particularly to identify 0 Left word stress shifts as non-standard since, as mentioned earlier, English regularly licenses leftward stress shift for the purpose of discourse-level contrastive stress (Field, 2005).

Yet these LD accuracy and LD RT findings in conjunction with such research raised the following question: To what extent *was* the L1 English listeners' definition of a "correctly pronounced English word" broad enough to accommodate the 0 Left and/or 0 Right Hierarchy-defined non-standard stress pronunciations that offer solely suprasegmental word stress cues?

Post-hoc analysis by reverse-coding L1 listeners' Hierarchy-defined *in*accurate LDs for the 0 Left and 0 Right categories as *accurate* and their Hierarchy-defined *accurate* LDs for these two suprasegmentally-demarcated categories as *in*accurate

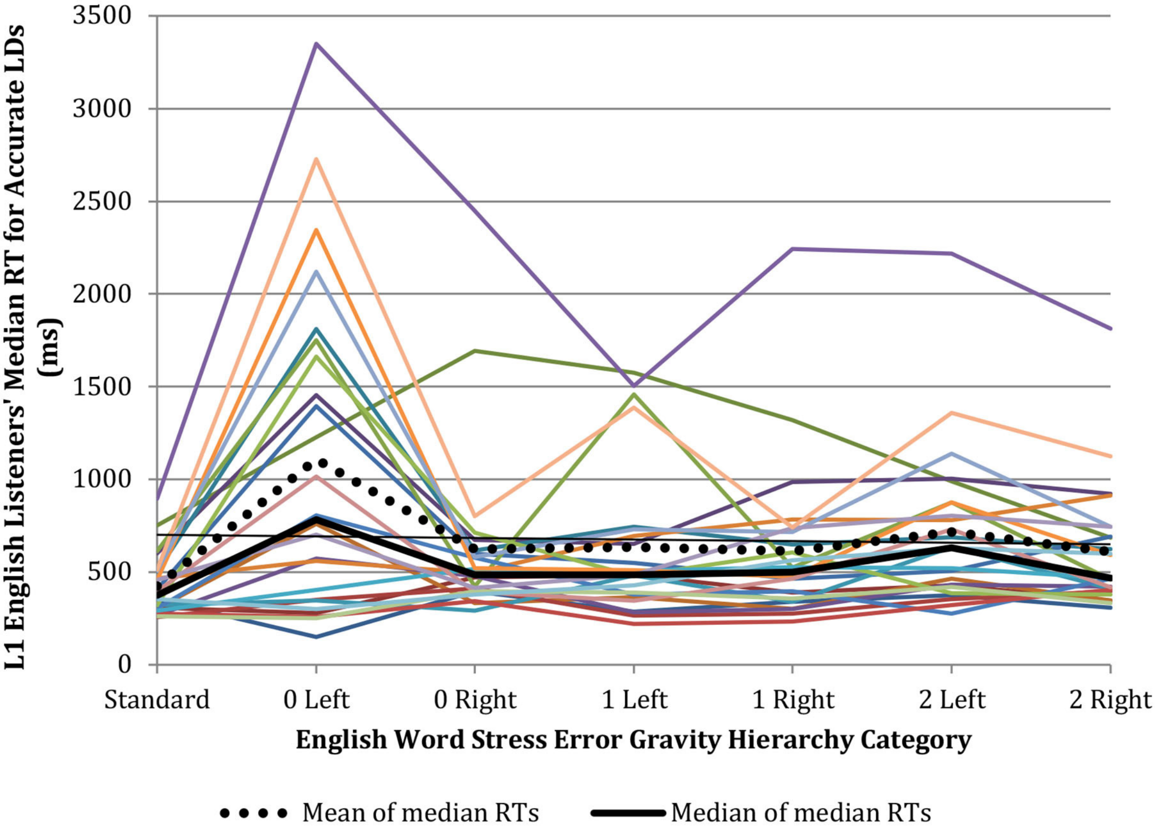


FIGURE 1 | Parallel coordinate plot of median accurate auditory LDRT for L1 English listeners with 2+ accurate LDs per Hierarchy category ($n = 24$).

allowed us to model this question. However, we removed in our reverse-coded LD RT analysis the most suprasegmentally sensitive L1 English listeners who made either zero or only one 0 Left or 0 Right Hierarchy-defined inaccurate LD. Bonferroni-corrected pairwise comparisons confirmed that for the remaining typical L1 English participants ($n = 32$), there was no significant difference in their median RT for indicating that standard stress pronunciations in comparison to 0 Left mispronunciations represented "correctly pronounced English word(s)." For the 0 Right stimuli, this *post-hoc* analysis revealed that their median 0 Right accurate LD RT *and* inaccurate LD RT were significantly slower than their median RT for accurate standard stress LDs. In sum, for the most part, L1 listeners did not hesitate to classify 0 Left pronunciations as "correctly pronounced English words," but they struggled to determine whether 0 Right mis-stressings had been pronounced correctly or incorrectly, no matter what their ultimate decision. Thus, L1 listeners' sensitivity to the suprasegmental correlates of English lexical stress depended on the direction of stress shift.

L2 English Listeners' LD Accuracy and LD RT

The L2 listeners' LD accuracy data visually follow a similar pattern to that of the L1 listeners (**Figure 2**), though apparently from a lower baseline and, as is frequently the case in studies involving L1 and L2 language users (Whelan, 2008), with much greater variability.

A significant within-subjects ANOVA with a Greenhouse-Geisser correction, $F_{(3.39, 101.81)} = 37.75$, $p < 0.001$, and very large partial η^2 effect size show that 56% of the variance in L2 English listeners' LD accuracy data can be attributed to Hierarchy category. This effect size is impressive given that, for the 1 Left and 2 Left categories, L2 listeners' scores range all the way from 0 to 100%. Bonferroni-corrected pairwise comparisons, displayed in **Table 3**, show this large effect size is due to L2 listeners' performance with the standard stress and 0 Left Hierarchy categories being reliably different from all other Hierarchy categories and the 1 Left category reliably different from all other categories except the immediately adjacent categories 0 Right and 1 Right.

In terms of LD RT, L2 English listeners' median accurate 0 Left LD RT represents an ~850 millisecond increase in processing time over their median accurate LD RT for all other non-standard stress categories. It thus visually appears (**Figure 3**) that in cases when L2 listeners *were* able to make a Hierarchy-defined accurate LD with the 0 Left pronunciations, they paid an RT cost to do so. However, while within-subjects ANOVA with a Greenhouse-Geisser correction run on L2 English listeners' LD RT data was significant $F_{(4.12, 90.57)} = 7.34$, $p < 0.001$, and had a large partial η^2 effect size of 0.25, Bonferroni-corrected pairwise

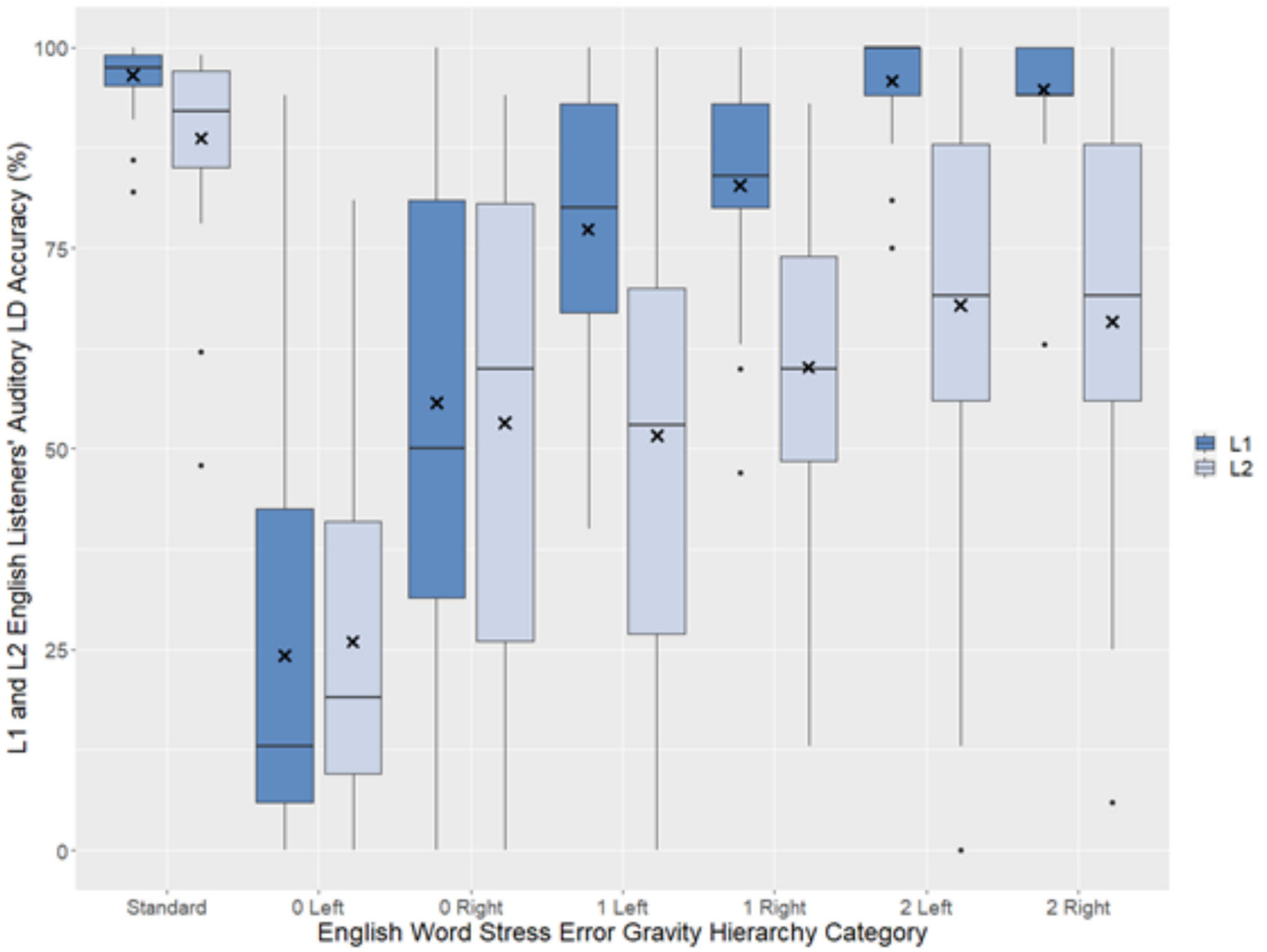


FIGURE 2 | Auditory LD accuracy by Hierarchy category for L1 (*n* = 38) and L2 (*n* = 31) English listeners. "X" marks the sample mean and center lines the median of listeners' individual mean LD accuracy, box limits indicate the interquartile range, whiskers contain all sample values within 1.5 times the interquartile range, and outliers are represented by dots.

TABLE 3 | Percentage of mean difference between pairs of hierarchy categories in L2 listeners' LD accuracy.

	Standard	0 Leftward	0 Rightward	1 Leftward	1 Rightward	2 Leftward	2 Rightward
Standard	0	0.63*	0.36*	0.37*	0.25*	0.21*	0.23*
0 Leftward		0	0.27*	0.25	0.37	0.42*	0.40
0 Rightward			0	0.17	0.10	0.15	0.13
1 Leftward				0	0.12	0.17*	0.15*
1 Rightward					0	0.05	0.03
2 Leftward						0	0.02
2 Rightward							0

*Bonferroni-corrected post-hoc pairwise comparisons. *Shows the difference is significant at the 0.05 level.*

comparisons indicate L2 English listeners' *only* significant accurate auditory LDRT result, perhaps due to their overall wide variability, is that predicted by Dual Route Cascaded Model finding, namely that accurate "Yes" responses are faster and less variable than accurate "No" responses (Coltheart et al., 2001; Cutler, 2012).

L1 vs. L2 English Listeners' LD Accuracy and LD RT

While the visual similarity in L1 and L2 listeners' LD accuracy data seen in **Figure 2**—and even more unmistakably in **Figure 4**—is intriguing, as mentioned earlier, it was impossible to test the significance of this potential LD accuracy difference because the L1 vs. L2

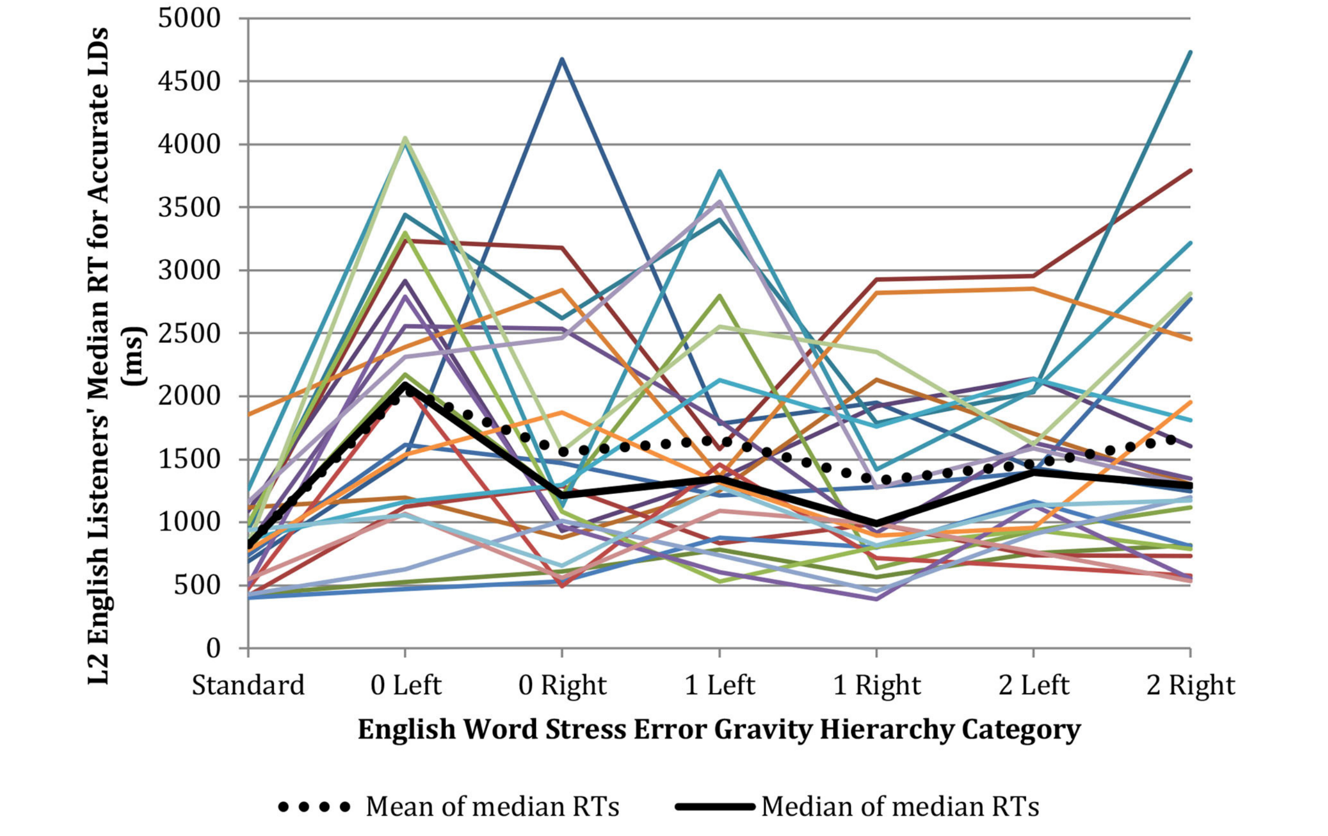


FIGURE 3 | Median accurate auditory LDRT for L2 English listeners with 2+ accurate LDs per Hierarchy category (n = 23).

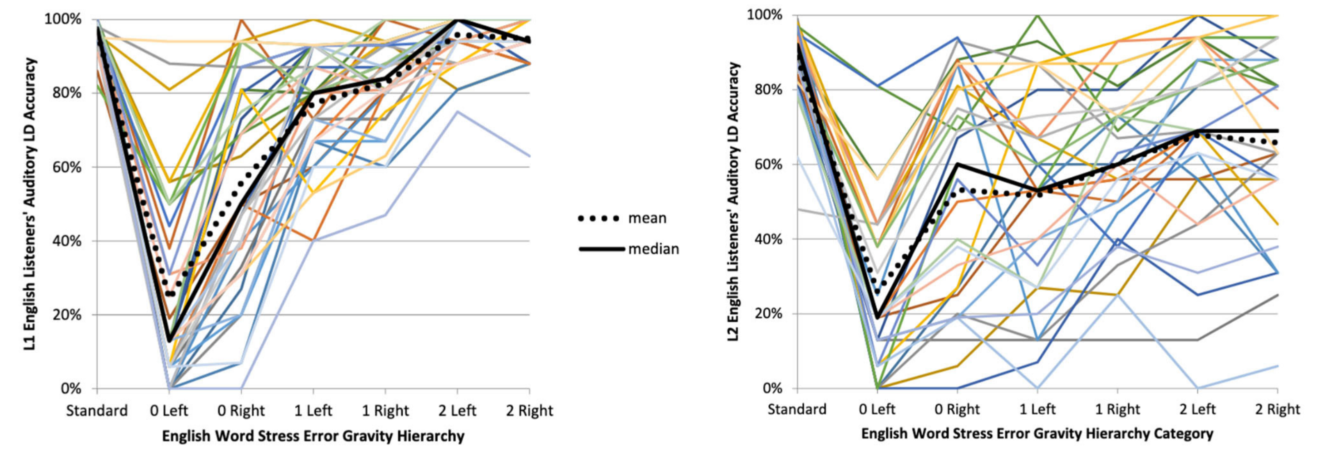


FIGURE 4 | Auditory lexical decision (LD) accuracy by English Word Stress Error Gravity Hierarchy category for L1 (n = 38, left) and L2 (n = 31, right) English listeners.

listener data strongly violated ANOVA's homogeneity of variance assumption.

What should be noted from **Figure 2** about L1 and L2 listeners' LD accuracy, however, is that L2 listeners' interquartile range barely overlaps with that of L1 English listeners for all non-standard stress categories in which an English word stress error induces one or more concomitant vowel errors. In other words, the L2 English listeners did not merely follow L1 English listeners' performance from a lower baseline. Rather, the further an English word stress error fell from the standard stress category

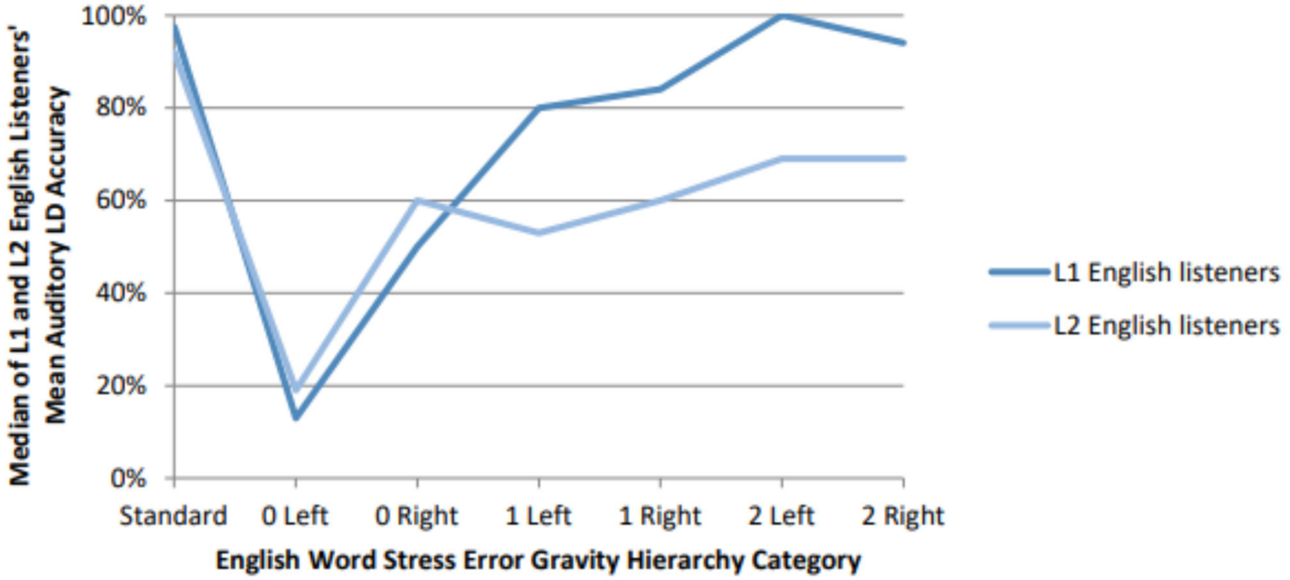


FIGURE 5 | Median of L1 (n = 38) and L2 (n = 31) English listeners' mean auditory LD accuracy by Hierarchy category.

of the Hierarchy, the more L2 listeners' LD accuracy was hurt in comparison to that of L1 listeners.

Also, one additional point for future research should be noted. For most Hierarchy categories, as one might expect, L2 listeners' mean and median LD accuracy is lower than that of L1 listeners. However, the L2 listeners' mean LD accuracy in **Figure 4** for the 0 Left and 0 Right categories almost exactly mirrors that of L1 listeners—and L2 listeners' median LD accuracy in **Figure 5** actually *exceeds* that of L1 listeners. How can this be?

Post-hoc analysis of the L2 English listeners' Language Background Questionnaire data suggests this anomaly may be explained by the fact that the "L2 listener" group subsumed both those from pitch-contrastive and non-pitch-contrastive L1s. Specifically, 17 of our L2 listeners were from either a tonal or pitch-accent L1 (Chinese $n = 11$, Vietnamese $n = 4$, Lao $n = 1$, and Japanese $n = 1$) and 14 were from a non-tonal, non-pitch-accent L1 (Arabic $n = 3$, Korean $n = 3$, Malay $n = 2$, Spanish $n = 2$, Czech $n = 1$, Indonesian $n = 1$, Turkish $n = 1$, and Urdu $n = 1$). Largely in accord with L2 listeners' self-assessed English listening and speaking proficiency (**Table 4**), the pitch-contrastive L1 listeners' LD accuracy appears generally lower across Hierarchy categories than that of the non-pitch-contrastive L1 listeners—a finding one of our reviewers has suggested may be due to English including several short (lax) vowels that are *not* part of many East Asian languages' vowel inventory, making it difficult for speakers of these languages to accurately determine whether English words containing these short vowels have or have not been correctly pronounced.

Specifically, the *only* two English Word Stress Error Gravity Hierarchy categories where the tonal or pitch-accent L1 listeners apparently outperform not only their non-tonal, non-pitch-accent L1 peers (**Figure 6**), but also L1 English listeners (**Figure 5**) are the two categories where only the suprasegmental cues to non-standard stress—including the pitch cue—were available. In other words, as is characteristic of L2 speech processing generally (Cutler, 2012), retaining their L1 speech processing strategy of closely attending to the pitch cue apparently served pitch-contrastive L1 listeners well for these two Hierarchy categories. While this study's small sample size for pitch-contrastive vs. non-pitch-contrastive L1 listeners made it impossible to test the significance of their apparent LD accuracy differences, future research investigating this apparent phenomenon would be of interest.

Word Identification Accuracy

The English Word Stress Error Gravity Hierarchy predicts that L1 and L2 English listeners should generally be able to recognize a speaker's intended word for English words pronounced with standard stress or pronounced with non-standard stress that is marked only suprasegmentally (Bond and Small, 1983; Cutler and Clifton, 1984; Cutler, 1986; Small et al., 1988; Fear et al., 1995; Jenkins, 2000; Cooper et al., 2002; Field, 2005). However, the further a non-standard stress pronunciation falls from the standard stress category of the Hierarchy, the more intelligibility is expected to decrease and therefore the less accurate word identification (WI) accuracy should become. Inaccurate WI was defined in this study as either (1) not attempting at all to spell a speaker's intended word or (2) spelling a real English word other than what the speaker intended. Typos consisting only of added non-alphabetic characters were counted as instances of accurate WI. Other misspellings were deleted from the data prior to WI accuracy analysis, since it was oftentimes impossible to decide objectively whether they represented (1) listeners' misspelling of the speaker's intended word that they had in fact accurately identified or (2) listeners' attempt to spell phonetically what they had heard, a response likely on at least some occasions because

TABLE 4 | Responses to Language Background Questionnaire's Self-Assessment of English Listening and Speaking Proficiency.

"Now I'm going to read several statements and I want you to tell me how often each expresses your current English listening/speaking proficiency on a scale of 'never,', 'rarely,' 'sometimes,' 'most of the time' or 'always.'"*	**Pitch-contrastive L1 speakers (*n* = 16)†**		**Non-pitch-contrastive L1 speakers (*n* = 14)**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**
I understand English speakers talking with one another (except maybe when they're talking about weird topics I know nothing about in any language, including my native language!)	3.63	0.86	4.00	0.65
Native English speakers understand my English pronunciation	4.00	0.61	3.86	0.64
Native English speakers *easily* understand my English pronunciation	3.56	0.79	3.79	0.67
Other highly proficient non-native English speakers whose native language is different from mine understand my English pronunciation	3.81	0.88	3.93	0.70
Other highly proficient non-native English speakers whose native language is different from mine *easily* understand my English pronunciation	3.44	0.70	3.57	0.73
I can easily say what I *need* to say in English	3.88	0.60	4.14	0.74
I can easily say what I *want* to say in English	3.50	0.50	3.93	0.88
My English accent matches my ideal (or "dream") English accent	2.75	0.83	2.93	0.80

**Listeners' rank-ordered "never"–"always" responses were, respectively, converted during analysis to the numerical range 1–5.*

†The Language Background Questionnaire data for one L1 Chinese participant was apparently somehow not saved or otherwise lost.

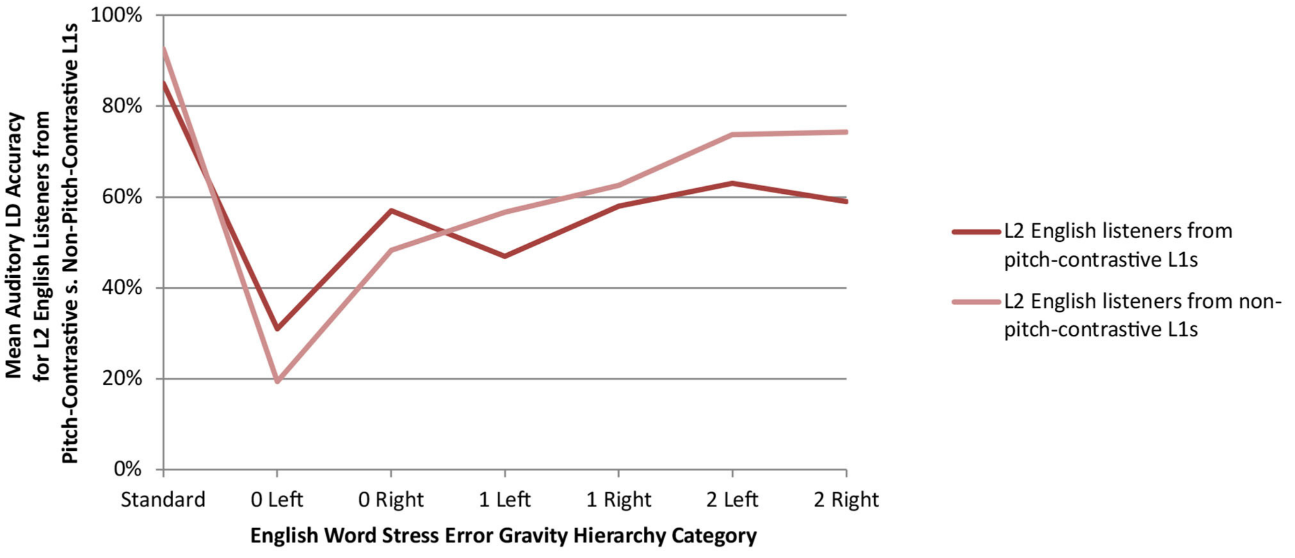


FIGURE 6 | Mean auditory LD accuracy for L2 English listeners' from pitch-contrastive L1s (*n* = 17) and non-pitch-contrastive L1s (*n* = 14) by Hierarchy category.

most non-standard pronunciations had been modeled on the standard pronunciation of a derivationally-related word family member and therefore likely sounded somewhat familiar to listeners. After all, while some misspellings appeared to be minor misspellings, e.g., "affectionite" for "affectionate" pronounced as [ˈæfɛkʃənət] and other misspellings were interesting in terms of this study's research questions, e.g., "mejestic" spelled instead of "majesty" for the pronunciation [məˈʤɛsti], we decided not to assume, in the absence of clear evidence, that a listener who typed, for example, "lugsurious" for "luxurious" pronounced as [ˈl∧gʒəriəs] successfully retrieved the speaker's intended word but simply misspelled it. In addition, because the WI task asked listeners to "Please type the English word you think the speaker was trying to say," making clear that the speaker had (mis)pronounced a real English word, L2 listeners may have assumed that inability to identify the speaker's intended word would signal inadequate English proficiency on their part and therefore wished to save face by attempting to spell something rather than admit they were unable to accurately identify the speaker's intended word by not attempting any spelling at all (All WI responses are summarized in **Table 5** and their underlying raw data are available from Richards, 2016, pp. 179–245).

L1 English Listeners' WI Accuracy

We observed deterioration in L1 listeners' WI accuracy, our intelligibility proxy, with the non-standard stress categories

TABLE 5 | Percent of WI response tokens submitted vs. not submitted to statistical analysis.

	Standard pronunciation potentially as…							Nonstandard pronunciation as…						
L1 (n = 38)	**0L**	**0R**	**1L**	**1R**	**2L**	**2R**	**Total**	**0L**	**0R**	**1L**	**1R**	**2L**	**2R**	**Total**
Submitted to statistical analysis as "correct"	0.94	0.91	0.85	0.83	0.91	0.87	0.89	0.93	0.90	0.87	0.81	0.84	0.84	0.87
Submitted to statistical analysis as "incorrect"	0.01	0.01	0.04	0.02	0.00	0.03	0.02	0.01	0.02	0.03	0.08	0.09	0.08	0.05
Unattempted	0.01	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.07	0.05	0.05	0.03
Spelled a REAL English word, but not the speaker's INTENDED word	0.00	0.01	0.03	0.01	0.00	0.03	0.01	0.00	0.01	0.02	0.01	0.04	0.03	0.02
Misspellings not submitted to statistical analysis	0.05	0.09	0.11	0.15	0.09	0.10	0.10	0.06	0.08	0.09	0.11	0.07	0.08	0.08
L2 (n = 31)														
Submitted to statistical analysis as "correct"	0.74	0.79	0.67	0.62	0.72	0.77	0.72	0.78	0.76	0.67	0.57	0.64	0.64	0.68
Submitted to statistical analysis as "incorrect"	0.13	0.06	0.13	0.16	0.10	0.09	0.11	0.11	0.11	0.16	0.23	0.20	0.22	0.17
Unattempted	0.09	0.04	0.05	0.13†	0.06	0.05	0.07	0.08	0.08	0.11	0.20†	0.16	0.18	0.13
Spelled a REAL English word, but not the speaker's INTENDED word	0.04	0.01	0.07	0.03	0.04	0.04	0.04	0.03	0.03	0.05	0.03	0.04	0.04	0.04
Misspellings not submitted to statistical analysis	0.13	0.15	0.20	0.23	0.19	0.14	0.17	0.11	0.13	0.17	0.20	0.16	0.14	0.15

†Many lexical criteria commonly used in psycholinguistics studies to ensure the cognitive equivalence of stimuli were matched across this study's Hierarchy categories, but it should be noted that the stimuli used for all other categories besides 1 Right were also chosen in light of years of ESL teaching experience. Unfortunately, very few potentially 1 Right non-standard stress manipulations could be found, making it impossible to exclude words which this study's L2 participants may very well not have known, e.g., "ridicule."

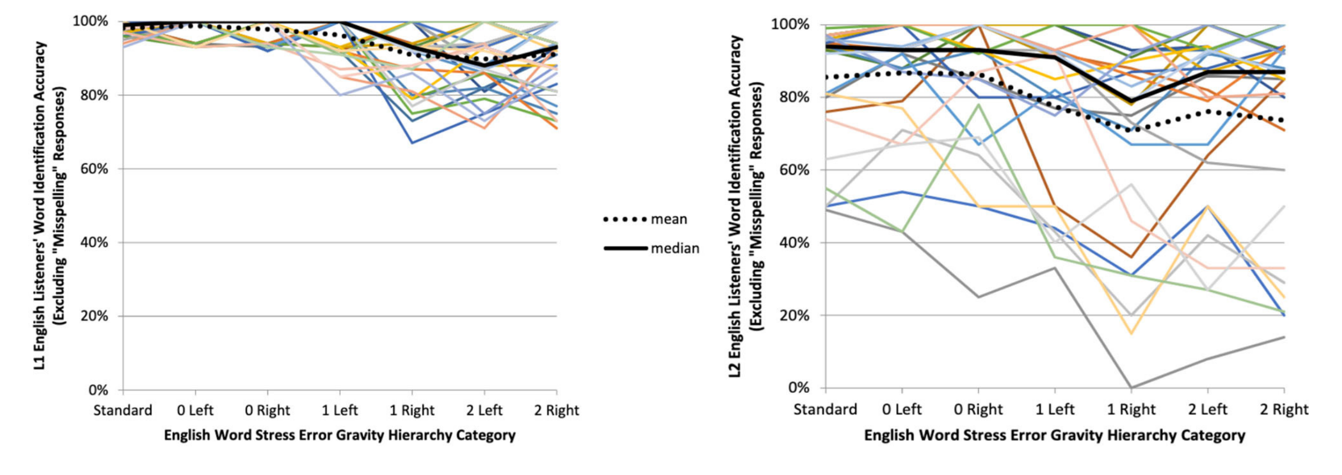


FIGURE 7 | (Typed) word identification accuracy (excluding misspellings) by Hierarchy category for L1 (n = 38, left) and L2 (n = 31, right) English listeners.

furthest from standard stress (**Figure 7** left). Within-subjects ANOVA with a Greenhouse-Geisser correction applied to L1 listeners' (typed) WI accuracy data indicates their WI accuracy varied significantly across Hierarchy categories, $F_{(3.59, 132.96)} = 18.76$, $p < 0.001$, partial $\eta^2 = 0.34$. This large partial η^2 effect size shows that 34% of the variance in L1 English listeners' WI accuracy can be attributed to Hierarchy category. It is only at the 1 Right non-standard stress category that L1 listeners began exhibiting significant deterioration in WI accuracy, either not attempting at all to spell the speaker's intended word or spelling a real English word other than that which the speaker intended.

L2 English Listeners' WI Accuracy

Within-subjects ANOVA with a Greenhouse-Geisser correction indicates L2 English listeners' WI accuracy varied significantly across Hierarchy categories, $F_{(4.12, 123.70)} = 10.69$, $p < 0.001$, partial $\eta^2 = 0.26$. Despite the variability in L2 English listeners' WI accuracy evident in **Figure 5** (right), this partial η^2 effect size shows that 26% of the variance in their WI accuracy is attributable to Hierarchy category. Like for L1 listeners, the 1 Right non-standard stress Hierarchy category is where the L2 listeners began to exhibit significant deterioration in WI accuracy—though unlike the L1 listeners, this was true for L2 listeners relative only to L2 standard and 0 Left category performance.

L1 vs. L2 English Listeners' Word Identification Accuracy

Both L1 and L2 English listeners experienced the greatest deterioration in intelligibility with the Hierarchy categories farthest from standard stress. However, while the visual similarity in L1 and L2 listeners' WI accuracy data seen in **Figure 7** is interesting, it was again impossible to test the significance of this potential difference because of how the L1 vs. L2 listener data strongly violated ANOVA's homogeneity of variance assumption. Although several (i.e., the logit, arcsine square root, and folded square root) transformations were attempted, no transformation succeeded at rendering this study's data homogenous.

Unattempted spellings may be the clearest possible indicator of unintelligibility as they occur only when listeners, despite being assigned no penalty for guessing, were nevertheless unwilling or unable to attempt identifying the speaker's intended word. For both L1 and L2 listeners, the number of unique words (types) and percent of total words (tokens) they declined to identify sharply increases at the 1 Right category (**Table 5**). However, the L2 listeners experienced substantially reduced intelligibility relative not only to L1 listeners, but even to their own standard stress performance. The L2 listeners therefore were not performing merely from a lower baseline than their L1 listener counterparts, but rather were impacted to an even greater degree (cf., Jenkins, 2000, 2002).

The Hierarchy and Word Stress Error Processing

The aim of the English Word Stress Error Gravity Hierarchy is to provide a means of predicting how listeners are likely to process any given word stress error. Many studies have noted the impact of vowel quality on L1 and L2 English listeners' word stress error processing (Bond and Small, 1983; Cutler and Clifton, 1984; Cutler, 1986; Small et al., 1988; Fear et al., 1995; Cooper et al., 2002; Field, 2005) and that direction of stress shift also impacts listener understanding (Cutler and Clifton, 1984; Field, 2005). This study adds to this research as follows.

In terms of our L1 listeners' LD accuracy data, within-within-subjects ANOVA with a Greenhouse-Geisser correction found a significant interaction between the number of vowel errors and direction of stress shift, $F_{(1.91, 57.4)} = 22.21, p < 0.001$. Specifically, although both number of vowel errors and direction of stress shift significantly affected L1 listeners' LD accuracy, only the number-of-vowel-errors factor did so across the entire Hierarchy. That is, as in Field (2005), direction of stress shift was a statistically significant factor only where non-standard stress errors were not simultaneously inducing vowel errors.

In terms of L1 listeners' LD RT (with the 24 most sensitive L1 listeners who made 2+ accurate LDs for the 0 Left category), a within-within-subjects ANOVA with a Greenhouse-Geisser correction also found a significant interaction between the number of vowel errors and direction of stress shift, $F_{(1.36, 31.29)} = 9.07, p = 0.002$. Specifically, the more sensitive L1 English listeners who, at least on occasion, succeeded in making accurate 0 Left "No" LDs significantly slowed from their accurate standard stress "Yes" LDRT baseline to do so, but within the remaining 0 Right – 2 Right Hierarchy categories L1 listeners' accurate "No" LDRTs were statistically equivalent. Interestingly, reverse coding L1 English listeners' 0 Left and 0 Right LDs under the hypothesis that their definition of a "correctly pronounced English word" was, in most cases, broad enough to accommodate non-canonical stress so long as it was instantiated *only* suprasegmentally. In contrast, for both Hierarchy-defined accurate *and* inaccurate LDs with the 0 Right stimuli, L1 English listeners' median reaction times were significantly slower relative to their accurate standard stress LDRT.

In terms of L1 listeners' word identification (WI) accuracy, a within-within-subjects ANOVA with a Greenhouse-Geisser correction found a significant interaction between the number of vowel errors and direction of stress shift, $F_{(1.79, 66.22)} = 7.23$, $p = 0.002$. L1 English listeners were equally accurate in identifying a speaker's intended word whether that word was pronounced with standard stress or 0 Left, 0 Right or 1 Left non-standard stress. It was only at the 1 Right – 2 Left Hierarchy categories that L1 English listeners exhibited significant deterioration in WI accuracy. Thus, direction of stress shift did affect listeners' WI accuracy, but its impact was not stable across the Hierarchy.

In terms of L2 listeners' LD accuracy, a within-within-subjects ANOVA with a Greenhouse-Geisser correction found a significant interaction between the number of vowel errors and direction of stress shift, $F_{(1.91, 57.4)} = 22.21$, $p < 0.001$. As with the L1 listeners, the L2 English listeners' mean LD accuracy was significantly lower for the 0 Left vs. 0 Right non-standard stress pronunciations, but both their 1 Left vs. 1 Right as well as 2 Left vs. 2 Right LD accuracy exhibited no significant difference. As with the L1 listeners, direction of stress shift mattered for the L2 listeners only where non-standard stress errors did not involve vowel errors.

In terms of L2 listeners' LD RT, a within-within-subjects ANOVA with a Greenhouse-Geisser correction found no significant interaction between the number of vowel errors and direction of stress shift, $F_{(1.94, 42.76)} = 2.76$, $p > 0.05$. Although only L1 listeners showed a direction-of-stress-shift-modulated effect in terms of *accurate* LDRT, the L2 listeners did show a direction-of-stress-shift-modulated effect when *in*accurately labeling 0 Left vs. 0 Right non-standard stress pronunciations as instantiating a "correctly pronounced English word."

In terms of L2 listeners' WI accuracy, a within-within-subjects ANOVA with a Greenhouse-Geisser correction found no significant interaction between the number of vowel errors and direction of stress shift, $F_{(1.89, 56.55)} = 1.34$, $p = 0.27$. In regard to the respective main effects of these two factors, however, both were significant. Unsurprisingly, number of vowel errors had the greatest impact, $F_{(1.93, 57.96)} = 17.88$, $p < 0.001$, partial $\eta^2 = 0.37$. Nevertheless, direction of stress shift fell just shy of the 0.14 rule-of-thumb partial η^2 effect size boundary separating "medium" vs. "large" effects, $F_{(1, 30)} = 4.8$, $p = 0.04$, partial $\eta^2 = 0.138$. Specifically, L2 listeners (like L1 listeners) were equally accurate in identifying a speaker's intended word whether that word was pronounced with standard stress or 0 Left, 0 Right or 1 Left non-standard stress. It was only at the 1 Right – 2 Left Hierarchy categories that the L2 listeners (like the L1 listeners) exhibited significant deterioration in WI accuracy. In other words, direction of stress shift *did* affect both listener groups' WI accuracy, but its impact was not stable across the Hierarchy.

In sum, both number of vowel errors and direction of stress shift impacted L1 and L2 English listeners' English word stress error processing. The impact of number of vowel errors and direction of stress shift, however, varied across both Hierarchy categories and dependent variables. The Hierarchy should therefore prove a useful tool for L2 pronunciation teaching and testing, as it provides an easy way of assessing the likely error gravity on any given word stress error.

DISCUSSION

Both L1 and L2 English listeners' word stress error processing largely followed the proposed English Word Stress Error Gravity Hierarchy, with stronger influence from the numbers of vowel changes and weaker influence from the direction of stress shift. As indexed by lexical decision (LD) accuracy, L1 and L2 English listeners frequently struggled to identify as non-standard the mis-stressings containing no vowel errors, even though the experiment instructions explicitly stated they would hear a series of "correctly and incorrectly pronounced English words." That is, listeners in this study identified non-standard stress largely by the presence or absence of vowel errors, just as has been found to be the case by many previous studies (Bond, 1981, 1999; Bond and Small, 1983; Cutler and Clifton, 1984; Cutler, 1986, 2015; Small et al., 1988; Fear et al., 1995; van Leyden and van Heuven, 1996; Cooper et al., 2002; Field, 2005; Cutler et al., 2007; Zhang and Francis, 2010).

However, the L2 English listeners in this study did not follow L1 listeners' performance merely from a lower baseline. Rather, the further an English word stress error fell from the standard stress Hierarchy category, inducing one or more concomitant vowel errors, the more L2 listeners' auditory LD accuracy was hurt relative to L1 listeners. In addition, in terms of reaction time (RT) for accurate LDs, L1 and L2 English listeners both followed the predicted Hierarchy trajectory—but L2 listeners characteristically required from half a second to a full second longer to search their mental lexicons for the pronunciation they had heard in order to make an accurate LD regarding whether a *word spoken in isolation* instantiated a "correctly pronounced English word." Finally, in terms of word identification (WI) accuracy, this study's intelligibility proxy, the L2 listeners again were not working merely from a lower baseline, but rather had higher rates than L1 listeners of unattempted word identification when non-standard stress induced vowel errors.

In other words, even in a non-discourse context, the further a word stress error falls from the English Word Stress Error Gravity Hierarchy's standard stress category, the more likely both L1 and L2 English listeners are to mis-segment the speech string, be led down a garden path forcing additional rounds of mental lexicon lookup, with the result of at least slowed processing (reduced comprehensibility) and perhaps failure to recover the speaker's intended word at all (unintelligibility) (Bond, 1981, 1999; Bond and Small, 1983; Cutler, 2012, 2015; Isaacs and Trofimovich, 2012).

It is true listeners may be able to use context to identify a mispronounced or otherwise unfamiliar word. However, L2 listeners face an uphill battle in taking advantage of context for many reasons. First, the process of acquiring an L1 is the process of becoming highly skilled at attending to the cohort of features most efficiently serving perception and production and becoming equally skilled at suppressing the processing of redundant (or L1-defined "meaningless") features regarding which attention would waste processing resources. Unfortunately, maximally efficient subconscious strategies for processing the L1 so finely honed in the process of childhood language acquisition frequently have just the opposite effect when applied to an L2, where language features matching those one has learned during L1 acquisition to "tune out" are often those on which attention must be focused if the L2 is to be perceived and processed most efficiently (Cutler et al., 1983, 1986, 1992; Otake et al., 1993, 1996a,b; Cutler and Otake, 1994; Cutler, 1997; Kim et al., 2008). This impacts L2 listeners in that their default misapplication of L1 speech segmentation strategies to the L2 stream of speech characteristically renders slow and sometimes completely unsuccessful the word boundary identification that necessarily precedes mental lexicon lookup that necessarily precedes the context-building required for recovering the meaning of mispronounced words! In addition, context is not always particularly helpful for L2 listeners due to their less robust vocabulary and much stronger tendency than L1 speakers to hear phantom words rather than real words (Broersma and Cutler, 2008). Syntactic complexity and cultural unfamiliarity can further exacerbate L2 listeners' difficulty in identifying mispronounced words from context. These issues are compounded when the input contains *multiple* unrecognized forms, as less understood context from which the meaning of unknown forms can be inferred can make guessing from context untenably demanding for not only L2 listeners, but also L1 listeners (Schmitt, 2000; Nation, 2001; Folse, 2004; Field, 2008).

In sum, this study has replicated the findings of Cooper et al. (2002) and Cutler et al. (2007) in its investigation of how listeners map non-standard stress pronunciations onto their (presumably) standard stress mental lexicon prototypes. In the current study, both L1 and L2 English listeners from non-tonal L1s largely processed the suprasegmental correlates of English lexical stress merely as phonetic detail, i.e., as allophonic variation. The *only* correlate of lexical stress that these listeners consistently used to distinguish differences in word stress was not suprasegmental, but segmental. Our phonological understanding of English word stress errors—and our pedagogy—must therefore recognize that the non-canonical use of the suprasegmental correlates of English lexical stress is generally processed as acceptable allophonic variation by both L1 and L2 English listeners. It is only the vowel quality correlate of English lexical stress that is consistently processed categorically. Therefore, the traditional labeling of any non-canonical shift in English word stress as representing an error, regardless of whether the stress shift induces a vowel quality change, is problematic.

Traditionally, any non-canonical shift in English word stress has been treated as an error, without regard to whether the mis-stressing creates a change in vowel quality. However, L1 and L2 English listeners both frequently failed to recognize mis-stressings with no vowel errors as being non-standard. According to the WI accuracy results, neither L1 nor L2 English listeners show any deterioration in intelligibility with 0 Left and 0 Right non-canonical stress pronunciations. Instead, for these advanced L1 and L2 English listeners, non-standard English word stress was largely *defined* by the presence or absence of vowel errors (cf., Bond and Small, 1983; Cutler and Clifton, 1984; Cutler, 1986; Small et al., 1988; Fear et al., 1995; Cooper et al., 2002; Field, 2005). In the Hierarchy categories farthest from standard stress, word stress errors could and did induce vowel errors that significantly reduced intelligibility (Munro and Derwing, 1995).

Teaching Implications

A motivation for this study was to address the importance of word stress in L2 English learning and teaching. There is evidence that the accurate production of English word stress is crucial for intelligibility and comprehensibility. L1 English listeners use stressed syllables to identify the beginnings of words in speech (Cutler and Norris, 1988; Cutler and Butterfield, 1992), and incorrect word stress can cause listeners to completely misunderstand intended words (Benrabah, 1997; Zielinski, 2008). Incorrect word stress has also been found to lead to loss of comprehensibility (Slowiaczek, 1990; Isaacs and Trofimovich, 2012). Because of the critical role of vowel quality in determining stress, stress errors that change vowel quality result in both L1 and L2 English listeners struggling to identify words being spoken (Field, 2005). Unfortunately, due to the preference of English for alternating stressed and unstressed syllables (Liberman and Prince, 1977), non-standard word stress commonly triggers *multiple* vowel quality errors because stress exchange causes ordinarily reduced vowels to become clear, while adjacent ordinarily clear vowels become reduced.

Yet it is also the case that word stress errors do not invariably harm intelligibility and comprehensibility (Levis, 2018). For example, word stress minimal pairs such as INsight/inCITE and *INsult/inSULT* seem not to result in loss of intelligibility. Similarly, when stress errors do not involve changes in vowel quality (e.g., *CONcentrate* said as *concenTRATE*), there is no evidence for loss of intelligibility and little evidence for impaired comprehensibility (Slowiaczek, 1990). Field (2005) and this study have demonstrated that L2 English listeners are similarly affected by misplaced word stress but from a lower baseline due to causes such as the continued use of L1 processing strategies inefficient for English, phantom word activation, and a more limited vocabulary.

As a result, teaching English word stress to L2 learners is at the same time both critical and unimportant. Word stress is critical to intelligibility and comprehensibility in words where stress errors result in a change of vowel quality, with the tendency toward multiple changes in vowel quality likely to have greater impact than a single change. But there have also been influential arguments against teaching word stress that must be addressed. In perhaps the most influential, Jenkins (2002) argues that "the placement of word stress… varies considerably across different L1 varieties of English" resulting in "a need for receptive flexibility" (p. 98) but not productive accuracy. Jenkins' argument involves a number of implicit, problematic claims. First, Jenkins (2000), Jenkins (2002) says that word stress patterns vary considerably, which in light of findings about the importance of word stress for L1 English listeners would suggest there is marginal mutual intelligibility across varieties of English. This is clearly not the case. Berg (1999) says that about 1.7% of multisyllabic words have different stress patterns between AmE and BrE, for a total of about 930 words. Teachers should be aware of these differences, but having over 98% of words stressed similarly represents enormous agreement across varieties.

Second, Jenkins (2000), Jenkins (2002) implies that if L1 speakers of English vary considerably yet understand each other, L2 speakers will also be understood despite variations in word stress patterns. However, McCrocklin (2012) argues that "word stress affects a number of other important features, such as vowel quality and length… These features are listed as core features in Jenkins' proposal and were thus shown within her data to impact intelligibility" (p. 252). Thus, Jenkins' conclusion that word stress "rarely causes intelligibility problems… and where it does so, always occurs in combination with another phonological error" in fact does *not* imply word stress rarely causes intelligibility problems, but rather that since vowels (as well as other aspects of Jenkins' proposed Lingua Franca Core) are strongly contingent on accurate word stress, English word stress is critical for the intelligibility of L2 speech by L2 listeners (Deterding, 2013; Lewis and Deterding, 2018).

Jenkins (2000, p. 39) is right in saying that *comprehensive* analysis of all possible English word stress rules and their exceptions is "far too complex for mental storage by students and teachers alike." However, it is not necessary to teach the full system since a relatively small number of word stress rules cover most academic English vocabulary (Murphy and Kandil, 2004; McCrocklin, 2012; Richards, 2016). As a result, materials developers and teachers of need to know specific underlying word stress patterns which facilitate perception and production of standard English word stress patterns for known and novel words (Nation, 2001; Aitchison, 2012; Cutler, 2012). If these word stress regularities are learned, only the small number of relevant exceptions need to be learned individually—a much more manageable task. Strategies such as condensing L2 learners' exposure to the similar-sounding words from which L1 listeners have acquired their implicit knowledge of English word stress patterns via rhyme-based (e.g., "-tion," "-ssion" and "-cian") pattern "flooding" (see Richards, 2016, p. 257ff.) are particularly recommended for helping produce stream-of-speech automaticity. Key academic word list words such as *ANalyze, aNAlysis,* and *anaLYtical* carry a high semantic load in academic and professional communication. Unfortunately, these are precisely the type of words that are most likely to result in vowel changes when they are mis-stressed, resulting in slowed understanding at best and loss of understanding at worst. Whether the speaker is a graduate student, business executive, healthcare worker, or one of many others whose job depends on clear communication, accurate stress of essential vocabulary can make speech more intelligible, especially when the words are longer than two syllables.

Limitations and Future Directions

This study, in order to examine the effects of word stress errors across the Hierarchy, admittedly used a task that is unlikely to show up in normal communication. As such, it is uncertain how well the results reflect how listeners respond to word stress errors in communicative contexts. The study also is limited in its use of L2 listeners. Although there were enough L2 listeners to statistically test the research questions, the listeners came from a wide variety of L1 backgrounds, and as a result, it is unclear whether the wide variation in L2 scores came from variations in the L1s of the listeners or other unexplored factors.

The study raises questions about how intelligibility and comprehensibility would be affected in more authentic communicative contexts. Our results especially showed that greater numbers of vowel changes led to less efficient understanding and processing by both L1 and L2 listeners. This suggests a correlation between spoken test scores and word stress accuracy, and examinations of spoken language test scores for populations such as international teaching assistants, who frequently use words from the academic wordlist (Coxhead, 2000) may show that types of word stress errors correlate with test scores. Additionally, we assumed that typical word stress errors happened due to analogy with other related words (Guion et al., 2003), but it would be helpful to have better data on the patterns of word stress errors that occur in real speech. A corpus of L2 speech that elicited varied multi-syllabic words would be useful in identifying the extent to which such analogical errors occur.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board, Iowa State University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

This article is based on the dissertation research of MG. As such, her work forms the foundation of the research. JL is her dissertation supervisor. In this paper, he contributed most heavily to reformulating the introduction, literature review, discussion, and teaching implications. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.628780/full#supplementary-material

REFERENCES

Aitchison, J. (2012). *Words in the Mind: An Introduction to the Mental Lexicon*, 4th ed. Malden, MA: John Wiley & Sons.

Anderson-Hsieh, J., Johnson, R., and Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Lang. Learn*. 42, 529–555. doi: 10.1111/j.1467-1770.1992.tb01043.x

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behav. Res. Methods* 39, 445–459. doi: 10.3758/BF03193014

Barca, L., Burani, C., and Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behav. Res. Methods Instrum. Comput*. 34, 424–434. doi: 10.3758/BF03195471

Benrabah, M. (1997). Word stress - a source of unintelligibility in English. *IRAL Int. Rev. Appl. Linguis. Lang. Teach*. 35:157. doi: 10.1515/iral.1997.35.3.157

Berg, T. (1999). Stress variation in British and American English. *World Englishes* 18, 123–143. doi: 10.1111/1467-971X.00128

Bond, Z. S. (1981). Listening to elliptic speech: pay attention to stressed vowels. *J. Phon*. 9, 89–96. doi: 10.1016/S0095-4470(19)30929-5

Bond, Z. S. (1999). *Slips of the Ear: Errors in the Perception of Casual Conversation, 1st Edn.* San Diego, CA: Academic Press.

Bond, Z. S., and Small, L. H. (1983). Voicing, vowel, and stress mispronunciations in continuous speech. *Percept. Psychophys*. 34, 470–474. doi: 10.3758/BF03203063

Broersma, M., and Cutler, A. (2008). Phantom word activation in L2. *System* 36, 22–34. doi: 10.1016/j.system.2007.11.003

Brown, G. (1990). *Listening to Spoken English, 2nd Edn.* London: Longman.

Brysbaert, M., New, B., and Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behav. Res. Methods* 44, 991–997. doi: 10.3758/s13428-012-0190-4

Brysbaert, M., Warriner, A. B., and Kuperman, V. (2014). Concreteness ratings for 40,000 generally known English word lemmas. *Behav. Res. Methods* 46, 904–911. doi: 10.3758/s13428-013-0403-5

Cambridge University Press (2015). *Cambridge Dictionaries Online*. Available online at: http://dictionary.cambridge.org/us/ (accessed March, 2016).

Celce-Murcia, M., Brinton, D., and Goodwin, J. M. (2010). *Teaching Pronunciation: A Course Book and Reference Guide*. New York, NY: Cambridge University Press.

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. Cambridge, MA: The MIT Press.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev*. 108, 204–256. doi: 10.1037/0033-295X.108.1.204

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

Coxhead, A. (2000). A new academic word list. *TESOL Q*. 34, 213–238. doi: 10.2307/3587951

Cutler, A. (1986). Forbear is a homophone: lexical prosody does not constrain lexical access. *Lang. Speech* 29, 201–220. doi: 10.1177/002383098602900302

Cutler, A. (1997). The syllable's role in the segmentation of stress languages. *Lang. Cogn. Process*. 12, 839–846. doi: 10.1080/016909697386718

Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words*. Cambridge, MA: The MIT Press.

Cutler, A. (2015). "Lexical stress in English pronunciation," in: *The Handbook of English Pronunciation,* eds M. Reed and J. M. Levis (Hoboken, NJ: John Wiley & Sons, Inc), 106–124. Available online at: http://onlinelibrary.wiley.com/doi/10.1002/9781118346952.ch6/summary

Cutler, A., and Butterfield, S. (1992). Rhythmic cues to speech segmentation: evidence from juncture misperception. *J. Memory 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. (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 (London: Lawrence 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. Memory 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 Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *J. Exp. Psychol. Hum. Percept. Perform*. 14, 113–121. doi: 10.1037/0096-1523.14.1.113

Cutler, A., and Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *J. Memory Lang.* 33, 824–844. doi: 10.1006/jmla.1994.1039

Cutler, A., Wales, R., Cooper, N., and Janssen, J. (2007). "Dutch listeners' use of suprasegmental cues to English stress," in: *Proceedings of the 16th International Congress of Phonetics Sciences (ICPhS 2007)*, eds J. Trouvain and W. J. Barry (Dudweiler: Pirrot), 1913–1916.

Dalton, C., and Seidlhofer, B. (1994). "Is pronunciation teaching desirable? Is it feasible?," in: *Proceedings of the 4th International NELLE Conference,* ed T. S. Sebbage (Hamburg, Germany: NELLE), 14–15.

Dauer, R. M. (2005). The lingua franca core: a new model for pronunciation instruction? *TESOL Q.* 39, 543–550. doi: 10.2307/3588494

Dechert, H. W. (1984). "Individual variation in speech," in: *Second Language Productions,* eds H. W. Dechert, D. Möhle, and M. Raupach (Tübingen, Germany: Gunter Narr Verlag), 156–185.

Deterding, D. (2013). *Misunderstandings in English as a Lingua Franca: An Analysis of ELF Interactions in South-East Asia*, Vol. 1. Berlin: Walter de Gruyter.

Dickerson, W. B. (1989). *Stress in the Speech Stream: The Rhythm of Spoken English*. Urbana: University of Illinois Press.

Dupoux, E., Sebastián-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

Fear, B. D., Cutler, A., and Butterfield, S. (1995). The strong/weak syllable distinction in English. *J. Acoust. Soc. Am*. 97:1893. doi: 10.1121/1.412063

Field, J. (2005). Intelligibility and the listener: the role of lexical stress. *TESOL Q.* 39, 399–423. doi: 10.2307/3588487

Field, J. (2008). Revising segmentation hypotheses in first and second language listening. *System* 36, 35–51. doi: 10.1016/j.system.2007.10.003

Folse, K. S. (2004). *Vocabulary Myths: Applying Second Language Research to Classroom Teaching*. Ann Arbor, MI: University of Michigan Press.

Guion, S. G. (2005). Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Stud. Second Lang. Acquis*. 27, 503–533. doi: 10.1017/S0272263105050230

Guion, S. G., Clark, J. J., Harada, T., and Wayland, R. P. (2003). Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Lang. Speech* 46, 403–426. doi: 10.1177/00238309030460040301

Guion, S. G., Harada, T., and Clark, J. J. (2004). Early and late Spanish–English bilinguals' acquisition of English word stress patterns. *Bilingualism Lang. Cogn*. 7, 207–226. doi: 10.1017/S1366728904001592

Isaacs, T., and Trofimovich, P. (2012). Deconstructing comprehensibility: identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Stud. Second Lang. Acquis*. 34, 475–505. doi: 10.1017/S0272263112000150

Jansz, D. (n.d.). *IPA Phonetic Transcription of English Text*. Available online at: http://lingorado.com/ipa/ (accessed November 4, 2015).

Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: OUP Oxford.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Appl. Linguist*. 23, 83–103. doi: 10.1093/applin/23.1.83

Kenworthy, J. (1987). *Teaching English Pronunciation*. London: Longman.

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

Kuperman, V., Estes, Z., Brysbaert, M., and Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *J. Exp. Psychol. Gen*. 143, 1065–1081. doi: 10.1037/a0035669

Lee, G., Shin, D. J., and Garcia, M. T. M. (2019). Perception of lexical stress and sentence focus by Korean-speaking and Spanish-speaking L2 learners of English. *Lang. Sci*. 72, 36–49. doi: 10.1016/j.langsci.2019.01.002

Levis, J. M. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. Cambridge: Cambridge University Press.

Lewis, C., and Deterding, D. (2018). Word stress and pronunciation teaching in English as a Lingua Franca contexts. *CATESOL J*. 30, 161–176.

Liberman, M., and Prince, A. (1977). On stress and linguistic rhythm. *Linguist. Inq*. 8, 249–336.

Lo, S., and Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Front. Psychol*. 6:1171. doi: 10.3389/fpsyg.2015.01171

Maczuga, P., O'Brien, M. G., and Knaus, J. (2017). Producing lexical stress in second language German. *Die Unterrichtspraxis Teach. German* 50, 120–135. doi: 10.1111/tger.12037

McCrocklin, S. (2012). "The role of word stress in English as a lingua franca," in: *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference,* eds J. Levis and K. LeVelle (Ames, IA: Iowa State University), 249–256.

Merriam-Webster (2015). *Merriam-Webster*. Available online at: http://www.merriam-webster.com/ (accessed March, 2016).

Munro, M. J., and Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn*. 45, 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x

Murphy, J., and Kandil, M. (2004). Word-level stress patterns in the academic word list. *System* 32, 61–74. doi: 10.1016/j.system.2003.06.001

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge, United Kingdom: Cambridge University Press.

Ortega-Llebaria, M., Gu, H., and Fan, J. (2013). English speakers' perception of Spanish lexical stress: context-driven L2 stress perception. *J. Phonetics* 41, 186–197. doi: 10.1016/j.wocn.2013.01.006

Otake, T., Hatano, G., Cutler, A., and Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *J. Memory Lang*. 32, 258–278. doi: 10.1006/jmla.1993.1014

Otake, T., Hatano, G., and Yoneyama, K. (1996a). "Speech segmentation by Japanese listeners," in: *Phonological Structure and Language Processing: Crosslinguistic Studies,* eds T. Otake and A. Cutler (Berlin: Mouton), 183–201.

Otake, T., Yoneyama, K., Cutler, A., and Lugt, A., van der. (1996b). The representation of Japanese moraic nasals. *J. Acoust. Soc. Am*. 100, 3831–3842. doi: 10.1121/1.417239

Oxford University Press (2015). *Oxford Dictionaries*. Available online at: http://www.oxforddictionaries.com/us/ (accessed March, 2016).

Richards, M. G. (2016). *Not All Word Stress Errors Are Created Equal: Validating an English Word Stress Error Gravity Hierarchy* (Doctoral dissertation) Iowa State University.

Schmitt, N. (2000). *Vocabulary in Language Teaching.* Cambridge: Cambridge University Press.

Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Lang. Speech*. 33, 47–68. doi: 10.1177/002383099003300104

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

Soto-Faraco, S., Sebastián-Gallés, N., and Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *J. Memory Lang*. 45, 412–432. doi: 10.1006/jmla.2000.2783

van Heuven, W. J. B., Mandera, P., Keuleers, E., and Brysbaert, M. (2014). SUBTLEX-UK: a new and improved word frequency database for British English. *Q. J. Exp. Psychol*. 67, 1176–1190. doi: 10.1080/17470218.2013.850521

van Leyden, K., and van Heuven, V. (1996). Lexical stress and spoken word recognition: Dutch vs. English. *Linguist. Netherlands* 13, 159–170. doi: 10.1075/avt.13.16ley

Whelan, R. (2008). Effective analysis of reaction time data. *Psychol. Rec*. 58, 475–482. doi: 10.1007/BF03395630

Zhang, Y., and Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *J. Phonet*. 38, 260–271. doi: 10.1016/j.wocn.2009.11.002

Zielinski, B. W. (2008). The listener: no longer the silent partner in reduced intelligibility. *System* 36, 69–84. doi: 10.1016/j.system.2007.11.004

Conflict of Interest: 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.

The handling editor declared a past co-authorship with one of the authors JL.

frontiers
in Communication

GENERAL COMMENTARY
published: 07 October 2021
doi: 10.3389/fcomm.2021.721053

Commentary: "Vowel Quality and Direction of Stress Shift in a Predictive Model Explaining the Varying Impact of Misplaced Word Stress: Evidence From English" and "Exploring the Complexity of the L2 Intonation System: An Acoustic and Eye-Tracking Study"

Alison McGregor *

McGraw Center for Teaching and Learning-English Language Program, Princeton University, Princeton, NJ, United States

Keywords: L2 pronunciation, prosodic structure, word stress, English intonation, vowel quality

OPEN ACCESS

Edited by:
Oliver Niebuhr,
University of Southern Denmark,
Denmark

Reviewed by:
Mary Grantham O'Brien,
University of Calgary, Canada

****Correspondence:***
Alison McGregor
alison.mcgregor@princeton.edu

Specialty section:
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: *05 June 2021*
Accepted: *13 September 2021*
Published: *07 October 2021*

Citation:
McGregor A (2021) Commentary: "Vowel Quality and Direction of Stress Shift in a Predictive Model Explaining the Varying Impact of Misplaced Word Stress: Evidence From English" and "Exploring the Complexity of the L2 Intonation System: An Acoustic and Eye-Tracking Study". Front. Commun. 6:721053. doi: 10.3389/fcomm.2021.721053

A Commentary on

Vowel Quality and Direction of Stress Shift in a Predictive Model Explaining the Varying Impact of Misplaced Word Stress: Evidence From English
by Monica Ghosh and John M. Levis (2021). Front. Commun. 6:56. doi: 10.3389/fcomm.2021.628780

Exploring the Complexity of the L2 Intonation System: An Acoustic and Eye-Tracking Study
by Di Liu and Marnie Reed (2021). Front. Commun. 6:627316. doi: 10.3389/fcomm.2021.627316

The aim of this commentary is to propose word prosody training as scaffolding for learning the English intonation system. Drawing on Ghosh and Levis (2021), I discuss the pedagogical implications of research on vowel quality in relation to word stress instruction and the nested nature of vowels within syllables, which make up prosodic words. In light of Liu and Reed, (2021) findings on the structural complexity of intonation (i.e., interrelated and interacting components), I hope to demonstrate the applicability of L2 phonology research to improve prosodic structure pedagogy in the context of L2 English pronunciation.

According to Ghosh and Levis (2021), word stress errors that introduce concomitant vowel errors highlight a critical role played by vowel quality in listener processing of multi-syllabic words. Pedagogically, how should these findings inform classroom practices? First, does the finding on vowel quality establish a stronger case for prerequisite vowel training in order to promote word-level intelligibility? If so, would increased emphasis on vowel quality entail spending more time on the perception of clear versus reduced English vowels and/or the production mechanisms often missing in students' articulatory settings to make English vowels, especially the reduced vowel (i.e., the schwa)? And what about the need to address relative length—by which I mean English vowel lengths contrasted with the learners' L1 vowel lengths? Based on my own teaching experience, it is quite apparent that without explicit instruction on similarities/differences and the relative nature of vowel length, L2 learners are often ill-equipped to recognize these subtleties.

TABLE 1 | Proposed Prosodic Structure Pathway.

Prosodic Structure Prerequisites	**Description**	**Rationale**
Articulatory settings	Provide instruction on the default mouth position for English; contrast L1 versus English settings; practice perception and production of the schwa	The settings enable the production of the schwa, thus supporting word-level rhythm, which underlies overall English rhythm
Vowels	Highlight length and relative length (L1 versus English vowel length); practice perception and production of vowel quality	According to L1 listener data, word- level intelligibility is influenced by central/reduced vowels
Prosodic Structure Scaffolding		
Word prosody	For one-syllable words, practice perception and production of: • segmentals with an emphasis on vowel quality For two-syllable words, focus perception and production on: • vowel quality (clear versus reduced) • word-level prosody (including pitch range, pitch level, and duration) For three or more syllable words, target: • words as the unit of analysis (as a prerequisite for thought groups) • vowel quality (clear versus reduced) • word-level prosody (including pitch range, pitch level, duration, and pitch contours across syllables)	Structural complexity of English intonation is problematic for L2 Mandarin-English speakers; use word prosody to scaffold structural complexity of intonation learning
Phrase and sentence stress	Focus on: • thought groups as unit of analysis • primary sentence stress (pitch range, pitch level, and duration) • rhythm (clear versus reduced vowels) • pitch contour(s)	Evidence shows primary sentence stress contributes to the degree of intelligibility
Utterance-level prosody	Practice: • thought groups (or paragraph) as unit of analysis • primary sentence stress (pitch range, pitch level, and duration) • rhythm (clear versus reduced vowels) • pitch contours	Prosody plays a significant role in encoding meaning

Second, from a phonetics/phonology crossroad perspective, what is the relationship between vowel length inherent in vowel quality and the prosodic cue of duration created by a vowel nested within a stressed syllable of a multi-syllabic word or in a sentence? This relationship does highlight the importance of vowel quality training; however, it only addresses length/duration, omitting features related to pitch. In order for learners to develop word-level prosody as a foundation for utterance-level prosody, some traditionally taught stress characteristics such as pitch range/level/change are required. I would like to argue, therefore, that word-level prosody training (as opposed to just vowel quality or just stress characteristics) provides an opportunity for students to begin intonation skill development as part of English prosody learning.

This brings me to Liu and Reed (2021). As they indicate, the signaling of contrastive and implicational information by Mandarin-English L2 speakers relies more on the acoustic feature of intensity, while L1 English speakers use pitch range, pitch level, and duration. From an assessment perspective, these findings are not surprising, since speech rater comments on L2 Mandarin-English speakers' speech often include descriptions such as "plodding", "choppy", or "tone-like". This is certainly in part due to the use of intensity more than duration and pitch movement for signaling stress. From a classroom research perspective, however, I am curious about the training of the Mandarin-English L2 speaking participants. Had they ever had perception or production instruction on either word- or sentence-level stress in which the difference between tone and English stress was made explicit? Did they receive feedback on their use of intensity versus the use of pitch (range and level) and duration for word stress or primary sentence stress? Anecdotally, a majority of my L1 Mandarin speaking graduate students in the past 15 years have mostly lacked explicit instruction on word- or sentence-level stress. Similarly, there is typically a lack of structural or functional complexity awareness on their part. Yet, most make considerable progress in their production with explicit instruction, supportive feedback, and scaffolded practice.

Let us turn now to the broader implications of L2 phonology research for L2 pronunciation pedagogy. If we accept language as a complex system (Larsen-Freeman, 2017) and specifically, the interconnectedness of components or parts, then the nested nature of prosodic structure presents itself as a pedagogical pathway. In other words, by embracing interconnectedness—vowels are nested in syllables, which are nested in words, then nested in intonation units and ultimately, utterances (Fox, 2000)—instruction can move beyond the segmental versus suprasegmental debate (Zielinski, 2015). This would also address the Liu and Reed finding that their L1 Mandarin English learners largely failed at producing contrastive/implicational intonation. Accordingly, word prosody training during vocabulary instruction can be used to scaffold the structural complexity of intonation required as learners' proficiency levels advance. This approach paves the way for phrase- and sentence-level practice before moving on to contextualized utterance- and discourse-level practice. With structural complexity integrated in such a way, more time would be available for addressing the functions of intonation.

Initially inspired by readings on prosodic hierarchy (Nespor and Vogel, 1986) and decades of grappling with learner needs, as well as in light of the research findings mentioned above, I set forth in **Table 1** a proposed prosodic structure pathway for teaching and learning English prosody. For prerequisite skills, the pathway starts with articulatory settings and vowels. Essentially, the process includes an emphasis on vowel quality, but continues to integrate word stress characteristics with the rationale that word-level prosody is a training ground for the structural complexity of English intonation. Although empirical research is needed to test the efficacy of this approach, further study is warranted based on successful classroom outcomes with high intermediate to low advanced L2 Mandarin-English speakers.

Both Ghosh and Levis (2021) and Liu and Reed (2021) found L1 and L2 differences. The former identified a lower baseline and greater variation related to lexical stress errors in L2 listeners compared to L1 listeners. The latter found feature differences for encoding contrastive and implicational information. These findings reinforce the value in and need for cross-linguistic comparison studies investigating the perception and production of target features. Pedagogically, there remains an opportunity to improve the efficacy of pronunciation pedagogy through strategic application of the L2 phonology knowledgebase. May the future foster both.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

This publication was supported by the Princeton University Library Open Access Fund. I would also like to thank the reviewers of this commentary for their close and thoughtful reading.

REFERENCES

Fox, A. (2000). *Prosodic Features and Prosodic Structures*. Oxford: Oxford University Press.

Ghosh, M., and Levis, J. M. (2021). Vowel Quality and Direction of Stress Shift in a Predictive Model Explaining the Varying Impact of Misplaced Word Stress: Evidence from English. *Front. Commun.* 6, 628780. doi:10.3389/fcomm.2021.628780

Larsen-Freeman, D. (2017). "Chapter 1. Complexity Theory," in *Complexity Theory and Language Development: In Celebration of Diane Larsen-Freeman*. Editors L. Ortega and Z. Han (Amsterdam, Netherlands: John Benjamins), 11–50. doi:10.1075/lllt.48.02lar

Liu, D., and Reed, M. (2021). Exploring the Complexity of the L2 Intonation System: An Acoustic and Eye-Tracking Study. *Front. Commun.* 6, 627316. doi:10.3389/fcomm.2021.627316

Nespor, M., and Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications.

Zielinski, B. (2015). "The Segmental/suprasegmental Debate," in *The Handbook of English Pronunciation*. Editors M. Reed and J. Levis (Hoboken, NJ, USA: Wiley), 397–412. doi:10.1002/9781118346952.ch22

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Communication

CONCEPTUAL ANALYSIS
published: 14 April 2021
doi: 10.3389/fcomm.2021.639889

Evidence-Based Design Principles for Spanish Pronunciation Teaching

Laura Colantoni[1], *Paola Escudero*[2,3]*, *Victoria Marrero-Aguiar*[4] *and Jeffrey Steele*[5]

[1]Department of Spanish and Portuguese, University of Toronto, Toronto, ON, Canada, [2]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, NSW, Australia, [3]Australian Research Council Centre of Excellence for the Dynamics of Language, Australian National University, Canberra, ACT, Australia, [4]Departamento de Lengua Española y Lingüística General, Universidad Nacional de Educación a Distancia, Madrid, Spain, [5]Department of Language Studies, University of Toronto Mississauga, Mississauga, ON,Canada

In spite of the considerable body of pedagogical and experimental research providing clear insights into best practices for pronunciation instruction, there exists relatively little implementation of such practices in pedagogical materials including textbooks. This is particularly true for target languages other than English. With the goal of assisting instructors wishing to build effective evidence-based instructional practices, we outline a set of key principles relevant to pronunciation teaching in general, illustrated here via Spanish in particular, drawing on previous pedagogical research as well as methods and findings from experimental (applied) linguistics. With the overall goal of enabling learners to move toward greater intelligibility, these principles include the importance of perceptual training from the onset of learning, a strong prosodic component, the use of contextualized activities, and a focus on segmental and prosodic phenomena with a high functional load as well as those that are shared across target language varieties. These principles are then illustrated with innovative perception and production exercises for beginner, university-level learners of Spanish. We conclude with a discussion of ways in which the pedagogical principles exposed here can be extended beyond the production of individual activities to the design of a broader pronunciation curriculum.

Keywords: pronunciation instruction, focus on perception, Spanish, contextualized learning, segments, prosody, functional load, evidence-based principles

Edited by:
*Mary Grantham O'Brien,
University of Calgary, Calgary, AB, Canada*

Reviewed by:
*Angela George,
University of Calgary, Calgary, AB, Canada*
*Ashley Roccamo,
American University, Washington, DC, United States*

Correspondence:
*Paola Escudero
paola.escudero@westernsydney.edu.au*

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *10 December 2020*
Accepted: *03 February 2021*
Published: *14 April 2021*

Citation:
Colantoni L, Escudero P, Marrero-Aguiar V and Steele J (2021) Evidence-Based Design Principles for Spanish Pronunciation Teaching. Front. Commun. 6:639889. doi: 10.3389/fcomm.2021.639889

INTRODUCTION

With a few exceptions (e.g., Gilbert, 2005; certain recent methods, see *Profile of Widely Used Textbooks in Europe* section), L2 pronunciation textbooks typically mirror traditional introductory phonetics textbooks, adopting a structure-based organization (consonants and vowels followed by prosody). Moreover, instruction often involves decontextualized, word-level exercises (e.g., minimal pairs) with a strong focus on accent reduction, that is, on helping learners to become more native-like. Such practices run counter to the now well-established general principles that pronunciation instruction should focus first and foremost on increased intelligibility[1] as opposed to native-like accuracy (e.g., Munro and Derwing, 1995; Levis, 2005; Munro and Derwing, 2011; Levis, 2018; Levis,

[1] Following Levis (2005); Levis (2018); Levis (2020), we use 'intelligibility' to refer to both the ease and accuracy with which a speaker's interlocutor understands what is being said. This use collapses the distinction sometimes made between 'comprehensibility' and 'intelligibility' (e.g., Munro and Derwing, 1995).

2020) and that prosody merits equal attention to segmentals (e.g., Field, 2005; Gilbert, 2008). Clearly, there is work to be done to help the creators of pronunciation instructional materials as well as instructors in general benefit more widely from the insights provided by pedagogical and experimental research[2]. In the case of instructors of languages other than English, this gulf is arguably wider.

To assist instructors interested in developing effective pronunciation materials, we set two general goals. First, following the call in Derwing and Munro (2015)[3], drawing on both pedagogical research (e.g., Derwing and Munro, 2015; Sicola and Darcy, 2015; Levis, 2018; Rao, 2019) as well as the findings of experimental (applied) linguistics, we propose a set of five evidence-based principles applicable to the teaching of any language that are capable of enabling learners to move toward greater intelligibility; some of these are well established, others are new. As concerns our second goal, with the aim of expanding the discussion of such principles beyond English, we illustrate these principles via another widely spoken and taught language, Spanish. The first principle proposes that, on the assumption that perception leads production (e.g., Flege, 1995; Escudero, 2009; Baese-Berk, 2019; Goodin-Mayeda, 2019), initial instruction should involve considerable perception-based activities. Moreover, such activities should go beyond traditional listen-repeat tasks typical of the audiolingual method and draw on recent findings from experimental and classroom-based research (e.g., the rhythmic beat gestural training in Gluhareva and Prieto, 2017). Second, given that intelligibility and fluency are intimately related (e.g., Levis, 2005; Saito, 2011; Lin and Francis, 2014), initial instruction should incorporate larger prosodic structures such as rhythm and intonation as opposed to focusing on segments alone (de la Mota, 2019). Third, even with lower proficiency learners, practice should be contextualized in keeping with the principle that language should be learned and practised in the same contexts as in normal communicative use (e.g., Lightbown, 2007; Mora and Levkina, 2017). Given the overarching focus on intelligibility, the fourth and fifth evidence-based principles espoused are that greater time-on-task should be given to features that have a higher functional load (e.g., Brown, 1988; Munro and Derwing, 2006; Dupoux et al., 2008; Derwing and Munro, 2014), and that a primary focus should be placed on segmental and consonantal features shared by (the majority of) the varieties of the target language. Features that do not impede intelligibility should be left for instruction targeting more advanced learners.

In the remainder of this article, we first outline and motivate the five core evidence-based principles outlined above that we argue should be central to the teaching of the pronunciation of any language (*Evidence-Based Principles of Pronunciation Curriculum Design* section). To illustrate the disconnect between evidence-based principles and many actual pronunciation teaching materials, we then turn to an analysis of the most commonly used Spanish pronunciation textbooks in North America and Europe (*Assessment of Current Practices in Spanish Pronunciation Textbooks* section). We highlight that, although efforts are made to expose learners to dialectal variation and contextualized materials (e.g., Morgan, 2010; Schwegler and Ameal-Guerra, 2019), most textbooks follow the traditional structure of introductory phonetics textbooks, circumscribe the teaching of prosody to a single chapter, and provide limited evidence for dialogue with current (applied) linguistic research. We then turn to demonstrating how the guiding principles can shape the creation of innovative materials via perception and production activities targeting beginner, university-level[4] learners (*Putting Principles Into Practice: Sample Perception and Production Activities* section). We conclude with a discussion of how to extend these principles to the design of a broader pronunciation curriculum.

EVIDENCE-BASED PRINCIPLES OF PRONUNCIATION CURRICULUM DESIGN

In this section, we review both theoretical and experimental evidence for the five design principles espoused in the current framework.

The Importance of Perception-Focused Instruction

Although L2 pronunciation research tends to focus more on learners' L2 speech production, the wide availability of cross-linguistic speech perception research has led to perception-based explanations for L2 pronunciation difficulties (Colantoni et al., 2015). Specifically, the most influential models that aim at explaining learner's difficulties in attaining native-like L2 speech, namely the Perceptual Assimilation Model (PAM; Best, 1995; Best and Tyler, 2007), the Speech Learning Model (SLM; Flege, 1995; Flege and Bohn, 2021), and the Second Language Linguistic Perception Model (L2LP; Escudero, 2005; van Leussen and Escudero, 2015; Elvin and Escudero, 2019; Yazawa et al., 2020), are perception-based. All of these models adopt the assumption that, in the same way that young children's perceptual knowledge overwhelmingly surpasses their ability to produce their first words, L2 learners' abilities are greater in perception than in pronunciation. Two of these theoretical models, namely the SLM and L2LP, propose and demonstrate with empirical evidence that L2 perception accuracy is a precursor to L2 production accuracy (Flege, 1995; Flege et al., 1997; Escudero, 2005; Escudero, 2007).

[2] The disconnect between pedagogical and experimental research and instructional materials is not unique to pronunciation but, arguably, characteristic of much second language teaching.

[3] Derwing and Munro (2015) make the most elaborated claim re the need for evidence-based instruction, a call made elsewhere including for Spanish pronunciation (Lord and Fionda, 2013: 525).

[4] The activities presented are arguably well suited for beginners learning in any instructed context. We focus on university-level learners given that this is the population with which we are most experienced.

Recent lab-based studies have shown that perception-based training indeed has positive effects on L2 production but not vice versa (Baese-Berk and Samuel, 2016; Baese-Berk, 2019). Moreover, classroom-based studies comparing the efficacy of perception- and production-based methods for L2 production training have concluded that perception-based methods yield the best results for both segmental and suprasegmental features (see the meta-analysis in Lee et al., 2020). With respect to L2 Spanish pronunciation in particular, Goodin-Mayeda's (2019) proposal, which follows perception-based L2 speech models, emphasizes the connection between perception and production and the prominent role of perception in L2 Spanish pronunciation learning. In terms of classroom practice, perception training should include a key role for explicit instruction where "learners' attention must be explicitly drawn to the differences in the L2 and the L1 via form-focused instruction (FFI), and errors in the learners' L2 production would benefit from explicit corrective feedback" (Lee et al., 2020, p.3). However, other studies have shown that methods that rely on "implicit" or "ambiguous" learning without corrective feedback also result in significant phonetic learning at the segmental and word levels (Wanrooij et al., 2013; Escudero and Williams, 2014; Ong et al., 2017; Tuninetti et al., 2020), although for very difficult L2 contrasts, "attentive" listening (with a task that draws attention to auditory stimuli), rather than "passive" (with no task performed while listening to an array of sounds), yields better results (Ong et al., 2015). In our proposal for perception-based L2 Spanish pronunciation activities (*Production of Spanish /a e o/* section), we suggest using explicit and implicit methods that emphasize the important role of both prosody and contextualized speech, as per our next two design principles.

The Importance of Prosody

Two commonalities of much L2 pronunciation instruction are an (initial) primary focus on segments, and practice with isolated, often short, words (*Assessment of Current Practices in Spanish Pronunciation Textbooks* section for discussion with reference to Spanish textbooks; see e.g., Gilbert (2005); Gilbert (2008) for illustrations of alternative practices). Such a practice is understandable if one wishes to make materials accessible and "doable", at least when working with lower-proficiency learners. However, a primary focus on individual words goes against the now well-established importance of the teaching and learning of prosody to pronunciation learning.

Numerous studies have demonstrated that, equally or sometimes more so than segmentals, prosody is relevant to improving all dimensions of L2 speech, including intelligibility (e.g., Anderson-Hsieh and Koehler, 1988; Derwing et al., 1998; Field, 2005; Warren et al., 2009; Isaacs and Trofimovich, 2012), accentedness (e.g., Anderson-Hsieh et al., 1992; Kang, 2010; Polyanskaya et al., 2017), and perceived fluency (e.g., Derwing et al., 1998; Saito et al., 2018).

A focus on prosody may also lead to improvement with segmentals. Indeed, cross-linguistically, for many phonological and phonetic phenomena, there is an interaction between the two. For example, in English, vowel quality is conditioned by lexical stress: vowels are reduced and produced with a schwa-like quality in unstressed syllables (e.g., Fry, 1965; Delattre, 1969; Beckman, 1986) with vowel reduction being a cue to stress (e.g., Beckman, 1986; Howell, 1993) and important to establishing rhythm (e.g., Roach, 1982). In the case of Spanish pronunciation instruction, Piñeros (2019) provides another example of the relevance of considering segmental-prosody interactions, arguing for the importance of teaching nasal assimilation using prosody, since nasal assimilation is sensitive to prosodic constituency: in particular, it applies within the intonational phrase and is blocked at a prosodic break. As Zielinksi (2015) highlights, "the segmental/suprasegmental debate is based on a false dichotomy." (*p*. 409).

Accordingly, with the goal of improving learners' intelligibility, pronunciation activities should regularly and consistently incorporate larger prosodic structures than individual words from the very onset of learning (e.g., Kjellin, 1999; Gilbert, 2005; de la Mota, 2019 as well as *Production of Spanish /a e o/* section for discussion and illustrations of best practices). Research on L2 prosody has also demonstrated that it is possible to determine which features will contribute more to intelligibility vs. accentedness (e.g., Kang, 2010; Polyanskaya et al., 2017), including how particular prosodic features interact with learner proficiency level (e.g., Anderson-Hsieh and Koehler, 1988; Li and Post, 2014; Saito et al., 2016; Saito et al., 2018).

The Importance of Contextualized Speech

As mentioned previously, pronunciation instruction often involves decontextualized speech, with listening and production exercises focusing on isolated words or short phrases. There are three reasons to argue for a contrasting approach involving activities that place a higher priority on contextualized speech.

First, in keeping with the general principle that learners should be provided with authentic input (e.g., Villegas Rogers and Medley, 1988; Gilmore, 2007), instructional materials should reflect natural speech, which is contextualized by nature (e.g., Bowen, 1972; Isaacs, 2009 for exposition of this claim in the context of L2 pronunciation instruction). It is important to keep in mind that context has effects on the particular phonetic segmental and prosodic variants that learners must come to approximate. As mentioned in *The Importance of Prosody* section, Spanish nasal assimilation is sensitive to prosodic constituency, occurring within but not across intonational groups (e.g., *llega*[ŋk]*ansados* "they arrive tired" vs. *cuando llega*[n#k]*omen* "when they arrive, they eat"; Piñeros, 2019). Arguably, of all the pronunciation aspects that a textbook should cover, intonation is the one that is most sensitive to contextual aspects given that its functions range from expressing emphasis to indicating question type.

A second pedagogical principle that supports the call for the use of contextualized speech is that language should be learned and practised in the same contexts as those encountered in normal communicative use (e.g., Lightbown, 2007; Mora and Levkina, 2017). This principle is consistent with the goal of helping learners to acquire the automatized linguistic knowledge necessary for fluent speech (e.g., Gatbonton and Segalowitz, 1988) and the combined form-meaning-focused activities advocated for in communicative frameworks for pronunciation teaching (e.g., Celce-Murcia et al., 2010; Sicola and Darcy, 2015).

Finally, in terms of the results of experimental research, numerous studies have shown that instruction with a prosodic, as opposed to segmental, focus can lead to relatively superior performance (e.g., Derwing et al., 1998; Hardison, 2005). For example, Derwing et al. (1998) compared the effects of instruction with a segmental vs. prosodic focus, the latter targeting features such as lexical stress, intonation, and speech rate. While both types of instruction resulted in improvements in their intermediate-proficiency English-speaking learners' comprehensibility and accentedness with read sentences, with narratives, the global focus alone led to improvements in comprehensibility and fluency.

A Focus on Features With High Functional Load

Many researchers have proposed that greater time-on-task should be given to features that have a higher functional load (e.g., Brown, 1988; Munro and Derwing, 2006; Dupoux et al., 2008; Derwing and Munro, 2014). [Martinet (1978): 129] defines functional load as "the number of [lexical] pairs that would be complete homonyms once the opposition is lost". For example, in Spanish, /s/ and /θ/ (e.g., *casó* /kaso/ "he married" vs. *cazó* /kaθo/ "he hunted") would be much more likely to fuse into one phoneme than /p-b/. Indeed, the *Minimal Pair Finder* tool (Mairano and Calabrò, 2016, http://phonetictools.altervista.org/minimalpairfinder) presents 724 minimal pairs involving /s/-/θ/ vs. 3,463 minimal pairs for /p/-/b/.

In the context of deciding what to teach, the logic behind considering functional load is that not all pronunciation aspects are equally important for intelligibility at each stage of development (e.g., Brown, 1988). For example, it is important to begin by teaching contrasts that are frequent in the language (*Targeting Features and Segments Shared by the Majority of the Varieties of the Target Language* section), such as those involving Spanish vowels, and then progress to those that are less frequent, such as the tap-trill contrast (e.g, ['kaɾo] "expensive" vs. ['karo] "car"). As acknowledged by Brown (1988), measuring functional load is not a trivial task. If two sounds are contrastive, one must ask how many minimal pairs are distinguished by the presence/absence of these sounds, and whether both members of the opposition are equally frequent and/or likely to appear in different positions in the word (e.g., syllable onsets vs. codas). When evaluating functional load, it is also important to consider whether to use databases of written or oral corpora, and whether the corpora represent one or multiple varieties. In the case of Spanish, interested readers can conduct a quick search using the *Corpus del español* (https://www.corpusdelespanol.org) and discover that the relative frequency of lexical items varies not only across modalities (written vs. oral) but also across dialects and time.

As concerns functional load in Spanish, examining the frequency counts of individual sounds (e.g., Guirao and García Jurado, 1990; Arias Rodríguez, 2016) and syllables (e.g., Moreno Sandoval et al., 2008) allows for the formulation of several generalizations. First, the vowels /a e o/ are by far the most frequent sounds. Second, the list of the ten most frequent sounds is rounded out by the vowel /i/ and the consonants /t d k s n ɾ/[5]. Finally, in keeping with the importance of prosody argued for in the preceding section, relative frequency is affected by stress. For example, certain vowels are more frequent in unstressed than in stressed syllables (e.g., the relative frequency of /a/ in stressed and unstressed syllables is 4 and 9.3%, respectively; Arias Rodríguez, 2016).

While using functional load as a metric for determining which structures should receive greater focus during pronunciation instruction is appealingly intuitive, it is not without problems. In particular, this concept fails to address suprasegmental features. Given the importance of prosody, this is not an inconsequential limitation. In order to compute functional load for suprasegmentals, several questions can be asked. For instance, how many minimal pairs does a language have at the utterance level (intonation) compared to the lexical level (stress)? As outlined earlier, prosody contributes to intelligibility (e.g., Munro and Derwing, 2006), but how is prosodic intelligibility impacted by functional load? In an attempt to be coherent with our proposal of building an evidence-based pronunciation curriculum, we suggest conservatively that lexical stress and sentence-type intonation should be incorporated into the notion of functional load. From a typological point of view, Spanish is a stress and intonation language (e.g., Jun, 2015) where lexical word contrasts depend on which syllable is realized with longer duration (Ortega Llebaria and Prieto, 2011) and, possibly, higher fundamental frequency (i.e., pitch). Moreover, the function of lexical stress differs in the nominal and verbal paradigms (e.g., Hualde, 2014): whereas stress patterns in nouns can be contrastive (e.g., *sábana* ['saβana] "sheet" vs. *sabana* [sa'βana] "savannah"), within the verbal paradigm, differences in stress patterns serve to realize inflectional features such as mood, tense, and person (e.g., *tome* ['tome] "s/he drinks SUBJ" vs. *tomé* [to'me] "I drank"). The use of tonal variations at the sentence level is also contrastive. In most varieties, a sentence like *Viene* "s/he comes" realized with falling intonation is interpreted as a statement whereas the same sentence with a rising intonation is interpreted as a question. Intonation is critical: there are no additional lexical (e.g., English-type *do*-support) or syntactic differences (word order) that serve to signal differences in sentence type.

In summary, choices concerning what should be taught (most) should not be based primarily on sounds that are difficult to produce, such as the Spanish trill /r/ that is, ironically, among the 10 least frequent segments regardless the corpus consulted, but rather on the realization of vowels, /s/, and sonorants (*Targeting Features and Segments Shared by the Majority of the Varieties of the Target Language* section), which, in addition to being frequent, also encode grammatical features such as gender, number, and person. Furthermore, such sounds should be

[5] The relative frequency patterns described here may vary depending on the source consulted.

taught in different stress conditions and inserted into different sentence types so that students can learn to discriminate the tonal movements used to encode lexical stress from those that are relevant at the sentence level (i.e., to signal questions vs. statements).

Targeting Features And Segments Shared By The Majority of The Varieties of The Target Language

Second language learners typically interact with speakers of different varieties of the target language, as well as with other non-native speakers. In the case of widely spoken languages (including English and Spanish) that are characterized by both great inter-dialectal variation and a body of learners with a wide range of first languages, this leads to there being a great degree of inter-speaker variability in pronunciation. Such variability has consequences for intelligibility. Focusing on the case of English as an international language (that is, English as spoken between non-native speakers), Jenkins (2000; 2002) proposes that instruction should focus on a set of common features central to assuring intelligibility, labeled the Lingua Franca Core (LFC). Attempts to characterize a panhispanic norm have been made for Spanish. For decades, linguists have tried to define the common base shared by educated speakers across the Spanish-speaking world[6] (e.g., Rosenblat, 1967; Alvar, 1991; Lope Blanch, 1993a; Lope Blanch, 1993b; Balmaseda Maestu, 2000; Andión-Herrero, 2008; Gómez Font, 2013; Moreno Cabrera, 2008 or Mar-Molinero and Paffrey, 2011 for a critical view). Although a consensus has not been reached, it is important to highlight that Spanish varieties are highly mutually intelligible, since they share a large percentage of their lexicon and grammar. Still, variation is widespread both at the level of phonological inventory and, particularly, phonetic realization. Several studies emphasize the need to incorporate dialectal variation into the foreign language classroom (Schoonmaker-Gates, 2017) including in Spanish (Casado and Andión, 2014; Bárkányi and Fuertes Gutiérrez, 2019; Zárate-Sández, 2019).

The Spanish phonological system has five vowels that generally maintain their timbre in all syllabic positions, and 15 phonemic consonants shared to a large extent by all Spanish speakers. There are two additional phonemes (/θ/ and /λ/), which are only found in a small set of varieties (see Hualde, 2014 for their cross-dialectal distribution), and two rhotic sounds. A quick examination of standard phonology and dialectology textbooks used in North America (e.g., Lipski, 1994; Hualde, 2014), reveals that generalizing across varieties is challenging. Truly, there is hardly a segmental or suprasegmental feature of Spanish phonology that has not been described as variable[7]. The degree of variability, however, differs by feature. There is widespread consensus that vowels are less variable than consonants, a situation which contrasts with English. Moreno Fernández (2000) only mentions two instances of vocalic variability: the weakening and loss of unstressed vowels in voiceless contexts in the Mexican highlands and Andean regions (e.g., *antes* [ˈants] instead of [ˈantes] “before”; *cafesito* [kafˈsito] instead of [kafeˈsito] “coffee”), and vowel lengthening in Dominican Spanish. There are, however, other instances of variability, such as the laxing of low and mid vowels as a consequence of the lenition of word-final /s/ in Andalusian Spanish (e.g., Henriksen, 2017: *perros* [ˈperos] > [ˈperɔ] > [ˈpɛrɔ] “dogs”), which could be discussed in more advanced courses. In spite of these few instances of variability, in contrast to English, Spanish is characterized by its lack of unstressed vowel reduction: stressed and unstressed vowels have the same quality but may differ in duration. This is an important feature to highlight when teaching pronunciation and should be emphasized right from the beginning of the learning process, as per the many studies that have demonstrated improvement when this feature is taught (Lord, 2005; Lord and Fionda, 2013; Long et al., 2018; Martínez Celdrán and Elvira-García, 2019).

Although individual vowels are relatively stable, vocalic sequences are highly variable across Spanish dialects with a clear preference for the diphthongization of mid vowels in Latin America (Garrido, 2008; Colantoni and Hualde, 2016) when compared to Spain, triggering perceptual confusion between words like *palear* [paleˈar] “to shovel” and *paliar* [paˈljar] “to ease” since both are pronounced [paˈljar]. Given that this process applies to sequences within and across words, it deserves attention, as, in the latter case, it introduces variability into the pronunciation of word-final vowels. In general, the realization of vowels across words, which may range from diphthongization to fusion, needs to be discussed, since Spanish, in contrast to English (e.g., Davidson and Erker, 2014), tends to resyllabify vowels across words (Hutchinson, 1974; Alba, 2006; Hualde et al., 2008). This resyllabification may lead to perceptual confusion; this is particularly problematic in word-final position since, as highlighted earlier, these final vowels encode grammatical information including agreement, person, and tense.

Turning to the consonantal system, several segments are relatively less variable: the realization of /ptfmn/ is characterized by minor cross-dialectal differences. In contrast, /ʎ/ is disappearing, still used in bilingual Catalan-Spanish communities and by older generations in particular areas of Spain and America (Gómez and Molina Martos, 2013). As concerns the latter, accordingly, it is important to make learners aware of the extremes in the continuum (from the palatal glide [ˈkaje] *calle* “street” to the post-alveolar voiceless fricative [ˈkaʃe]), variation that the *Plan Curricular del Instituto Cervantes* recommends presenting at the intermediate (CEFR B) level, since this variability may pose comprehension problems for both L1 and L2 speakers (MacLeod, 2012)). The other palatal in the system, /ɲ/, also shows signs of depalatalization in some Spanish dialects, where it is being replaced by a sequence of a glide + alveolar nasal. This realization, however, poses fewer problems for intelligibility and comprehension than the palatal fricative variants (Kochetov and Colantoni, 2011; Bongiovanni, 2015).

[6] This would be the normative Spanish proposed by such organizations as the Real Academia Española and the Academias de la Lengua of all Spanish-speaking countries.

[7] We refer the reader to Hualde (2014) and Real Academia Española and Asociación de Academias de la Lengua Española (2011) for in-depth discussion of phonetic variation across the Spanish-speaking world.

The voiced stops /b d g/ have similar characteristics in all varieties, with differences only in the distribution of their allophones. Generally, stop realizations are found in absolute word-initial position or following a nasal, except in the interior of Mexico and in the highlands of Colombia where they occur even between vowels, especially across words (Canfield, 1962; Montes Giraldo, 1975; Lipski, 1994; Michnowicz, 2009). However, stop realizations of /b d g/ should prove less problematic for learners than extreme weakening or deletion, since stop maintenance mirrors the orthographic form. Instead, weakening and deletion, a frequent process in many Spanish-speaking areas (Moreno Fernández, 2000), may impact intelligibility. Laterals and rhotics are characterized by a large degree of variability across Spanish varieties. However, intervocalic laterals and taps are realized in a similar way, and thus, should be targeted before the same segments in codas or complex onsets. Indeed, laterals and rhotics alternate in codas in many Spanish varieties. In intervocalic position, the tap and the trill alternate (['koɾo] "*chorus",* ['koro] "*I run*"). In all other positions in the word, the two segments are in complementary distribution. Although there is a large degree of variability in the actual realization of the trill (e.g., Blecua, 2008), all varieties maintain an opposition between tap and trill rhotics in intervocalic position. Another contrast involving taps that is maintained across varieties and that is usually ignored is the /d ɾ/ opposition in intervocalic position. Attention to this contrast is particularly relevant for learners whose L1 is an English variety in which coronal stops are flapped in this context.

Fricatives are extremely variable across Spanish dialects to the point that Peninsular and Latin American varieties differ in the number of fricative phonemes. Whereas in the former varieties there is an opposition between /s/ and /θ/ (e.g., ['kasa] "house", ['kaθa] "hunting"), in the latter, the opposition has been reduced to /s/. Since most of the Spanish-speaking world has merged both phonemes (independently of variability in /s/ realization), it may be advisable to begin by focusing on /s/ realizations and to turn to the realization of the /s/-/θ/ opposition at upper levels. Although /s/ in onsets is relatively stable, the weakening of coda /s/ is one of the most-well studied phenomena in Spanish dialectology and sociolinguistics (e.g., Cedergren, 1978; Terrell, 1978; Hammond, 1980; Lipski, 1984; Lipski, 1985; Torreira and Ernestus, 2012). For our purposes, it is important to point out that teaching /s/ maintenance in codas makes a contribution to learners' acquisition of the Spanish nominal and verbal systems. In addition to /s/, Spanish has a dorsal fricative /x/, which may show realizations ranging from a tense uvular fricative in Spain to a lax aspirated variant in the Caribbean. Weakly aspirated realizations may be perceived as vocalic sequences, and thus, pose a problem for learners (e.g., *cejas* [sexas] "eyebrows" can be understood as *seas* [seas] "you are, SUBJ"). Thus, such dialectal variation may likely need to be discussed in upper-intermediate and advanced courses.

As concerns the suprasegmental level, the main and most important similarity is in the placement and realization of lexical stress. There are indeed very few words that have different stress patterns across varieties (see Hualde, 2014, Chapter 10 for examples). Since stress is important for lexical retrieval and for the learning of verbal morphology, it should be taught from the very beginning. At the syllable level, it may be important to discuss certain sandhi (i.e., reduction) phenomena, which may facilitate intelligibility, such as resyllabification mentioned above. Although lack of resyllabification may fail to hinder intelligibility, it may delay comprehension. Thus, it is well motivated to dedicate first efforts to familiarizing beginners with those great points of coincidence common to all varieties of the target language. As concerns sentence intonation, cross-dialectal comparisons (e.g., Sosa, 1999) suggest that all dialects have the same prosodic realization of declaratives and interrogatives, namely, the first peak is always relatively higher in interrogatives than in declaratives. Varieties do differ in the realization of nuclear contours, particularly in questions. As concerns phrasing, and if we have English learners in mind, it is also worth stressing that, in Spanish, subjects tend to be phrased independently of the verb phrase. Moreover, within noun phrases, as with sentences in general, the nuclear accent tends to be on the final constituent (Estebas-Vilaplana and Prieto, 2010; Gabriel et al., 2010). This means, for example, that in noun + adjective phrases, the nuclear stress falls on the adjective, whereas in adjective + noun phrases, the nuclear stress is placed on the noun. Contrastive pitch accents on the first element of the noun phrase are rarely heard in Spanish.

ASSESSMENT OF CURRENT PRACTICES IN SPANISH PRONUNCIATION TEXTBOOKS

When working toward an evidence-based curriculum which seeks to train teachers and learners alike, we need to turn to the existing textbooks, as well as to recent literature on Spanish phonetics, phonology, and pronunciation teaching, in order to determine which practices are established and which of these are consonant with the principles espoused here. We discuss textbooks for the North American and European markets separately, which target L1 English learners vs. learners with a wider variety of L1 backgrounds, respectively.

Profile of Widely Used Textbooks in North America

There are four textbooks that are widely used in North America: 1) *Spanish pronunciation* (Dalbor, 1980); 2) *Fonética y fonología Española* (Schwegler and Ameal-Guerra, 2019); 3) *Sonido y Sentido* (Guitart, 2004); and 4) *Sonidos en contexto* (Morgan, 2010). All of these books are clearly written with an American, English-speaking audience in mind and, for the most part, follow a traditional organization presenting first consonants and vowels (the order differs by textbook) and then prosody[8]. Morgan (2010) and Schwegler and Ameal-Guerra (2019) are the only textbooks that are accompanied by on-line resources. All four textbooks

[8] Both Morgan (2010) and Schwegler and Ameal-Guerra (2019) depart slightly from this structure, and discuss some aspects of prosody before introducing vowels and consonants.

include a variety of exercises, most aimed at developing students' production rather than perception. To this end and in order to familiarize students with different Spanish varieties, three of the textbooks (Guitart, Schwegler and Ameal-Guerra, and Morgan) have recordings which are made available digitally via a CD (Guitart) or through a website (Schwegler and Ameal-Guerra, Morgan). All of these books address the problem of which variety to teach, including lengthy discussions concerning dialectal or sociolectal variation (Schwegler and Ameal-Guerra and Morgan, in particular), although recordings do not always feature speakers of different varieties, and incorporate additional information regarding the history of Spanish and/or of the Spanish spoken in the United States.

In addition to these textbooks, in a recent volume devoted to reflections on the teaching of Spanish pronunciation, Rao (2019) speaks to the need of developing a pronunciation curriculum for Spanish instructors. Moreover, Rao addresses the importance of having a conversation concerning which sounds should be prioritized in teaching and which variety should be taught.

Profile of Widely Used Textbooks in Europe

There is a scarce supply of teaching materials for Spanish pronunciation in the European market and, with few exceptions, they are not particularly innovative (unlike teacher training manuals, that include excellent books, such as Gil Fernández, 2007; Gil Fernández, 2012 or Cortés Moreno, 2002 for prosody).

Textbooks from well-known publishers, such as Edelsa (González Hermoso and Romero Dueñas, 2002a; González Hermoso and Romero Dueñas, 2002b) or Anaya (Nuño Álvarez and Franco Rodríguez, 2001), begin with the presentation of vowels, followed by consonants, and end with syllables, stress and intonation (mainly of declarative and interrogative sentences). Exercises are very limited in nature, being of the type listen and repeat/write/complete/search for the intrusive sound or minimal pair discrimination.

A notable exception is Padilla (2015) *La pronunciación del español. Fonética y enseñanza de lenguas* (University of Alacant), whose declared purpose is "to improve the dialogue between theoretical phonetics and the teaching of pronunciation". This textbook focuses both on speech perception and production. In addition to segments, it incorporates stress, rhythm, and intonation, and also discusses the conversational and kinetic components, linking the teaching of rhythm and intonation with everyday conversation dialogues, and paying attention to the visual and gestural component (gestures of the face, movements of the hands, etc.). This textbook also includes an interesting comparison between the phono-articulatory and the verbo-tonal methods, and ends with a didactic proposal with exercises "in a protocol of phono-cognitive performance". This protocol is built upon two cornerstones: the particular phonetic mechanisms and the more general cognitive processes of acquisition. This text is sequenced in six phases: presentation of the model, mechanical perception, mechanical production, reflection and contrast, conscious perception and, finally, conscious production.

General Spanish as a Foreign Language textbooks, such as *ELE actual* (SM publisher), '*Español*' 2000, *Diverso* (SGEL) include pronunciation sections very closely linked to spelling (i.e., with a clear focus on the segmental level) with few units devoted to lexical stress or the intonation patterns of basic sentence types. The types of exercises included are similar to those found in general pronunciation textbooks, namely, 1) listen (to recordings or the instructor's pronunciation) and identify (sometimes using minimal pairs); 2) listen and repeat or write; 3) read aloud (classic literary texts, in some textbooks). Arguably, *Difusión* is the commercial publisher making the largest efforts to update its pronunciation teaching offerings; in its Spanish teaching methods (*Gente joven, nueva edición; Aula Internacional, Socios*), the suprasegmental level receives extensive attention, all units include content targeting the segmental level, and, in some cases, dialectal variation is addressed.

In summary, there are commonalities and exceptions when we compare the textbooks available in both markets. Textbooks on both sides of the Atlantic share, to a large extent, the organization of the contents and the way in which they are presented, albeit they differ in the L1s addressed.

PUTTING PRINCIPLES INTO PRACTICE: SAMPLE PERCEPTION AND PRODUCTION ACTIVITIES

We now turn to demonstrating the full implementation of these principles in perception and production activities targeting beginner, university-level learners of Spanish. The choice of such a population is motivated by the fact that it allows us to illustrate most of our principles, particularly the focus on frequent and relatively low-variability structures, efficiently. It also allows us to explain how the complexity of more basic exercises can be increased to address the needs of more proficient learners. The goal of our exercises is to practice the perception and production of word-final unstressed vowels, /a e o/ in particular. The reasons for focusing on these vowels are numerous. First, they are not acquired easily: adult learners of Spanish and heritage speakers alike diverge from baseline speakers in their perception and production of these vowels (Mazzaro et al., 2016; Colantoni et al., 2020). Second, in *A Focus on Features With High Functional Load* and *Targeting Features and Segments Shared by the Majority of the Varieties of the Target Language* sections, we highlighted that these vowels, particularly in unstressed position, are among the most frequent segments in Spanish and are realized in a similar fashion across dialects. Moreover, these vowels encode crucial morphosyntactic information, such as gender and person/tense/mood. As such, their accurate perception will facilitate the acquisition of key components of Spanish grammar, and their accurate production will have an impact on intelligibility. Finally, as highlighted in *Targeting Features and Segments Shared by the Majority of the Varieties of the Target Language* section, these vowels are realized differently when pronounced in absolute word-final position

TABLE 1 | Suggested words for perception exercises targeting the Spanish /a e o/ contrast.

/a o/		/a e/		/e o/	
Noun	**Verb**	**Noun**	**Verb**	**Noun**	**Verb**
pala-palo “shovel-stick”	coma-como “s/he eats, SUBJ - I eat”	nena-nene “girl-boy”	beba-bebe “s/he drinks, SUBJ - s/he drinks”	base -vaso “base - glass”	parte-parto “s/he leaves - I leave”
hoja-ojo “leaf-eye”	duerma-duermo “s/he sleeps, SUBJ - I sleep”	jefa-jefe “female boss-male boss”	cena-cene “s/he dines - s/he dines, SUBJ”	hombre-hombro “man-shoulder”	cante-canto “s/he sings, SUBJ - I sing”
niña-niño “girl-boy”	diga-digo “s/he says, SUBJ - I say”	bota-bote “boot-boat”	cuenta-cuente “s/he tells - s/he tells, SUBJ”	leche-lecho “milk-bed”	hable-hablo “s/he speaks, SUBJ - I speak”
gata-gato “female cate-male cat”	pueda-puedo “s/he can, SUBJ - I can”	tela-tele “fabric-TV set”	toma-tome “s/he drinks - s/he drink, SUBJ”	calle-callo “street-callus”	Sueñe-sueño “s/he dreams, SUBJ - I dream”
pata-pato “female duck-male duck”	mira-miro “s/he sees, - I see”	clienta-cliente “female client-male client”	mueva-mueve “s/he moves, SUBJ - s/he moves”	pase-paso “pass-step”	teme-temo “s/he fears - I fear”

vs. when followed by another vowel-initial word. Thus, practicing them in isolation vs. in context is relevant, since intelligibility may be compromised if an isolated focus alone is adopted.

Perception of Spanish /a e o/

The goal of this exercise is to increase learners’ accuracy in the discrimination and identification of the vowel pairs /a o/, /a e/, and /e o/. We propose to do this by progressing from the discrimination and identification of isolated words to the identification of words in context. As explained in the *Targeting Features and Segments Shared by the Majority of the Varieties of the Target Language* section, final vowels in isolation are less variable than when occurring in sequences. Inspired by Gluhareva and Prieto (2017) and Lee (2020), so as to make these final vowels more prominent, 1) we will propose a warm-up exercise in which we use rhythm to enhance the stress patterns, and 2) we will present the stimuli with falling and rising contours, since the latter context makes them more perceptible. In this way, students will also practice the prosodic cues to sentence types. If the learners’ L1 is a tonal language, we recommend that teachers make them aware that tonal variations in Spanish convey sentence meaning rather than lexical meanings, since they may tend to associate the different prosodic contours with the latter (Ortega Llebaria et al., 2015).

We will use the materials presented in **Table 1**. For students to be familiarized with or reminded of the stress patterns and the correlates of stress (e.g., duration rather than vowel quality), instructors will use clapping to emphasize the trochaic pattern of all words, as in Gluhareva and Prieto (2017). Instructors could also read the words, exaggerating the longer duration of the stressed syllable. After this warm up, instructors can present the words, which could have been previously recorded by the instructor or by other native Spanish speakers. Target words can be presented in pairs and students will be asked if the words are the same or different. Here, the instructor may want to present this as an individual or rather as a group activity with a competitive component (e.g., the group with more accurate responses wins) to increase learners’ motivation. In order to make the exercise more difficult, instructors may either choose to use triplets (i.e., an ABX discrimination task) instead of pairs or have stimuli recorded by different speakers, since it has been shown in training studies that increasing speaker variability has a positive impact on accurate perception, in spite of making the exercise more difficult at the beginning of testing (e.g., Logan et al., 1991; Logan and Pruitt, 1995). The instructor can also vary the temporal distance between the presentation of the words (i.e., the interstimulus interval, ISI). Perception studies have shown that longer ISIs target phonological rather than auditory perception because shorter intervals between target stimuli enable acoustic listening rather than listening with learned phonemic categories (Flege and MacKay, 2004; Escudero et al., 2009).

Once learners can discriminate the final vowels, we will work on their identification. For that purpose, a variety of exercises can be used. The easiest one is to ask learners to transcribe what they hear; learners can also be presented with two or three orthographic transcriptions on a computer screen and be asked to choose the correct one. Alternatively, with depictable nouns, images could be presented on a screen and learners asked to choose the correct image. With beginner learners with a sufficient grasp of present tense forms, which are typically taught early on, accuracy with final vowels in verbal forms could be tested by asking learners to write down the appropriate subject for high frequency verbs (for example, when they hear *parto* “I leave”, they would be expected to write *yo* “I” as opposed to *él/ella* “s/he”), keeping in mind that, if we do this, it may be difficult to distinguish perception skills from the knowledge of the grammar.

To practice vowels in sequences, discrimination and identification activities can be designed by recording the words in **Table 1** followed by adjectives in the case of nouns and direct objects or other modifiers in the case of verbs. For example, to design a discrimination exercise, students could listen to pairs of stimuli such as *hoja* azul “blue leaf” vs. *ojo azul* “blue eye” or *como alfajores* “I eat sandwich cookies” vs. *come alfajores* “he eats sandwich cookies”, and be asked to indicate whether these phrases are the same or different. In an identification experiment, they could see two pictures and be asked to choose the appropriate one.

Production of Spanish /a e o/

We propose two exercises to practice the production of Spanish /a e o/ here. The goal of the first exercise will be to practice the production of these vowels in isolated words: by doing so, we will target vowel quality in insolation and make sure that learners are producing the correct vowel rather than, for example, a schwa. Keeping it in mind that this exercise complements those proposed in the *Perception of Spanish /a e o/* section, we will once again target beginner students. We will add suggestions for instructors so that they can manipulate the complexity in order to adapt these exercises for more advanced students.

In the first production exercise, students will work in pairs. Using digital flashcards, Student one will receive the words listed in **Table 1**. Student one will pick a word to read aloud. Student two will have to write/type the word. Once the students have moved through the set, students will compare notes and discuss the types of errors witnessed. For example, if the transcriber is not sure about vowel quality in many of the words, this implies that Student one is not making a (sufficient) difference between the target vowels. Students can further investigate in which words misperceptions occurred and see if they can identify any phonological context that explains where difficulties were found. Additionally, if students are familiar with acoustic analysis techniques, they could measure vowel formants in Student one's productions.

The second exercise involves the production of the same vowels, this time in short sentences so that students can practice these vowels in context. The instructor should remind students that these vowels are produced contiguously without a pause or the insertion of a glottal stop, unlike in English, for example. In order to practice nouns ending in vowels, there will be pictures depicting each of the options. Once again, students may work in pairs with one student reading sentences such as those in (1–3), and another student choosing the appropriate image. To practice verbs ending with vowels, one student may read a sentence, such as those in (4–5), and the other one may write down the appropriate pronoun (sentences may also be depicted).

(1) Tiene tela/tele "S/he has fabric/a TV set"
(2) Cava con pala/palo "S/he digs with a shovel/stick"
(3) De niña/de niño, andaba en bicicleta "When I was a female/male child, I used to ride a bike"
(4) Cena pronto/Cene pronto "S/he dines early/s/he dines (SUBJ) early"
(5) Hablo tranquilo/Hable tranquilo "I speak quietly/S/he speaks (SUBJ) quietly"

EXPANDING EVIDENCE-BASED PRINCIPLES TO CURRICULUM DESIGN

In this article, we have proposed five evidence-based pronunciation instruction principles targeting both what should be taught – segmental as well as prosodic features, particularly those that have a high functional load and are shared across varieties – and how, namely, via contextualized perception and production activities targeting not only individual words but also larger prosodic units. In illustrating the application of these principles, we proposed structured perception and production activities for beginner L2 learners of Spanish. What we have outlined here is only the first step in the larger process of creating an evidence-based pronunciation curriculum, whether it be for the teaching of pronunciation within broader "four skills" classes or rather for courses focused on pronunciation alone. The overall learning objective of such a curriculum would not change – to help learners move toward ever increasing intelligibility. What remains to be done is to determine how our evidence-based principles can be applied to this larger project. We outline here a set of three important questions that instructors must ask themselves when designing such a curriculum, questions that are shaping our own work on the development of an evidence-based Spanish pronunciation textbook.

The first factor to consider is target language proficiency. To this point, we have touched on this issue tangentially. However, following the general pedagogical principle of developmental readiness, it is usual practice to implement a progressive curriculum in which structures to be learned are spread across proficiency levels with scaffolding allowing learners to improve continuously assisted by consciousness-raising instruction. Our first question is thus: what segmental and prosodic structures should be taught at what levels? Elaborating an in-depth, evidence-based answer to this is no small feat. Some of the principles evidenced here provide a partial answer. For example, when discussing the features that are relatively stable across Spanish dialects, we have made a case regarding which segmental and suprasegmental aspects should be taught first. We have also underlined the importance of certain phonological features for learning morphosyntactic aspects of the language; including such phonological features in the Spanish pronunciation curriculum will thus allow learners to bootstrap from phonology to morphosyntax and make their overall learning more successful. Empirical research in (applied) linguistics also provides insights. In keeping with the evidence-based nature of the pronunciation instruction advocated for here, we underline the importance of aligning instructional practice with learning sequences. It is now well established that developmental sequences exist for many areas of linguistic ability (e.g., Meisel et al., 1981; Gleason and Ratner, 1989; Clark, 2003 for general discussion of such stages; Colantoni et al., 2015 for examples from L2 speech research)[9]. Moreover, it is possible to test pronunciation effectiveness empirically in both classroom- and laboratory-based instructional and training studies (see Lee et al., 2015 for a meta-analysis; Lord and Fionda, 2013: 517–522 for a summary of studies of the effects of pronunciation instruction on Spanish learners of different proficiency levels). Consequently, proficiency-level-appropriate pronunciation instruction practices can be informed both by

[9] One might also wish to turn to progressive learning and assessment frameworks such as the European Common Framework of Reference for Languages for insights into pedagogical sequencing. Some caution is, however, warranted in basing instructional practices on such frameworks: various researchers have questioned their evidence-based nature including the extent to which they align with learning sequences (Hulstijn et al., 2010 for general discussion) or have demonstrated empirical divergences between such frameworks and real-world language use (Kusseling and Lonsdale, 2013 for vocabulary profiles).

evidence-based L2 developmental sequences and studies designed to measure the effect of instruction on learners of different proficiency levels.

When considering the issue of relative importance or sequencing, it is not only the phonetic and phonological structures as modulated by learner proficiency that must be weighted. A second central question to the development of an effective pronunciation curriculum is what the relative weighting given to each of the individual principles should be. For example, as we illustrated in our sample exercises, practicing sounds that are frequent and have a high functional load, such as /a e o/, comes at the expense of teaching these vowels in context, namely, in the smaller contexts of words, in the perception exercises, so as to allow learners to discriminate and identify the elements that we are working on, and later, in the production exercises, in larger contexts such as phrases and sentences. Thus, we need to take into account two competing principles, namely, functional load and context, in order to facilitate learning with the latter principle becoming more important following the initial stage of perceptual learning.

Finally, there is the question of how the principles we suggest are best implemented, particularly in the context of real world classrooms. While research on best practices in instruction exists (e.g., Wrembel, 2007; Derwing and Munro, 2015; Levis, 2018), this is a question for which we currently need more evidence. Luckily, the growing number of publications targeting the effects of different factors including instructional type (e.g., Saito and Plonsky, 2019) as well as conferences and workshops on L2 pronunciation instruction (e.g., Pronunciation in Second Language Learning and Teaching, PSSLT) demonstrate that answers to this final question are already being offered.

AUTHOR CONTRIBUTIONS

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

FUNDING

PE's work and article publication fees were funded by an Australian Research Council Future Fellowship (FT160100514). VMA acknowledges the Ministry of Science, Innovation and Universities of Spain for the grant from the "Programa de Estancias de Movilidad de profesores e investigadores en centros extranjeros de enseñanza superior e investigación 2019" that favored the contacts that led to this work.

ACKNOWLEDGMENTS

We thank Deeahn Sako for editorial help with reference and table formatting as well as reading the revised version for consistency.

REFERENCES

Alba, C. M. (2006). "Accounting for variability in the production of Spanish vowel sequences," in Selected Proceedings of the 9th Hispanic Linguistics Symposium. Editors N. Sagarra and A. J. Toribio, Pennsylvania State University, November 10–13, 2005 (Somerville, MA: Cascadilla Press), 273–285.

Alvar, M. (1991). *El español de las dos orillas [The Spanish of the two shores]*. Majadahonda, Spain: Fundación MAPFRE.

Anderson-Hsieh, J., Johnson, R., and Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentais, prosody, and syllable structure. *Lang. Learn.* 42 (4), 529–555. doi:10.1111/j.1467-1770.1992.tb01043.x

Anderson-Hsieh, J., and Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Lang. Learn.* 38 (4), 561–613. doi:10.1111/j.1467-1770.1988.tb00167.x

Andión Herrero, M. A. (2008). Modelo, estándar y norma: conceptos imprescindibles en el español L2/LE [Model, standard and norm: essential concepts in Spanish L2/FL]. *Rev. Esp. Lingüíst. Apl.* 21, 9–26. doi:10.1075/resla

Arias Rodríguez, I. (2016). Cálculo de frecuencias de aparición de fonemas y alófonos en español actual utilizando un transcriptor automático [Computation of frequency of occurrence of phonemes and allophones in modern Spanish using an automatic transcription system]. *Loquens* 3, 1–29. doi:10.15517/am.v3i0.25204

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Atten. Percept. Psychophys.* 81, 981–1005. doi:10.3758/s13414-019-01725-4

Baese-Berk, M. M., and Samuel, A. G. (2016). Listeners beware: speech production may be bad for learning speech sounds. *J. Mem. Lang.* 89, 23–36. doi:10.1016/j.jml.2015.10.008

Balmaseda Maestu, E. (2000). Norma panhispánica y enseñanza del español [Pan-Hispanic norm and Spanish teaching]. Actas del coloquio internacional de la AEPE Available at: https://cvc.cervantes.es/ensenanza/biblioteca_ele/aepe/pdf/coloquio_2000/coloquio_2000_22.pdf (Accessed November 07, 2020).

Bárkányi, Z., and Fuertes Gutiérrez, M. (2019). Dialectal variation and Spanish language teaching (SLT): perspectives from the United Kingdom. *J. Spanish Lang. Teach.* 6 (2), 199–216. doi:10.1080/23247797.2019.1676980

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht, Netherlands: Foris Publication.

Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech perception and linguistic experience: issues in cross-language research*. Editor W. Strange (Walmgate, England: York Press), 171–204.

Best, C. T., and Tyler, M. (2007). "Nonnative and second-language speech perception," in *Language experience in second language speech learning: in honour of James Emil Flege*. Editors O. S. Bohn and M. J. Munro (Amsterdam, Netherlands: John Benjamins), 13–34.

Blecua, B. (2008). Los sonidos vibrantes: aspectos comunes y variación [Trills and taps: common features and variability] In *New trends in experimental phonetics: selected papers from the IV International Conference on Experimental Phonetics*. Editors A. Pamies Bertrán and E. Melguizo Moreno Granada, February 11–14, 2008. 1, 23–30.

Bongiovanni, S. (2015). Neutralización del contraste entre /η/ y /nj/ en el español de Buenos Aires: Un estudio de percepción [Neutralization of the /η/-/nj/ contrast in Buenos Aires Spanish: a perception study]. *Rev. Instit. lingüíst.* 27, 11–46. doi:10.34096/sys.n27.3187

Bowen, J. D. (1972). Contextualizing pronunciation practice in the ESOL classroom. *TESOL Q.* 6 (1), 83–94. doi:10.2307/3585862

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Q.* 22 (4), 593–606. doi:10.2307/3587258

Canfield, D. L. (1962). *La pronunciación del español en América: Ensayo histórico-descriptivo [Latin American Spanish pronunciation: a historical-descriptive essay]*. Bogota, Columbia; Instituto Caro y Cuervo.

Casado, C., and Andión, M. A. (2014). *Variación y variedad del español aplicadas a E-LE/L2* [Variation and variety of Spanish applied to E-LE/L2]. Madrid, Spain: UNED.

Cedergren, H. (1978). "En torno a la variación de la s final de sílaba en Panamá: análisis cuantitativo [On the syllable coda s in Panama: a quantitative analysis]," in *Corrientes actuales en la dialectología del Caribe hispánico*. Editor H. López Morales (Santiago, Chile: Editorial Universitaria), 37–49.

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., and Griner, B. D. (2010). *Teaching pronunciation: a course book and reference guide*. 2nd Edn. New York, NY: Cambridge University Press.

Clark, E. (2003). *First language acquisition*. New York, NY: Cambridge University Press.

Colantoni, L., and Hualde, J. I. (2016). "Constraints on front-mid vowel gliding in Spanish," in *The syllable and stress*. Editor R. Nuñez Cedeño (Berlin, Germany: Mouton de Gruyter), 1–28.

Colantoni, L., Martínez, R., Mazzaro, N., Pérez-Leroux, A. T., and Rinaldi, N. (2020). A phonetic account of Spanish-English bilinguals' divergence with agreement. *Lang.* 5 (4), 58. doi:10.3390/languages5040058

Colantoni, L., Steele, J., and Escudero, P. (2015). *Second language speech*. New York, NY: Cambridge University Press.

Cortés Moreno, M. (2002). *Didáctica de la prosodia del español: la acentuación y la entonación* [Didactics of Spanish prosody: stress and intonation]. Madrid, Spain: Editorial Edinumen.

Dalbor, J. (1980). *Spanish pronunciation. Theory and practice*. 3rd Edn. New York, NY: Holt, Rinehart and Winston.

Davidson, L., and Erker, D. (2014). Hiatus resolution in American English: the case against glide insertion. *Lang.* 90, 482–514. doi:10.1353/lan.2014.0028

de la Mota, C. (2019). "Improving non-native pronunciation: teaching prosody to learners of Spanish as a second/foreing language," in *Key issues in the teaching of Spanish pronunciation*. Editor R. Rao (London, United Kingdom: Routledge), 163–197.

Delattre, P. (1969). An acoustic and articulatory study of vowel reduction in four languages. *IRAL.* 7 (4), 295–326.

Derwing, T. M., and Munro, M. J. (2014). "Once you have been speaking a second language for years, it is too late to change your pronunciation," in *Pronunciation myths: applying second language research to classroom teaching*, Editors L. Grant, et al. Ann Arbor, Michigan: The University of Michigan Press, 34–55.

Derwing, T. M., and Munro, M. J. (2015). *Pronunciation fundamentals: evidence-based perspectives for L2 teaching and research*. Amsterdam, Netherlands: John Benjamins.

Derwing, T. M., Munro, M. J., and Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Lang. Learn.* 48 (3), 393–410. doi:10.1111/0023-8333.00047

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., and Peperkamp, S. (2008). Persistent stress 'deafness': the case of French learners of Spanish. *Cognition* 106 (2), 682–706. doi:10.1016/j.cognition.2007.04.001

Elvin, J., and Escudero, P. (2019). "Cross-linguistic influence in second language speech: implications for learning and teaching," *Cross-linguistic influence: from empirical evidence to classroom practice,* Editors M. J. Gutierrez-Mangado, M. Martínez-Adrián, and F. Gallardo-del-Puerto (Cham: . Springer), 1–20.

Escudero, P. (2005). *Linguistic perception and second language acquisition: explaining the attainment of optimal phonological categorization*. Amsterdam, Netherlands: Netherlands Graduate School of Linguistics.

Escudero, P. (2009). "Linguistic perception of "similar" L2 sounds," in *Phonology in perception*. Editors P. Boersma and S. Hamann (Berlin, Germany: Mouton de Gruyter), 151–190.

Escudero, P. (2007). "Second-language phonology: the role of perception," in *Phonology in context*. Editor M. Pennington (London, United Kingdom: Palgrave Macmillan), 109–134.

Escudero, P., Benders, T., and Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: the case of Dutch, German, and Spanish listeners. *J. Phon.* 37 (4), 452–465. doi:10.1016/j.wocn.2009.07.006

Escudero, P., and Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition* 133 (2), 408–413. doi:10.1016/j.cognition.2014.07.002

Estebas-Vilaplana, E., and Prieto, P. (2010). "Castilian spanish intonation," in *Transcription of intonation of the Spanish language*. Editors P. Prieto and P. Roseano (Munich, Germany: Lincom Europa), 17–48.

Field, J. (2005). Intelligibility and the listener: the role of lexical stress. *TESOL Q.* 39 (3), 399–423. doi:10.2307/3588487

Flege, J., and Bohn, O. (2021). "The revised speech learning model (SLM-r)," in *second language speech learning: theoretical and empirical progress*. Editor R. Wayland (New York, NY: Cambridge University Press), 3–83.

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 (4), 437–470. doi:10.1006/jpho.1997.0052

Flege, J. E., and MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Stud. Sec. Lang. Acq.* 26 (1), 1–34. doi:10.1017/s0272263104261010

Flege, J. E. (1995). "Second language speech learning," in *Speech perception and linguistic experience: issues in cross-language research*. Editor W. Strange (Walmgate, England: York Press), 233–277.

Fry, D. B. (1965). "The dependence of stress judgments on vowel formant structure," in Proceedings of the 5th International Congress of Phonetics Sciences, August 1965. Editors E. Zwimer and W. Bethge, Münster, August 16–22, 1964, (Basel, Switzerland: Karger), 306–311.

Gabriel, C., Feldhausen, I., Pešková, A., Colantoni, L., Lee, S. A., Arana, V., et al. (2010). "Argentinian Spanish intonation," in *Transcription of intonation of the Spanish language*. Editors P. Prieto and P. Roseano (Munich, Germany: Lincom Europa), 285–317.

Garrido, M. (2008). Diphthongization of non-high vowel sequences in Latin American Spanish. PhD dissertation. Champaign (IL): University of Illinois at Urbana-Champaign.

Gatbonton, E., and Segalowitz, N. (1988). Creative automatization: principles for promoting fluency within a communicative framework. *TESOL Q.* 22 (3), 473–492. doi:10.2307/3587290

Gil Fernández, J. (2007). Fonética para profesores de español: de la teoría a la práctica [Phonetics for Spanish teachers: from theory to practice]. Madrid, Spain: Arco Libros, 298–309.

J. Gil Fernández (Editor) (2012). *Aproximación a la enseñanza de la pronunciación en el aula de español [Approach to teaching pronunciation in the Spanish classroom]*. Madrid, Spain: Edinumen.

Gilbert, J. B. (2008). *Teaching pronunciation. Using the prosody pyramid*. New York, NY: Cambridge University Press.

Gilbert, J. (2005). *Clear speech*. 3rd Edn. New York, NY: Cambridge University Press.

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Lang. Teach.* 40, 97–118. doi:10.1017/s0261444807004144

Gleason, J. B., and Ratner, N. B. (1989). *The development of language*. New York, NY: Merrill.

Gluhareva, D., and Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Lang. Teach. Res.* 21 (5), 609–631. doi:10.1177/1362168816651463

Gómez Font, A. (2013). Español neutro, global, general, estándar o internacional [Neutral, global, general, standard or international Spanish]. *Aljamía. Revista de la Consejería de Educación en Marruecos* 24, 9–15.

R. Gómez and I. Molina Martos (Editors) (2013). *Variación yeísta en el mundo hispánico [Yeísta variation in the Hispanic world]*. Madrid, Spain: Iberoamericana Editorial Vervuert S.L.

González Hermoso, A., and Romero Dueñas, C. (2002a). *Tiempo para pronunciar [Time for pronunciation]*. Madrid, Spain: Edelsa.

González Hermoso, A., and Romero Dueñas, C. (2002b). *Fonética, entonación y ortografía. + de 350 ejercicios para el aula y el laboratorio [Phonetics, intonation and spelling]*. Madrid, Spain: Edelsa.

Goodin-Mayeda, E. (2019). "The role of perception in learning Spanish pronunciation," in *Key issues in the teaching of Spanish pronunciation*. Editor R. Rao (London, United Kingdom: Routledge), 254–26.

Guirao, M., and García Jurado, M. A. (1990). Frequency of occurrence of phonemes in American Spanish. *Revue québécoise de linguistique.* 19, 135–149.

Guitart, J. (2004). *Sonido y sentido [Sound and meaning]*. Washington, DC: Georgetown University Press.

Hammond, R. (1980). "Las realizaciones fonéticas del fonema /s/ en el español cubano rápido de Miami [The realizations of the /s/ phoneme in fast-spoken Cuban Spanish in Miami]," in *Dialectología hispanoamericana: estudios actuales*. Editor G. E. Scavnicky (Washington, DC: Georgetown University Press), 8–15.

Hardison, D. M. (2005). Contextualized computer-based L2 prosody training: evaluating the effects of discourse context and video input. *CALICO J.* 22 (2), 175–190. doi:10.1558/cj.v22i2.175-190

Henriksen, N. (2017). Patterns of vowel laxing and harmony in Iberian Spanish: data from production and perception. *J. Phon.* 63, 106–126. doi:10.1016/j.wocn.2017.05.001

Howell, P. (1993). Cue trading in the production and perception of vowel stress. *The J. Acoust. Soc. America* 94, 2063–2073. doi:10.1121/1.407479

Hualde, J. I. (2014). *Los sonidos del español [The sounds of Spanish]*. New York, NY: Cambridge University Press.

Hualde, J., Simonet, M., and Torreira, F. (2008). Postlexical contraction of non-high vowels in Spanish. *Lingua.* 118 1906–1925. doi:10.1016/j.lingua.2007.10.004

Hulstijn, J. H., Alderson, J. C., and Schoonen, R. (2010). "Developmental stages in second-language acquisition and levels of second-language proficiency: are there links between them?," in *Communicative proficiency and linguistics development: intersections between SLA and language testing research*. Editors I. Bartning, M. Martin, and I. Vedder (Colchester, United Kingdom: European Second Language Association), 11–20.

Hutchinson, S. (1974). *Parasession on natural phonology of the regional meeting of the Chicago linguistic society*. Chicago, IL: Chicago Linguistic Society, 752–762.

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada J.* 27 (1), 1–12. doi:10.18806/tesl.v27i1.1034

Isaacs, T., and Trofimovich, P. (2012). Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Stud. Second Lang. Acquis.* 34, 475–505. doi:10.1017/s0272263112000150

Jenkins, J. (2000). *The phonology of English as an international language*. London, United Kingdom: Oxford University Press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Appl. Linguist.* 23 (1), 83–103. doi:10.1093/applin/23.1.83

Jun, A. (2015). *Prosodic typology II*. London, United Kingdom: Oxford University Press.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* 38 (2), 301–315. doi:10.1016/j.system.2010.01.005

Kjellin, O. (1999). "Accent addition: prosody and perception facilitates second language learning,". Proceedings of LP'98 (Linguistics and Phonetics Conference) at Ohio State University. Editors O. Fujimura, B. D. Joseph, and B. Palek. Columbus, The Ohio State University, September 15–20, 1998, (Prague: The Karolinum Press), 2, 373–398.

Kochetov, A., and Colantoni, L. (2011). Coronal place contrasts in Argentine and Cuban Spanish: an electropalatographic study. *J. Int. Phon. Assoc.* 41, 313–342. doi:10.1017/s0025100311000338

Kusseling, F., and Lonsdale, D. (2013). A corpus-based assessment of French CEFR lexical content. *Can. Mod. Lang. Rev.* 69 (4), 436–461. doi:10.3138/cmlr.1726.436

Lee, B. J. (2020). Enhancing listening comprehension through kinesthetic rhythm training. *RELC J.* doi:10.1177/0033688220941302

Lee, B., Plonsky, L., and Saito, K. (2020). The effects of perception- vs. production-based pronunciation instruction. *System* 88, 1–13. doi:10.1016/j.system.2019.102185

Lee, J., Jang, J., and Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: a meta-analysis. *Appl. Linguist.* 36 (3), 345–366. doi:10.1093/applin/amu040

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Q.* 39, 367–377. doi:10.2307/3588485

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. New York, NY: Cambridge University Press.

Levis, J. (2020). Revisiting the intelligibility and nativeness principles. *JSLP* 6 (3), 310–328. doi:10.1075/jslp.20050.lev

Li, A., and Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm. *Stud. Second Lang. Acquis.* 36, 223–255. doi:10.1017/s0272263113000752

Lightbown, P. (2007). "Transfer appropriate processing as a model for classroom second language acquisition," in *Understanding second language process*. Editor Z. Han. (Bristol, United Kingdom: Multilingual Matters), 27–44.

Lin, M., and Francis, A. L. (2014). The relationship between fluency, intelligibility, and acceptability of non-native spoken English. *J. Acoust. Soc. Am.* 135 (4), 2227. doi:10.1121/1.4877285

Lipski, J. (1994). *Latin American Spanish*. New York, NY: Longman.

Lipski, J. (1984). On the weakening of /s/ in Latin American Spanish. *Zeitschrift für Dialektologie und Linguistik* 51, 31–43.

Lipski, J. M. (1985). /s/ in Central American Spanish. *Hispania* 68, 143–149. doi:10.2307/341630

Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *J. Acoust. Soc. Am.* 89 (2), 874–886. doi:10.1121/1.1894649

Logan, J. S., and Pruitt, J. S. (1995). "Methodological issues in training listeners to perceive non-native phonemes," in *Speech perception and linguistic experience: issues in cross-language research*. Editor W. Strange (Walmgate, England: York Press), 351–377.

Long, A. Y., Solon, M., and Bongiovanni, S. (2018). Context of learning and second language development of Spanish vowels. *Stud. Hispanic Lusophone Linguistics* 11 (1), 59–87. doi:10.1515/shll-2018-0003

Lope Blanch, J. M. (1993a). *Ensayos sobre el español de América [Essays on American Spanish]*. Mexico City, Mexico: Universidad Nacional Autónoma de México. Instituto de Investigaciones Filológicas.

Lope Blanch, J. M. (1993b). *Nuevos estudios de lingüística hispánica [New Studies in Hispanic Linguistics]*. Mexico City, Mexico: Universidad Nacional Autónoma de México.

Lord, G., and Fionda, M. I. (2013). "Teaching pronunciation in second language Spanish," in *The handbook of Spanish second language acquisition*. Editor K. L. Geeslin (Hoboken, New Jersey: Wiley & Sons), 514–529.

Lord, G. (2005). (How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania* 88 (3), 557–567. doi:10.2307/20063159

MacLeod, B. (2012). *The effect of perceptual salience on phonetic accommodation in cross-dialectal conversation in Spanish*. PhD dissertation. Toronto, ON: University of Toronto.

Mairano, P., and Calabrò, L. (2016). "Are minimal pairs too few to be used in L2 pronunciation classes?," in *La fonetica sperimentale nell'insegnamento e nell'apprendimento delle lingue straniere. Phonetics and language learning*. Editors R. Savy and I. Alfano (Milano, Italy: Officinaventuno), 255–268.

Mar-Molinero, C., and Paffey, D. (2011). "Linguistic imperialism: who owns global Spanish?," in *The handbook of Hispanic sociolinguistics*. Editor M. Díaz Campos (Hoboken, New Jersey: Wiley), 747–764.

Martinet, A. (1978). "Function, structure and sound change," in *Readings in historical phonology*. Editors P. Baldi and R. Werth (University Park, PA: The Pennsylvania State University Press), 121–159.

Martínez Celdrán, E., and Elvira-García, W. (2019). "Description of Spanish vowels and guidelines for teaching them," in *Key issues in the teaching of Spanish pronunciation*. Editor R. Rao (London, United Kingdom: Routledge), 17–39.

Mazzaro, N., Colantoni, L., and Cuza, A. (2016). "Age effects and the discrimination of consonantal and vocalic contrasts in heritage and native Spanish," in *Romance linguistics 2013: selected proceedings of the 43th linguistic symposium on romance languages*. Editors C. Tortora, M. den Dikken, L. Montoya, and T. O'Neill (Amsterdam, Netherlands: John Benjamins), 277–300.

Meisel, J. M., Clahsen, H., and Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Stud. Second Lang. Acquis.* 3 (2), 109–135. doi:10.1017/s0272263100004137

Michnowicz, J. (2009). "Intervocalic voiced stops in Yucatan Spanish: a case of contacts-induced language change?," in *Español en Estados Unidos y en otros contextos de contacto: sociolingüística, ideología y pedagogía [Spanish in the United States and in other contact contexts: sociolinguistics, ideology and pedagogy]*. Editors M. Lacorte and J. Leeman (Mexico City, Mexico: Iberoamericana), 67–84.

Montes Giraldo, J. J. (1975). Breves notas de fonética actual del español [Brief notes on modern Spanish phonetics]. Thesaurus. Boletin del Instituto Caro y Cuervo. *Bogotá* 30 (2), 338–339.

Mora, J. C., and Levkina, M. (2017). Task-based pronunciation teaching and research. *Stud. Second Lang. Acquis.* 39, 381–399. doi:10.1017/s0272263117000183

Moreno Cabrera, J. C. (2008). *El Nacionalismo lingüístico: una ideología destructiva [Linguistic nationalism: a destructive ideology]*. Madrid, Spain: Ediciones Península.

Moreno Fernández, F. (2000). *Qué español enseñar [The Spanish to be taught]*. Madrid, Spain: Arco Libros.

Moreno Sandoval, A., Toledano, D., de la Torre, R., Garrote, M., and Guirao, J. (2008). "Developing a phonemic and syllabic frequency inventory for spontaneous spoken Castilian Spanish and their comparison to text-based inventories", in Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, Marrakech, Morocco, 26 May–1 June, 2008.

Morgan, T. (2010). *Sonidos en contexto: una introducción a la fonética del español con especial referencia a la vida real [Sounds in context: an introduction to Spanish phonetics with reference to real life]*. London, United Kingdom: Yale University Press.

Munro, M. J., and Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* 45 (1), 73–97. doi:10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., and Derwing, T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Lang. Teach.* 44 (3), 316–327. doi:10.1017/s0261444811000103

Munro, M. J., and Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: an exploratory study. *System* 34, 520–531. doi:10.1016/j.system.2006.09.004

Nuño Álvarez, M. P., and Franco Rodríguez, J. R. (2001). Ejercicios de fonética [Phonetic exercises]. Madrid: Anaya ELE.

Ong, J. H., Burnham, D., and Escudero, P. (2015). Distributional learning of lexical tones: a comparison of attended vs. unattended listening. *PLoS One* 10 (7), e0133446. doi:10.1371/journal.pone.0133446

Ong, J. H., Burnham, D., Escudero, P., and Stevens, C. J. (2017). Effect of linguistic and musical experience on distributional learning of nonnative lexical tones. *J. Speech Lang. Hear. Res.* 60 (10), 2769–2780. doi:10.1044/2016_JSLHR-S-16-0080

Ortega-Llebaria, M., Nemoga, M., and Presson, N. (2015). Long-term experience with a tonal language shapes the perception of intonation in English words: how Chinese-English bilinguals perceive 'Rose?. Vs. 'Rose'. *Bilingualism: Lang. Cogn.* 20 (2), 1–17. doi:10.1017/S1366728915000723

Ortega-Llebaria, M., and Prieto, P. (2011). Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Lang. Speech* 54, 73–97. doi:10.1177/0023830910388014

Padilla, X. A. (2015). *La pronunciación del español. Fonética y enseñanza de lenguas [Spanish pronunciation. Phonetics and language teaching]*. Alicante, Spain: Publicacions de la Universitat d'Alacant.

Piñeros, C. (2019). "The polymorphism of Spanish nasal stops," in *Key issues in the teaching of Spanish pronunciation*. Editor R. Rao (London, United Kingdom: Routledge), 126–144.

Polyanskaya, L., Ordin, M., and Busa, M. G. (2017). Relative salience of speech rhythm and speech rate on perceived foreign accent in a second language. *Lang. Speech* 60 (3), 333–355. doi:10.1177/0023830916648720

Rao, R. (2019). *Key issues in the teaching of Spanish pronunciation*. London, United Kingdom: Routledge.

Real Academia Española and Asociación de Academias de la Lengua Española. (2011). Nueva gramática de la lengua española: fonética y fonología [New grammar of the Spanish language: phonetics and phonology], Vol. III. Madrid, Spain: Espasa.

Roach, P. (1982). "On the distinction between "stress-timed" and "syllable-timed" languages," in *Linguistic controversies*. Editor D. Crystal (London, United Kingdom: Edward Arnold), 73–379.

Rosenblat, Á. (1967). El futuro de la lengua [The future of our language]. *Revista de Occidente* 56, 155–192.

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: an instructed SLA approach to L2 phonology. *Lang. Aware.* 20, 45–59. doi:10.1080/09658416.2010.540326

Saito, K., Ilkan, M., Magne, V., Tran, M. N., and Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency [New grammar of the Spanish language: phonetics and phonology]. *Appl. Psycholinguist.* 39, 593–617. doi:10.1017/s0142716417000571

Saito, K., and Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: a proposed measurement framework and meta-analysis. *Lang. Learn.* 69 (3), 652–708. doi:10.1111/lang.12345

Saito, K., Trofimovich, P., and Isaacs, T. (2016). Second language speech production: investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Appl. Psycholinguist.* 37, 217–240. doi:10.1017/s0142716414000502

Schoonmaker-Gates, E. (2017). Regional variation in the language classroom and beyond: mapping learners' developing dialectal competence. *Foreign Lang. Ann.* 50 (1), 177–194. doi:10.1111/flan.12243

Schwegler, A., and Ameal-Guerra, A. (2019). *Fonética y fonología* españolas *[Spanish phonetics and phonology]*. Hoboken, New Jersey: Wiley.

Sicola, L., and Darcy, I. (2015). "Integrating pronunciation into the second language classroom," in *The handbook of English pronunciation*. Editors M. Reed and J. Lewis (New York, NY: Cambridge University Press, 471–487.

Sosa, J. M. (1999). *La entonación del español [Spanish intonation]*. Madrid, Spain: Ediciones Cátedra.

Terrell, T. D. (1978). Sobre la aspiración y elisión de /s/ implosiva y final en el español de Puerto Rico [On the aspiration and deletion of implosive /s/ in Puerto Rican Spanish]. *NRFH.* 27, 24–38. doi:10.24201/nrfh.v27i1.1705

Torreira, F., and Ernestus, M. (2012). Weakening of intervocalic /s/ in the Nijmegen corpus of casual Spanish. *Phonetica* 69 (3), 124–148. doi:10.1159/000343635

Tuninetti, A., Mulak, K. E., and Escudero, P. (2020). Cross-situational word learning in two foreign languages: effects of native language and perceptual difficulty. *Front. Commun.* 5, 109. doi:10.3389/fcomm.2020.602471

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

Villegas Rogers, C., and Medley, F. W., Jr. (1988). Language with a purpose: using authentic materials in the foreign language classroom. *Foreign Lang. Ann.* 21 (5), 467–478. doi:10.1111/j.1944-9720.1988.tb01098.x

Wanrooij, K., Escudero, P., and Raijmakers, M. E. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *J. Phonet.* 41 (5), 319–102. doi:10.1016/j.wocn.2013.03.005

Warren, P., Elgort, I., and Crabbe, D. (2009). Comprehensibility and prosody ratings for pronunciation software development. *Lang. Learn. Tech.* 13 (3), 87–102.

Wrembel, M. (2007). "Metacompetence-based approach to the teaching of L2 prosody: practical implications," in *Non-native prosody: phonetic description and teaching practice*. Editors J. Trouvain and U. Gut (Berlin, Germany: Mouton de Gruyter), 189–209.

Yazawa, K., Whang, J., Kondo, M., and Escudero, P. (2020). Language-dependent cue weighting: an investigation of perception modes in L2 learning. *Second Lang. Res.* 36 (4), 557–581. doi:10.1177/0267658319832645

Zárate-Sández, G. (2019). "Spanish pronunciation and teaching dialectal variation," in *Key issues in the teaching of Spanish pronunciation: from description to pedagogy*. Editor R. Rao (London, United Kingdom: Routledge), 201–217.

Zielinski, B. (2015). "The segmental/suprasegmental debate," in *The handbook of English pronunciation*. Editors M. Reed and J. Lewis (Hoboken, New Jersey: john wiley and sons), 397–412.

Conflict of Interest: 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.

frontiers in Communication

ORIGINAL RESEARCH
published: 22 April 2021
doi: 10.3389/fcomm.2021.636122

The Effects of ESL Immersion and Proficiency on Learners' Pronunciation Development

*Maria Kostromitina and Okim Kang**

Northern Arizona University, Flagstaff, AZ, United States

Despite the efforts of existing studies in the domain of L2 phonology to examine ESL learners' pronunciation development, little research has comprehensively demonstrated ESL learners' pronunciation improvement in academic immersion contexts. Similarly, few studies have focused on learners' proficiency levels linked to their developmental success. The current exploratory study investigated the changes of learners' pronunciation constructs as a result of their ESL program. Seventy-five newly arrived ESL students (25 in each proficiency; beginner, intermediate, and advanced) enrolled in an Intensive English Program in the United States provided their speech responses (to the placement and exit tests from the program). One hundred fifty speaking samples were linguistically analyzed for the following suprasegmental features: fluency (speech rates and pauses) and prosody (prominence and pitch range). Segmental features were analyzed by employing a functional load approach with randomly selected 90 speech files. Findings revealed different developmental patterns among phonological features and proficiency levels; that is, the upper-level learners improved more in fluency and prominence than the lower-level learners. Segmental changes were minimal, suggesting that both high functional and low functional load sounds involve a complex process in learning. Overall findings provide important implications for ESL curriculum planning and development: 1) intonation acquisition can be difficult; 2) skill improvement differs by proficiency level; and 3) level-specific curriculum may be needed.

Keywords: ESL immersion, L2 pronunciation, L2 proficiency, functional load, fluency

OPEN ACCESS

Edited by:
Andrew Sewell,
Lingnan University, China

Reviewed by:
Paul Stapleton,
Other, Toronto, ON, Canada
Dustin Crowther,
University of Hawaii at Manoa,
United States

****Correspondence:***
Okim Kang
okim.kang@nau.edu

Specialty section:
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: *30 November 2020*
Accepted: *19 February 2021*
Published: *22 April 2021*

Citation:
Kostromitina M and Kang O (2021) The Effects of ESL Immersion and Proficiency on Learners' Pronunciation Development. Front. Commun. 6:636122. doi: 10.3389/fcomm.2021.636122

INTRODUCTION

Recent developments in the field of second language (L2) speech have instigated a shift of the emphasis on intelligible speech over the unrealistic goal of sounding native-like (Derwing and Munro, 2005; Field, 2005). Following this trend, instead of a complete absence of a foreign accent, second language (L2) English speakers are encouraged to strive for clear and intelligible speech. Previous research on L2 speech has identified that functional load–based segmental deviations can predict speaker intelligibility (Kang et al., 2020) or distinguish proficiency levels (Kang and Moran, 2014). Other studies have demonstrated that suprasegmental features, including fluency, prominence, and intonation, have an even larger role in promoting L2 speakers' intelligibility and comprehensibility in oral communication (Derwing and Rossiter, 2003; Pickering, 2004; Kang et al., 2010).

At the same time, study abroad (SA) experiences or immersion (IM) contexts, especially combined with instruction, are known to be beneficial for L2 learners' pronunciation development (Stevens, 2002; Lord, 2010). However, this development can happen differently depending on certain learner factors. In a series of longitudinal studies with Mandarin and Slavic ESL learners in Canada, Derwing and colleagues investigated the development of comprehensibility, fluency, and accentedness of L2 speakers as well as the relationship between L1 and L2 fluency (Derwing et al., 2006; Derwing et al., 2008; Derwing et al., 2009). The studies demonstrated differences in the improvement of these L2 pronunciation features between the two groups, emphasizing the role of motivational, cultural, and interactional factors in L2 pronunciation development in immersion (Derwing and Munro, 2013).

Such individual differences can also predict learners' success in developing linguistic competence as a result of L2 immersion. For example, after studying abroad, advanced learners brought more formulaic expressions to communicative contexts whereas beginners attended primarily to meaning (Lafford, 2004). Segalowitz and Freed (2004) found that among Spanish L2 learners, an initial threshold level of basic word recognition and lexical access processing abilities was necessary for a significant oral proficiency development in the immersion context. In their study, the most important gains that learners of Spanish made abroad were in the domain of speech fluency.

Regardless of the empirical evidence about the benefits of immersion experience in language development, however, few studies have comprehensively demonstrated L2 learners' pronunciation improvement, particularly with a focus on their ESL immersion experience through a functional load approach. Similarly, how learners' proficiency is linked to their developmental success in L2 pronunciation has been largely unknown. The present study explores the effects of ESL immersion on learners' pronunciation improvement across different levels of learners' proficiency. The results of the study shed light on pronunciation development of L2 English learners and offer important implications for curriculum design of ESL immersion programs.

REVIEW OF LITERATURE

ESL Immersion and Learners' L2 Proficiency

When discussing linguistic gains in an immersion environment, it is important to understand what immersion entails. Freed et al. (2004) defined immersion as a combination of classroom-based learning with expected outside activities in the at home environment. In an ESL context, immersion is most commonly associated with Intensive English Programs (IEP) at universities. Such programs are different in their duration and are typically designed for learners from various proficiency levels. Beyond IEP classrooms, immersion often includes opportunities for interaction with the native speech community and out-of-class activities.

In light of IEP, research has been focused on identifying the most favorable period in learners' L2 development when they can gain the most from immersion, bringing forward the idea of "proficiency threshold." Despite the general consensus about the benefits of immersion, there is little agreement in the existing literature on the most beneficial time for a language learner to immerse in the target language. Attempts have been made to determine the optimal point in learners' proficiency to be immersed for it to result in noticeable L2 development. For example, Kang and Ghanem, (2016) found with the help of a nation-wide self-report survey that the intermediate level was somewhat more beneficial than other levels for immersion programs. Other researchers (e.g., Brecht et al., 1995; Martinsen, 2008; Collentine, 2009; Muñoz and Llanes, 2014) argued that beginning-level language learners demonstrate the greatest amount of improvement in oral and aural communication skills as a result of ESL immersion. In contrast, others (e.g., Davidson, 2010) called for more studies to focus on advanced-level learners to validate the current findings.

Notwithstanding the research discrepancies, low-level proficiency learners seem to benefit from immersion contexts. As Martinsen, (2008) pointed out, true beginners are likely to be slower to progress than learners with at least some experience with the target language indicating "there may be a minimal level of proficiency at which learning abroad is optimal" (p. 506). Additionally, for students' oral abilities to improve, learners need at least a basic level of word recognition and processing abilities (Segalowitz and Freed, 2004). When discussing the proficiency threshold in the context of ESL immersion, however, learners' linguistic competence is indeed a complex construct (Collentine, 2009). How learners' proficiency levels relate to their learning gains in the immersion context needs further validation.

L2 Pronunciation Improvement in the Immersion Context

Pronunciation in an L2 is commonly operationalized as "aspects of oral production of language, including segments, prosody, voice quality, and rate" (Derwing and Munro, 2015, p. 5). However, as Derwing and Munro (2015) note, in interactional situations, pronunciation encompasses broader dimensions of communication that include *accentedness*, or particular patterns of pronunciation that distinguish members of speech communities; *comprehensibility*, or the amount of effort with which a listener understands L2 speech; and *intelligibility*, or the degree to which a message is received as intended by a listener. Another speech dimension related to pronunciation is fluency, or the rate and degree of fluidity of speech, indicated by the absence of pauses or other disfluency markers.

In ESL immersion contexts, L2 pronunciation has received relatively consistent attention in the field of SLA. Overall findings point out that advances in pronunciation over time are a slow or unchanging process (Avello, 2010; Pérez-Vidal et al., 2011). Trofimovich and Baker (2006) examined the changes in English learners' speech rate as a result of their U.S. immersion experience over different time periods (i.e., three months, three years, and ten months) and showed that there was no significant difference in the speech rate among the learners. Despite the length of residency in the U.S., speech rate did not necessarily change for those learners. In fact, the authors suggested that certain suprasegmental features, including speech rate and tonal peaks, may never be learned to a native-like proficiency (Trofimovich and Baker, 2006).

More recently, Højen (2019) examined effects of short-term immersion (3–10 months) on adult L2 English pronunciation. In the study, native English speakers evaluated the accentedness in speech samples from three experimental groups: a native Danish au pair group with experience living in England, a native Danish control group with no such experience, and a native English reference group. For the experience group, speech samples were recorded twice, before and after the immersion. The results of the study revealed significant improvements when compared before and after the immersion with a great degree of variation. Interestingly, the pronunciation score for the immersion group was significantly correlated with the length of residence in England (r = 0.61) suggesting that an immersion of at least five months is needed for noticeable improvement in L2 pronunciation. More longitudinal research can confirm such developmental patterns.

Pronunciation improvement over time in an ESL immersion context can be confounded by learners' first language (L1) background, even though learners' L1 is not the primary focus of the current study. In a longitudinal study over ten months, Derwing et al. (2006) evaluated the progress in Slavic and Chinese Canadian immigrants' accent and fluency over a period of 10 months. Both groups only showed a small improvement in accentedness over time, but the two L1 groups had different results in terms of fluency. The Slavic learners demonstrated significant improvement in fluency, while the Mandarin group did not. Considering these same L1 groups over a two-year period, Derwing et al. (2008) found similar results in their later research. In a more recent study, Derwing and Munro (2013) examined the extent to which the same two groups of learners continued to make progress in the development of oral English skills after finishing their formal ESL training. Similarly to earlier findings, Slavic speakers improved comprehensibility and fluency using English outside of the classroom context while Mandarin speakers showed much less improvement. The authors proposed that MacIntyre's (2007) Willingness To Communicate framework could account for the differences underlying the performance of the two groups including ties to the L1 community, reluctance to initiate conversations, and lack of opportunities to interact in English.

While there is developmental variation among different L1 groups, fluency certainly seems to benefit the most in an immersion context (Freed et al., 2004; Segalowitz and Freed, 2004). Towell (2002) further added that the initial fluency of the learners determined learners' fluency gains; in other words, fluency improved most significantly at the lower levels than at the higher levels. Overall, it seems essential to systematically determine what pronunciation properties underlying comprehensibility, intelligibility, accentedness, and fluency of L2 speech are actually changeable or learnable over time especially in such an immersion context and to what extent these properties may improve as a result of immersion.

Functional Load-Based Segmental Features in L2 Pronunciation

Segmental analyses of pronunciation often involve an examination of deviations from the native baseline or substitution of sounds in L2 speech (e.g., fun spoken like fan, Isaacs and Trofimovich, 2012). Other studies (e.g., Kang and Moran, 2014) have categorized the segmental deviations in terms of their relative weight in predicting listeners' judgments. Not all segmental deviations can have equal effects on listeners' understanding (see also Fayer and Krasinski, 1987) and a more "nuanced approach" is needed (Isaacs and Trofimovich, 2012).

An effort that has been made to categorize segmental deviations is the Functional Load (FL) theory (Catford, 1987; Brown, 1991). Munro and Derwing (2006) further explain that segmental pairs (e.g., pet vs. bet or dis vs. this) are ranked based on factors including the probability that individual members of the minimal pair are valid, the frequency of the minimal pair, and the position of the segmental within a word. Thus, if an L2 speaker substitutes 'them' for 'dem', it is unlikely that their comprehensibility is severely affected in a negative way. In contrast, the/d/-/*p*/contrast has a high functional load in English meaning (e.g., day–pay). Munro and Derwing (2006) demonstrated that high FL divergences had larger effects on listeners' perceptions of accentedness and comprehensibility of L2 speech than low FL deviations thus providing preliminary support for the theory.

In an L2 assessment study, Kang and Moran (2014) classified segmental deviations according to their FL to determine their effect on oral assessment across four proficiency levels (B1–C2). The analysis of vowel and consonant substitution divergences through the FL approach detected a significant difference across proficiency levels in the high FL deviations. That is, with an increase in learners' proficiency, the amount of high FL deviations dropped significantly. However, changes in low FL deviations were not noticeable across levels.

In a recent study, Suzukida and Saito (2019) re-examined the FL approach to evaluating the effect of segmental divergences on L2 comprehensibility in two experiments with learners in EFL and immersion settings. In the first experiment, the speech of Japanese learners of English in EFL settings was assessed in terms of perceived comprehensibility by L1 English raters. The second experiment was slightly different with the speakers being Japanese learners of English with immersion experience. Their findings also showed that only high FL consonant substitutions negatively affected native listeners' comprehensibility judgments in both experiments significantly. Importantly, high FL consonant substitutions impeded raters' comprehensibility regardless of task conditions. Kang et al., 2020 recent study also confirmed that divergences in high FL vowels and consonants strongly predicted listener comprehension and intelligibility scores.

It is evident that FL-based segmental pronunciation features play a critical role in listener judgments and perceptions of L2 speech. However, while some studies analyzed the differences in FL deviations in speech of L2 learners from different proficiencies (e.g., Kang and Moran, 2014), there have been only a few studies that investigated the development of segmental features over time. In particular, Kang et al., 2021 (in press) recent study examined how EFL learners developed their speaking skills by analyzing fifty-two EFL learners' IELTS spoken responses over the period of three months. The study comprehensively analyzed segmental and suprasegmental pronunciation features but did not find significant improvements in segmentals. The authors called for further research.

Suprasegmental Features of L2 Pronunciation

An increasing number of studies have addressed the importance of suprasegmentals, such as fluency, stress, and intonation, in listeners' judgments of accentedness and comprehensibility of L2 speakers (Munro and Derwing, 2001; Isaacs, 2008; Kang et al., 2010). Research indicates that fluency, characterized by the speaking rate, number and length of pauses, and repair fluency, is linked to listeners' comprehension of speech and the overall evaluation of speakers' oral proficiency (Derwing et al., 2004; Tavakoli and Skehan, 2005; Iwashita et al., 2008). With regard to accentedness, Trofimovich and Baker (2006) indicated that accentedness ratings given by L1 English speakers to Korean learners of English were higher when the L2 speech was faster. In fact, several studies have suggested certain speaking rate thresholds, which make the perception of L2 speech more comprehensible and less accented (Isaacs, 2008; Munro and Derwing, 2001). Kang et al. (2020) recently reported that temporal fluency measures predicted listener comprehension and intelligibility scores. Other studies showed that pause frequency and duration affected accentedness and comprehensibility ratings (Trofimovich and Baker, 2006; Kang et al., 2010).

Other suprasegmental features that have been found especially indicative of L2 pronunciation development include nuclear stress and pitch range. It has been established that placing incorrect sentence stress or emphasizing every word in a run, regardless of its function or importance to the communicative purpose, can negatively affect listeners' comprehension (Juffs, 1990; Wennerstrom, 2000; Field, 2005). Similarly, pitch range that is too narrow can considerably diminish L1 listeners' comprehension of L2 speech (Pickering, 2001) and cause misunderstandings (Kang, 2012). Moreover, in conjunction with accentedness ratings and suprasegmentals, Kang (2010) found that pitch range alone explained 24% of the variance in accentedness ratings, with narrow pitch range being associated with stronger accents. Thus, the particular suprasegmental features that were measured in this study (fluency, stress, and pitch range) reflect previous research that has systematically shown the importance of these features for the perceptions of L2 speech accentedness and comprehensibility making them crucial in L2 pronunciation.

Similar to that of segmental research, however, the development of suprasegmental features has rarely been studied, especially from a longitudinal perspective. Kang et al., 2021 (in press) demonstrated EFL learners' fluency and intonation changes over 12 weeks, but their findings are very limited. Generally, research that has comprehensively examined the production of both segmental and suprasegmental features, especially pertaining to their development over time, is scarce. In addition, little is known about the way learners at different proficiency levels develop pronunciation skills in an immersion context. In an attempt to fill these gaps, the present exploratory study systematically examines the interplay of proficiency and immersion on the changes in L2 learners' FL-based segmental and suprasegmental pronunciation. The present study addresses the following research question:

> To what extent does ESL learners' pronunciation develop as a result of one-semester long ESL immersion across the proficiency levels?

1) In terms of functional load-based segmentals in terms of suprasegmentals (fluency, pitch, sentence stress).

METHODS

Participants

The study recruited seventy-five ESL students enrolled in listening and speaking courses in an Intensive English Program (IEP) at a southwestern university in the United States. There were 25 participants from each of the three proficiency levels in the program: beginning, intermediate, and advanced. These proficiency levels were determined through an in-house placement test, which mimicked the standardized iBT TOEFL test of English proficiency. Level 1, or beginner, corresponds to scores below 15 of the TOEFL iBT; Level 3, intermediate, corresponds to scores between 32–44; and Level 5, upper-intermediate, corresponds to scores between 57–69. The first language of 59 speakers was Arabic while the remaining 16 participants were native speakers of Chinese. Most of them (90% of the participants) just arrived in the U.S. and started the IEP program for the first time while the remaining 10% arrived slightly earlier (1–2 weeks).

Speaking Tasks and Collection of Speech Samples

There were two stages of speech sample collection. The first collection of the samples took place during the first week of classes (pre-immersion) as a placement test to determine the appropriate level for the learners in the IEP listening and speaking class. The second speech sample collection happened during week 15 (post-immersion) when the learners completed an exit test to demonstrate that they successfully finished the course and were ready to move up to the next level in the program or graduate from the IEP.

During both times of the speech collection, the participants completed the same speaking task. The task consisted of an oral prompt that asked the participants to speak on the following topic: Some students prefer university in their home country, while others prefer studying abroad. What do you prefer? Give

TABLE 1 | Summary of speakers response length before and after the immersion.

	Pre-immersion (time 1)			Post-immersion (time 2)		
	Beginner	Intermediate	Advanced	Beginner	Intermediate	Advanced
N	25	25	25	25	25	25
M (sec)	45.68	63.25	88.83	58.68	62.02	79.80
SD	21.12	13.74	18.29	18.05	12.12	24.62

TABLE 2 | Pronunciation feature analysis and measures.

Categories	Feature	Description
Segmental measures	High FL substitutions	This measure calculates the number of high functional load consonant and vowel substitutions in word initial, word medial, or word final positions. e.g., “day”–“bay”; “cat”–“cot”
	Low FL substitutions	This measure calculated the number of low functional load consonant and vowel substitutions in word initial, word medial, or word final positions. e.g., “this”–“zis”; “walking”–“wolking”
Fluency measures	Syllables per second	This is a measure of the mean number of syllables produced per second, calculated as the total number of syllables divided by the total length of the speech sample
	Number of silent pauses per second	This measure is the number of silent pauses per second, calculated as the total number of pauses (over 0.1 s) divided by the total length of the speech sample.
	Number of hesitation markers per second	This measure is calculated as the total number of filled pauses divided by the total length of the speech sample. Filled pauses include hesitation markers or fillers such as *uh* or *um* but do not include repetitions, restarts, or repairs.
Stress Measure	Space	This measure calculates the proportion of prominent words to the total number of words.
Intonation measure	Overall pitch range	This measure calculates the pitch range of the sample based on the F_0 minimum and maximum frequency point on all prominent syllables within the speech sample.

reasons and examples to support your opinion. The task was designed to elicit monologic speech samples from the speakers. While presenting the same task to the participants before and after the immersion could present a potential risk of task familiarity effect, the importance of increased comparability of the speech samples outweighed this concern. Each participant was permitted to ask questions regarding the content of the prompt and any unknown vocabulary and given 1.5 min to prepare their response. The participants were then advised to speak for about 1–2 min on the topic and were recorded the entire time. The final dataset included pre- and post-immersion speech samples from the 75 participants. There were a total of 150 speech files varying from 30 s to 2 min each; that is, each participant contributed two sound files to the dataset. **Table 1** below presents the summary of the descriptive statistics (*M* and *SD*) of the speech samples before and after the immersion.

Speech Analysis

To analyze the development patterns of the pronunciation of segmentals and suprasegmentals in the speech samples, the 150 sound files were first transcribed by three trained coders. Next, FL deviations in a randomly sampled subset of 90 sound files were analyzed, calculated, and averaged per minute. It was deemed appropriate to subsample the files for the FL analysis as this process required a much more meticulous examination of the speech samples; thus, the difference in the sample size between segmental and suprasegmental analyses happened due to the labor intensiveness of speech analysis involved in different types of speech features. This subset consisted of 45 speech samples collected before the immersion and 45 speech samples collected after the immersion produced by the same speakers (15 speech samples per proficiency level in both cases). The FL deviations were categorized into two groups: high FL divergences and low FL consonant deviations (Catford, 1987; Kang et al., 2020). Following Munro and Derwing (2006) and Kang and Moran (2014), we considered the substitutions that ranked between 51 and 100% in Catford’s framework as high FL divergences, and those below were regarded as low. To identify these deviations, the coders listened and transcribed the speech files noting all the instances when a speaker’s pronunciation deviated from Standard American English (SAE). The coders then used Catford’s FL framework to assign a functional load value (percentage) to the deviations. Additionally, we analyzed the suprasegmental measures for all of the 150 speech samples.

The complete list of measures is presented in **Table 2**. The specific segmental measures selected for the pronunciation analysis have been found to differ distinctively across proficiency levels (e.g., Kang and Moran, 2014). The suprasegmental measures chosen for this analysis are also grounded in the aforementioned previous research that emphasized their weight for comprehensibility and accentedness judgments of L2 speech and for L2 pronunciation overall (e.g., Pickering, 2001; Kang, 2010; Kang et al., 2010). In accordance with Kang (2010), the runs in the speech samples were operationalized as stretches of undisturbed speech delimited by pauses of 0.1 or longer assuming that this pause cut-off would be meaningful in L2 speech.

To prepare the speech samples for analysis, they were converted into. wav format with the help of Audacity, a free audio software (Audacity Team, 2020) and transcribed using standard conventions. Then, three trained coders analyzed the

FL-based deviations in the samples. The coders were unaware of the proficiency levels of the speakers or any other identifying information. One of the coders calculated inter-coder reliability post hoc by re-analyzing 10% of the speech and reaching over 85% agreement with each coder. All remaining differences were later discussed and resolved reaching 100% agreement among the raters. As for the suprasegmental features, Praat (Boersma and Weenink, 2020) was used to conduct temporal and acoustic analyses. Spectrograms with extracted pitch contours were used to identify the prominent syllables in runs and consequently measure the pitch on prominent syllables as well as the length and number of silent and filed pauses. Two coders performed analysis of the suprasegmental features after reaching an agreement of 90% or higher on a subset of data for all variables.

Statistical Analysis

The main research question in the present study addressed the change in FL-based segmental and suprasegmental pronunciation features of beginner, intermediate, and advanced ESL learners as a result of one semester-long immersion. To answer this research question, seven linear mixed effect models (LMEM) were performed with each of the seven pronunciation features in **Table 1** as dependent variables and proficiency and time pre-/post-immersion as independent variables. The three pre-determined proficiency levels in the study and the time of speech sample collection (Time 1 and Time 2) were entered as fixed factors. Participants and their L1 backgrounds were entered as the random factors. Linear mixed effects modeling was considered appropriate for the current analysis since it allowed to investigate the effects of immersion, proficiency, and their interaction on pronunciation features while controlling for the participant and L1 (Arabic and Chinese) background factors. All statistical procedures in the study were completed with the help of *R* (ver. 4.0.2), a free statistical environment (R Core Team, 2020). Before fitting the models onto the data, seven scatterplots were created to examine the data for violations of normality, linearity, and homoscedasticity. After winsorization of the outliers, the data met the assumptions.

RESULTS

The goal of the present study was to examine the change in segmental and suprasegmental features in speech of beginner, intermediate, and advanced English learners as a result of ESL immersion. The following section provides a detailed description of the results of the mixed effects models fitted on the data using one dependent variable at a time. First, the results of the segmental analysis are given with regards to the high and low FL deviations in the speech samples. Both types of FL deviations each included consonant and vowel substitutions. Then, we present the analysis summary of the five suprasegmental features that represented fluency (speaking rate, silent pauses, and filled pauses), stress (space), and intonation (pitch range) in our study.

Development of FL-Based Segmentals

The first two LMEMs fitted on the data focused on High and Low FL substitutions in the speech samples. Importantly, to address the development of FL-based segmentals, high FL vowels and consonants were merged as well as low FL vowels and consonants. While we initially attempted to analyze them separately, the analysis revealed similar deviation patterns among the high FL vowels and consonants and low FL vowels and consonants. To allow for a more robust sample, the categories were combined into two groups: high FL substitutions (vowels and consonants) and low FL substitutions (vowels and consonants). We paid particular attention to these deviation types in our analysis being guided by previous research findings about the effect of high FL deviations on listeners' perceptions and learners' oral performances (e.g., Munro and Derwing, 2006; Kang and Moran, 2014).

The descriptive statistics for the segmental features are given in **Table 3**. The table summarizes the means, standard deviations, and 95% confidence intervals of high and low FL deviations before and after immersion for each of the three proficiency levels in the study. **Figure 1** provides an additional visual representation of the distributions of these features. Note that the plot represents the distribution of High and Low FL deviations for each proficiency level before and after the immersion. The boxes in the middle of the figure represent the interquartile range, the dark line in the middle of each box is the median, and the blue dot is the mean. The outliers are indicated by the gray circles outside of the overall range.

High and Low FL Deviations

As can be seen in **Figure 1** and **Table 3**, intermediate students improved by reducing the amount of high FL consonant and vowel deviations; in contrast, beginning and advanced learners demonstrated a higher amount of such deviations in their speech samples after the ESL immersion. It is noteworthy that the intermediate group had the largest amount of high FL substitutions of the three groups in the pre-test, while beginners displayed the highest number of substitutions in the post-test. The advanced learners had fewer high FL deviations in the pre-test than the intermediate group; however, their scores were almost equal in the post-test. Comparing both types of deviations for each proficiency group, the general patterns revealed differences in the distribution of high and low FL divergences in the pre- and post-test conditions. **Figure 1** shows that while beginners improved on the low FL substitutions, their production of high FL deviations increased. In contrast, intermediate learners demonstrated an opposite trend increasing low FL but reducing high FL divergences after immersion. The advanced group did not seem to show prominent changes in FL substitutions. None of the differences were significantly different as indicated by the 95% CIs overlapping by more than half in **Table 3** (see Cumming, 2009 for a detailed discussion). The only marginally significant result emerged for the beginner group in the amount of high FL deviations, $t = 1.97$, $p = 0.51$.

After fitting the linear mixed effects model on the data to examine high FL errors, the summary *F*-statistic obtained for the model was not significant *F* (5,84) = 0.473, $p = 0.79$. No significant interaction between the two fixed effects (Time and Proficiency) was observed, *F* (2,42) = 1.69, $p = 0.19$ indicating that the immersion did not necessarily improve the participants' pronunciation of FL-based segmental features across proficiency levels with regard to high

TABLE 3 | Summary of high and low functional load deviations for each proficiency group pre- and post-immersion.

	High FL deviations											
	Beginner				Intermediate				Advanced			
	M	SD	95% CI		M	SD	95% CI		M	SD	95% CI	
			U	L			U	L			U	L
Pre-immersion	3.37	2.26	2.23	4.51	4.57	3.72	2.69	6.45	3.55	1.97	2.55	4.55
Post-immersion	4.37	2.54	3.08	5.66	3.64	2.22	2.52	4.76	3.81	4.20	1.68	5.94
	Low FL deviations											
	Beginner				Intermediate				Advanced			
	M	SD	95% CI		M	SD	95% CI		M	SD	95% CI	
			U	L			U	L			U	L
Pre-immersion	5.48	4.08	3.42	7.54	4.21	2.75	2.82	5.60	4.80	2.86	3.35	6.25
Post-immersion	4.55	3.25	2.91	6.19	4.80	4.50	2.52	7.08	4.60	4.34	2.40	6.80

CI = confidence interval, U = upper 95% CI, L = lower 95% CI.

functional load divergences. The main effects of Proficiency and Time (pre/post) were not significant ($p = 0.83$), meaning that neither the high FL-based segmental changes happened significantly over 15 weeks, nor did proficiency make a difference. Overall, the fixed effects of Proficiency and Immersion, together with the random effects, explained close to 57% of the variance in High FL substitutions (conditional $R^2 = 0.569$). However, the fixed factors by themselves accounted for only 2% of the variance (marginal $R^2 = 0.022$) suggesting that the majority of variance was explained by participant factors and their L1s. To calculate the relative variance explained by the random factors, we transformed the variance values into proportions. The random idiosyncrasies of participants explained additional 42.5% of the variance; however, only 14% of variance was explained by the two L1s in the sample. No post hoc analysis was performed because neither interaction effect nor main effect was significant in this model.

Another similar LMEM was computed to examine the effect of proficiency and immersion on the production of low functional load substitutions. The results summarized in **Table 3** above showed that the average number of low FL substitutions did not change noticeably for each group. Although the standard deviation in each group was quite large, on average, the beginner and advanced learners made fewer deviations of this nature in the post-immersion speech collection, and the intermediate group made low FL substitutions more frequently after the immersion. Similarly to high FL substitutions, the model did not reveal a significant interaction between the two fixed factors of proficiency and immersion, F (2,42) = 0.46, $p = 0.63$.

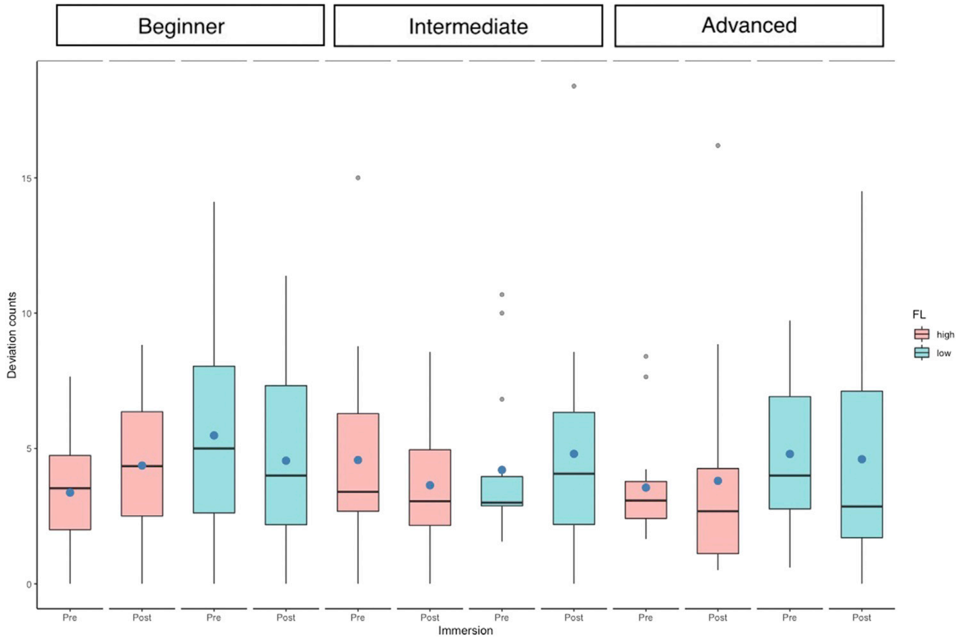


FIGURE 1 | Summary of results for segmental deviations per proficiency group before and after immersion.

TABLE 4 | Summary of speaking rate for each proficiency group pre- and post-immersion.

	Speaking rate (syllables per second)											
	Beginner				Intermediate				Advanced			
	M	SD	95% CI		M	SD	95% CI		M	SD	95% CI	
			U	L			U	L			U	L
Pre-immersion	2.16	0.50	1.96	2.36	2.35	0.53	2.14	2.56	2.50	0.37	2.35	2.65
Post-immersion	1.96	0.36	1.82	2.10	2.48	0.45	2.30	4.66	2.81	0.35	2.67	2.95

CI = confidence interval, U = upper 95% CI, L = lower 95% CI.

TABLE 5 | Summary of silent and filled pauses across proficiency levels pre- and post-immersion.

	Beginner				Intermediate				Advanced			
	M	SD	95% CI		M	SD	95% CI		M	SD	95% CI	
			U	L			U	L			U	L
						Silent pauses						
Pre-immersion	0.39	0.10	0.35	0.43	0.51	0.11	0.47	0.55	0.45	0.09	0.42	0.49
Post-immersion	0.46	0.09	0.43	0.49	0.49	0.08	0.46	0.52	0.47	0.12	0.42	0.52
						Filled pauses						
Pre-immersion	0.24	0.13	0.19	0.29	0.19	0.09	0.16	0.23	0.20	0.11	0.16	0.24
Post-immersion	0.15	0.10	0.11	0.19	0.24	0.11	0.20	0.28	0.26	0.15	0.20	0.32

CI = confidence interval, U = upper 95% CI, L = lower 95% CI.

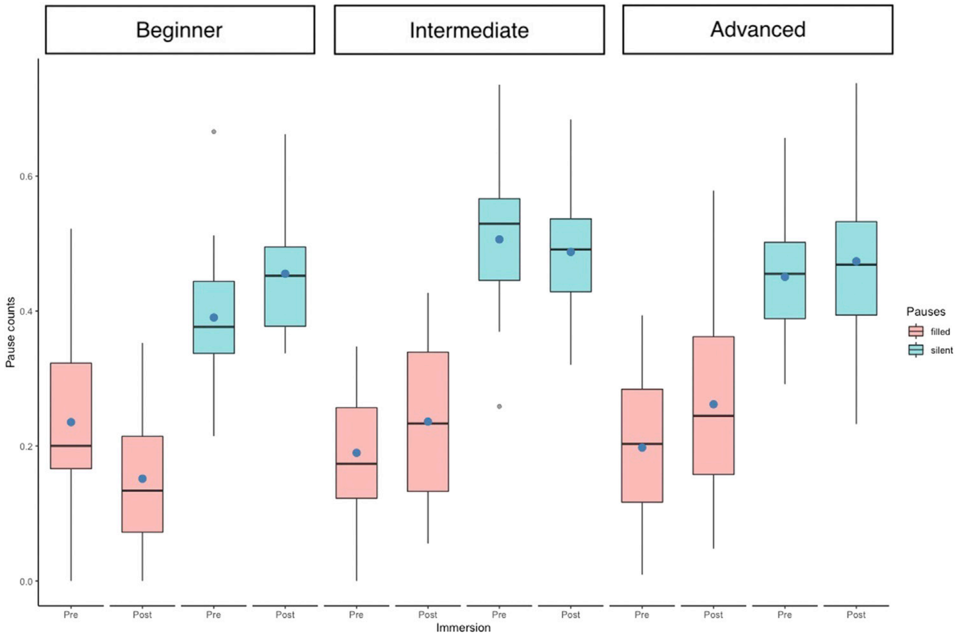


FIGURE 2 | Summary of results for silent and filled pauses for each proficiency group before and after immersion.

No significant main effects of time (pre/post), F (1,42) = 0.8834, p = 0.35, or Proficiency, F (2,42) = 0.10, p = 0.90 were detected either. Overall, this LMEM accounted for 72% of variance in the dependent variable (conditional R^2 = 0.718), although the fixed effects are responsible for only 1% of this variance (marginal R^2 = 0.009). The random participant factors explained over 70% of the variance. There was a small percentage of variance (3.2%) accounted for by participants' L1.

As stated earlier, additional LMEM analysis was done separately for high FL vowels and consonants as well as for low FL vowels and consonants. The emerged trends were similar to the main analysis presented here with none of the models being significant (F (5,84) = 1.574, p = 0.176 for high FL vowel deviations, F (5,84) = 0.229, p = 0.949 for low FL vowel deviations, F (5,84) = 0.849, p = 0.519 for high FL consonant deviations, and F (5,84) = 1.265, p = 0.287 for low FL consonant deviations).

Development of Suprasegmentals

The next five LMEM tested the effect of immersion on the development of suprasegmental features of pronunciation, namely, frequency measured by the speaking rate and the number of silent and filled pauses, stress measured by space, and intonation measured by pitch range. As stated earlier, the complete dataset of 150 speech files was used in building the five models. **Tables 4**, **5** as well as **Figure 2** present the summaries of descriptive statistics for the three fluency measures.

Syllables per Second

The summary of descriptive statistics for the speaking rate in **Table 4** indicates that the intermediate and advanced learners showed improvement in the speaking rate in the post-immersion speech collection whereas the beginners were slower. There is also a clear difference between the proficiency levels and their respective average speaking rate with beginners being the slowest and the advanced learners being the fastest.

The LMEM with speaking rate as a dependent variable was significant, as indicated by the summary statistic, F (5,150) = 12.47, $p < 0.01$. Moreover, the interaction between proficiency and time of testing (pre- and post-immersion) was also significant, F (2,75) = 8.54, $p < 0.01$. In order to find out which proficiency groups significantly improved this aspect of fluency after immersion, post hoc pairwise comparisons using Tukey HSD were calculated. The results revealed that advanced learners were significantly faster after immersion than before, $t = 3.49$, $p < 0.01$. This result revealed Cohen's d = 0.86, which is considered a medium to large effect based on meta-analytically determined effect size guidelines for applied linguistics (Plonsky and Oswald, 2014). However, the speech rate of intermediate students did not change much showing a small effect size, $t = -1.53$, $p = 0.65$, Cohen's $d = 0.35$. The slow-down in beginners' speech was also not significant, $t = 2.16$, p = 0.27 with small to medium Cohen's $d = 0.50$. Proficiency and immersion, together with the random factors, were able to explain 60.5% of the variance (conditional $R^2 = 0.605$) and the fixed factors by themselves accounted for almost half of that variance (29%, marginal R^2 = 0.286). The participant differences explained most of the variance that occurred due to random factors while L1 background contributed less than 1% to the model.

Number of Silent and Filled Pauses

Table 5 below summarizes the average number of silent and filled pauses in the speech samples normalized per second of speech. In terms of the silent pauses in the speech samples, **Table 4** illustrates that only the intermediate group demonstrated a decline in their amount after ESL immersion. The participants from the other two levels used more silent pauses in the post-test, although the increase was barely noticeable. Interestingly, the number of silent pauses per second seemed to increase from beginner to intermediate speakers before the immersion and became less prominent after the immersion indicating that the distribution of silent pauses in the speech of beginner and intermediate learners became more similar. In this study, a silence of over 0.1 s was considered a pause. It is possible that the intermediate and advanced learners produced more pauses, but they were much shorter than those of beginners. The analysis of filled pauses, also known as hesitation markers, in the speech samples across proficiency levels was inconclusive. While the beginner students used fewer filled pauses in their speech, both intermediate and advanced learners exhibited more filled pauses, as shown in **Table 5**.

Figure 2 below offers a visual comparison of the changes in the production of the two pause types by students from the three proficiency levels. It is noteworthy that silent pauses seem to be overall more frequent than filled pauses across all three groups in both pre- and post-tests.

The LMEM with silent pauses as a dependent variable was significant, F (5,150) = 4.18, $p < 0.01$. The model also uncovered a significant interaction between the fixed effects, F (2,75) = 4.218, p = 0.018. However, the examination of the results of post hoc Tukey HSD tests indicated that none of the groups displayed significant developmental changes that resulted from immersion, and the significant results in the fixed effects in the model were merely caused by the differences in the number of silent pauses between the levels in the pre- and post-tests. The model was able to explain a total of 61% (conditional $R^2 = 0.614$) of the variance in the number of silent pauses produced by the speakers with the fixed effects accounting for 8.5% of the total difference (marginal $R^2 = 0.085$). The results further indicated that both of the random factors explained almost 26% of the variance each.

The LMEM with filled pauses was statistically significant overall, F (5,150) = 3.059, $p < 0.01$, as well as the interaction between proficiency and the time of testing (pre/post), F (2,75) = 9.305, $p < 0.01$. The post hoc Tukey HSD tests indicated that only the beginner learners significantly improved their fluency by producing fewer filled pauses in the post-test, t = 3.122, p = 0.029, d = 0.77 (medium effect size). The model was able to explain 39% of the variance in the filled pauses across groups (conditional $R^2 = 0.39$) with the fixed factors of proficiency and immersion contributing to almost 9% of the variation (marginal $R^2 = 0.089$). The majority of the explained random variance in the data was accounted for by participant factors.

The results of the three LMEMs presented above focused on the fluency features of the speech samples, namely, syllables per second, number of silent pauses, and number of filled pauses. Based on our analysis, only advanced learners made substantial fluency progress, as indicated by significantly faster speech rate in the post-immersion test. Another significant change that was observed in the analysis was a decreased number of filled pauses in the speech samples produced by the beginner learners.

Space

In order to examine the change in prominence patterns in participants' speech after ESL immersion, another mixed effects model was built with space as the dependent variable. The boxplots in **Figure 3** show that all three groups used fewer prominent words in their speech in the post-test. In particular, the change of space in the intermediate group was particularly prominent followed by the advanced learners and beginners. The

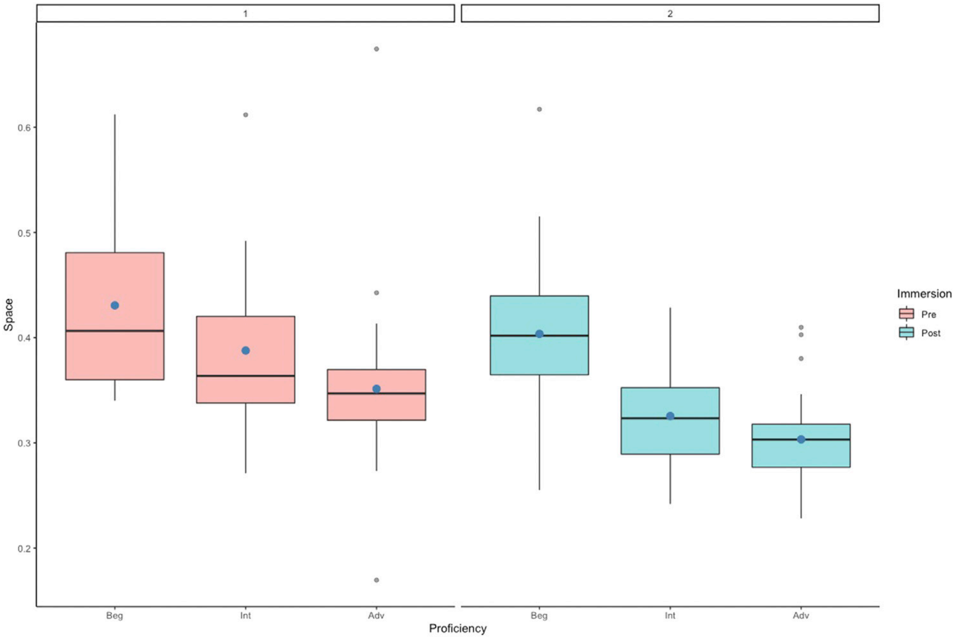


FIGURE 3 | Summary of results for space per proficiency group before and after immersion.

TABLE 6 | Summary of stress changes across proficiency levels pre- and post-immersion.

	Space											
	Beginner				**Intermediate**				**Advanced**			
	M	**SD**	**95% CI**		**M**	**SD**	**95% CI**		**M**	**SD**	**95% CI**	
			U	**L**			**U**	**L**			**U**	**L**
Pre-immersion	0.43	0.08	0.40	0.46	0.39	0.07	0.36	0.42	0.35	0.08	0.32	0.38
Post-immersion	0.40	0.07	0.37	0.43	0.32	0.05	0.30	0.34	0.30	0.04	0.28	0.32

CI = confidence interval, U = upper 95% CI, L = lower 95% CI.

figure also reveals that the intermediate and advanced groups contained some outliers who stressed as few as 10% of words in a run or as many as 70% in the pre-test. However, the outliers were grouped closer to the average in the post-test (in the advanced group) or disappeared completely (in the intermediate group). The descriptive results for space given in **Table 6** further exemplify that the stress changed across proficiency levels with more proficient learners stressing fewer words in a run and therefore improving their prominence.

Overall, the LMEM was significant, F (5,150) = 12.86, $p < 0.01$ as well as the interaction between the fixed effects, F (2,75) = 17.26, $p < 0.01$ suggesting that groups improved their prominence as a result of immersion. More specifically, according to the results of Tukey HSD tests, both intermediate and advanced groups significantly reduced their number of prominent words per run with a large and medium effect of immersion ($t = 3.858$, $p < 0.01$, $d = 1.15$ and $t = 3.148$, $p = 0.027$, $d = 0.79$ for intermediate and advanced groups, respectively). The change in space of the beginner group was not significant. The model accounted for nearly 52% of the variance (conditional $R^2 = 0.52$). Over half of the variance was due to the fixed factors, as indicated by a large marginal effect size ($R^2 = 0.29$). Most of the random factor variance was explained by the participant factors.

Pitch Range

The last linear mixed effects model in the data analysis involved the speakers' pitch range as the dependent variable. **Table 7** shows that the participant data in the three proficiency levels followed similar trends before and after the immersion with the advanced group demonstrating the widest range in contrast to the other two groups. It is noteworthy, however, that their pitch range was narrower in the post-test compared to the pre-test for this level. Similarly, in the post-test, the intermediate group also showed narrower while the beginners' pitch range was wider. The boxplots in **Figure 4** additionally illustrate that some of the speakers displayed pitch range as wide as almost 250 Hz in the advanced group.

The fitted linear mixed effects model was significant according to the summary statistic, F (5,150) = 2.324. The results of the linear mixed effects model detected a significant effect of

TABLE 7 | Summary of pitch range changes across proficiency levels pre- and post-immersion.

	Pitch Range											
	Beginner				Intermediate				Advanced			
	M	SD	95% CI		*M*	SD	95% CI		M	SD	95% CI	
			U	L			U	L			U	L
Pre-immersion	74.7	38.6	59.6	89.8	82.2	39.9	66.6	97.8	104	50.4	84.2	124
Post-immersion	76.5	38	61.6	91.4	80.5	36.6	66.2	94.8	99	41.8	86.2	115

CI = confidence interval, U = upper 95% CI, L = lower 95% CI.

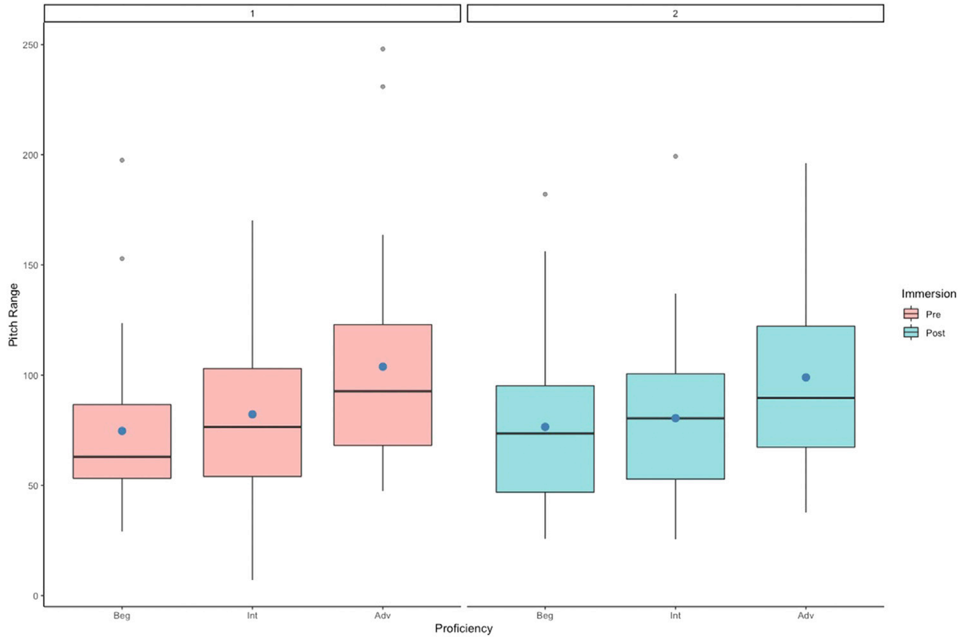


FIGURE 4 | Summary of results for pitch range per proficiency group before and after immersion.

proficiency F (2,75) = 3.54, $p = 0.03$ but not immersion F (1,75) = 0.05, $p = 0.82$ on pitch range. The interaction between the two effects was not significant. That is, the differences in the pitch range observed in the data occurred solely due to participants' proficiency and not as a result of ESL immersion. Overall, the model explained 61.5% (conditional $R^2 = 0.615$) of the variance in pitch range with the two fixed factors accounting for 7% (marginal $R^2 = 0.069$) of the variance and the remaining 54.5% falling mostly under the random effect of participants.

DISCUSSION AND CONCLUSION

The present study sought to offer a comprehensive investigation of the effects of a 15-week ESL immersion on learners' segmental and suprasegmental pronunciation features. Specifically, the study attempted to find out whether ESL immersion experience may play a role in reducing functional load divergences as well as making the participants' speech more fluent and intelligible. The results of the study revealed that immersion had no significant effect on segmental deviations across proficiency levels. In fact, both immersion and proficiency were able to explain only 1–2% of the difference in learners' production of high and low FL substitutions over time. Previous studies (e.g., Kang and Moran, 2014) focused on the difference in FL vowel and consonant substitutions across proficiency levels and found significant differences. In the current study, however, the focus was not on the differences between the levels but on whether or not FL-based segmental features change over time at each proficiency level; nevertheless, significant changes did not emerge.

In terms of changes in pronunciation at the suprasegmental level, the overall findings suggest that there are clear improvements in some of the suprasegmental aspects of L2 speech. In particular, the speaking rate of the advanced group increased significantly over the period of 15 weeks. It is widely known that speaking rate plays a crucial role in perceptions of L2 speech comprehensibility and accentedness. Indeed, there exists a curvilinear relationship between speech rate and listener ratings with speech that is too slow or too fast being rated less comprehensible and more accented (Munro and Derwing, 2001; Trofimovich and Baker, 2006). Although none of the groups in our study reached the optimal rate for accentedness and comprehensibility (4.76 and 4.23 syllables per second, respectively) suggested by Munro and Derwing (2001), the advanced learners made the most progress toward this goal. Another fluency feature that showed development in the study was filled pauses. In particular, beginner learners employed significantly fewer filled pauses in their speech after immersion. The finding supports recent research that showed that filled pauses can improve over time (e.g., Kang et al., 2021 in press). Furthermore, there was an improvement observed in intermediate and advanced learners' speech regarding their stress pattern (i.e., proportion of prominent words to the total number of words); that is, the number of prominent words significantly decreased over time for both of these groups. This decrease is especially beneficial for perceived accentedness of L2 speech. That is, the less a speaker stresses syllables in a sentence, the less accented they are found by the listeners (Kang, 2010). In contrast, the prosodic stress patterns of beginners did not improve. This finding lends potential support to the idea that the amount of L2 experience may influence the production of appropriate stress, as noted by Trofimovich and Baker (2006). It also is similar to Kang et al's., 2021 (in press) study where EFL learners improved their stress patterns most noticeably after 12 weeks of study in test preparation courses.

The two other suprasegmental features measured in our study (number of silent pauses and pitch range) did not exhibit substantial improvements after immersion. The overall pause structure of the participants' speech did not change with time for the three proficiency levels. The pitch range also stayed mostly the same across the groups. This result may indicate that the immersion experience does not affect all the suprasegmental features of speech in the same way and that some of these features, such as speaking rate and the distribution of silent pauses, may be more prone to improvement than others when students are immersed in an ESL environment. It is also possible that some suprasegmental features require more time and explicit instruction to develop noticeably (Levis, 2005). Taken together, the findings offer support to the idea that suprasegmental changes in L2 pronunciation may take less time than improvements in segmental features (Flege et al., 1997; Baker et al., 2001; Kang et al., 2021 in press).

An interesting finding that emerged as a result of the present study was the variance in the production of segmental and suprasegmental features that could be explained by participant factors in general and their L1 background. As a matter of fact, the possible individual differences among the participants accounted for the majority of the variance in the analyses, for wxample, over 50% in case of pitch range and 26% in the distribution of the silent pauses. The participants' L1 was not as pervasive of a predictor, although it did explain the other 26% of the variation in the use of silent pauses. These findings point at several inferences. First, the fact that the speakers' L1 explained so much of the variation in the silent pause distribution may be a sign of L1 transfer. Native speakers of Arabic and Mandarin may employ pauses differently in their first language and these patterns may be transferred from the participants' native language to English (e.g., Ortega-Llebaria and Colantoni, 2014). More importantly, however, the findings reinforce the influence of learners' individual differences in the acquisition of L2 speech. Research has repeatedly shown that both external factors such as age of acquisition and L1–L2 distance as well as the internal factors of motivation, attitude, and metalinguistic awareness are strongly connected to the learners' success in acquiring L2 pronunciation skills (Baker and Trofimovich, 2006; Derwing and Munro, 2013; Saito et al., 2020). Moreover, other community-based factors could potentially play a role in the development of the speakers' L2 pronunciation (Derwing and Munro, 2013). While the current findings cannot account for the specific outside activities and opportunities with the out-of-classroom community that might have affected the learners' pronunciation improvement, they do provide indirect support to the role of such activities for pronunciation development during immersion.

Taken together, the results in the present study yield evidence to three points related to ESL instruction, particularly in the context of immersion. First, explicit instruction is needed to improve L2 learners' pronunciation, especially on the segmental level. The participants in the study were students in an Intensive English Program that did not include a pronunciation class in its curriculum. The learners were enrolled in a listening and speaking class, which did not include targeted pronunciation activities beyond what was offered in the textbook. It has been previously shown that only in combination explicit pronunciation instruction can immersion be beneficial for learners (e.g., Lord, 2010). Moreover, the lack of a separate pronunciation class presents a challenge for learners' pronunciation development, especially since L2 textbooks are inconsistent in covering pronunciation (e.g., Derwing et al., 2012). It is not surprising that the only significant improvements demonstrated by the learners in this study were in fluency and sentence stress since L2 textbooks often weigh heavier toward suprasegmentals.

Second, comparing our results to previous research on ESL immersion, it appears that the duration of the immersion is another factor determining its effectiveness. It may be the case that the 15 weeks that the learners spent in the ESL environment in the present study was not enough to result in significant pronunciation improvement. For example, Trofimovich and Baker (2006) found that learners who spent 3 months abroad were significantly less fluent than those who were abroad for 3 years or more. Specifically, the authors found that the speakers' stress timing was related to the amount of L2 experience. This implies that the longer learners spend in the immersion context, the more prominent their pronunciation development is. On the other hand, Trofimovich and Baker also observed that the learners' production of L2 suprasegmentals was strongly correlated with their perceived accentedness no matter the duration of immersion. Since suprasegmentals present a learning challenge for the learners despite the length of their immersion experience, the need for explicit pronunciation instruction is reiterated.

Finally, the results presented here provide additional evidence to the developmental threshold hypothesis (Collentine, 2009). Each level in our study significantly improved their pronunciation by the end of their ESL immersion at least on some of the suprasegmental features, although this improvement was more noticeable among intermediate and advanced learners, as evidenced by the larger Cohen's d effect sizes. Importantly, the learners in the beginner group in our study were not "true" beginners. They were able to read the speaking prompt in the pre- and post-test and produced short but cohesive monologues in response to it. Therefore, it seems that the results of this study reinforce the idea that immersing students into the target language environment starts being effective only when the L2 learners reach at least basic literacy (Segalowitz and Freed, 2004) and that higher-level learners (intermediate and above) might improve their pronunciation more readily.

Despite an attempt in the present study to provide a comprehensive picture of pronunciation enhancement in the ESL immersion context, there are many questions that are left unanswered. For example, including additional segmental and/or suprasegmental features, such as tone choices, into the analysis could give a more fine-grained representation of specific changes in L2 English pronunciation. Another potential avenue of research could include the examination of interactive discourse (e.g., dialogues) in contrast to the monologues that were used in this study. It is oftentimes the case that suprasegmental features of conversations are different from monologic speech; thus, an investigation of changes in suprasegmentals employed by L2 learners in interactive discourse may shed additional light on prosodic development of L2 speech. Additionally, some individual differences can be further investigated through a qualitative approach as some of the speech properties (e.g., fluency features) can be idiosyncratic patterns instead of L2 proficiency or learning progression. In this research, we did not focus on the differences in segmental and suprasegmental deviations across learners' L1s. A potential qualitative study could investigate the effects of learners' L1 background on the types of pronunciation deviations produced by the learners. Finally, a word of caution needs be included with regard to the sampling procedures in the study. The number of speech samples for analyses was larger in case of suprasegmental analyses compared to segmental analysis. While linear mixed effects modeling performs equally well with smaller samples, we cannot exclude the slight possibility that the observed differences in suprasegmental features could have occurred due to the larger sample size. Moreover, the participants in the present study were recruited through convenience sampling. Although this type of sampling is typical for immersion studies, it would be beneficial for the domain if future research engaged in exploring pronunciation development in ESL immersion through random sampling.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Northern Arizona University IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

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

REFERENCES

Audacity Team (2020). *Audacity(R): free audio editor and recorder [computer application]*. Available at: https://audacityteam.org/ (Accessed September 10, 2020).

Avello, P. (2010). "The effect of study abroad on Catalan/Spanish bilinguals' production of English/i-ɪ/&/æ-ʌ/contrasting pairs," in 6th International conference on language acquisition-CIAL, September 8–10, 2010 Spain, Europe: Universitat de Barcelona.

Baker, W., Trofimovich, P., Mack, M., and Flege, J. E. (2001). "The effect of perceived phonetic similarity on non-native sound learning by children and adults," in Proceedings of the annual Boston University conference in language development. Editors B. Skarabela, S. Fish, and A. H-J. Do (Somerville, MA: Cascadilla Press), 36–47.

Baker, W., and Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: the role of individual differences. *Int. Rev. Appl. Linguistics Lang. Teach.* 44 (3), 231–250. doi:10.1515/iral.2006.010

Boersma, P., and Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Available at: http://www.praat.org/ (Accessed October 13, 2020).

Brecht, R., Davidson, D., and Ginsberg, R. (1995). "Predictors of foreign language gain during study abroad," in *Second Language Acquisition in a Study Abroad Context*. Editor B. Freed (Amsterdam, Netherlands: John Benjamins), 37–66.

Brown, A. (1991). "Functional load and the teaching of pronunciation," in *Teaching English Pronunciation: a Book of Readings*. Editor A. Brown (Abingdon, United Kingdom: Routledge), 221–224.

Catford, J. C. (1987). "Phonetics and the teaching of pronunciation: a systemic description of English phonology," in *Current Perspectives on Pronunciation: Practices Anchored in Theory*. Editor J. Morley (Alexandria, VI: TESOL), 87–100.

Collentine, J. G. (2009). "Study abroad research: findings, implications, and future directions," in *The handbook of Language Teaching*. Editors M. H. Long and C. J. Catherine (Hoboken, NJ: Wiley), 218–233.

Cumming, G. (2009). Inference by eye: reading the overlap of independent confidence intervals. *Stat. Med.* 28 (2), 205–220. doi:10.1002/sim.3471

Davidson, D. E. (2010). Study abroad: when, how long, and with what results? New data from the Russian front. *Foreign Lang. Ann.* 43 (1), 6–26. doi:10.1111/j.1944-9720.2010.01057.x

Derwing, T. M., Diepenbroek, L. G., and Foote, J. A. (2012). How well do general-skills ESL textbooks address pronunciation?. *TESL Can. J.* 30, 22–44. doi:10.18806/tesl.v30i1.1124

Derwing, T. M., and Munro, M. J. (2015). *Pronunciation fundamentals: evidence-based perspectives for L2 teaching and research*. Amsterdam, Netherlands: John Benjamins.

Derwing, T. M., and Munro, M. J. (2005). Second language accent and pronunciation teaching: a research-based approach. *TESOL Q.* 39 (3), 379–397. doi:10.2307/3588486

Derwing, T. M., and Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: a 7-year study. *Lang. Learn.* 63 (2), 163–185. doi:10.1111/lang.12000

Derwing, T. M., Munro, M. J., and Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Appl. Linguistics* 29 (3), 359–380.

Derwing, T. M., Munro, M. J., Thomson, R. I., and Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Stud. Second Lang. Acquis* 31 (4), 533–557. doi:10.1017/s0272263109990015

Derwing, T. M., and Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Appl. Lang. Learn.* 13 (1), 1–17.

Derwing, T. M., Rossiter, M. J., Munro, M. J., and Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Lang. Learn.* 54 (4), 655–679.

Derwing, T. M., Thomson, R. I., and Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System* 34 (2), 183–193. doi:10.1016/j.system.2006.01.005

Fayer, J. M., and Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Lang. Learn.* 37 (3), 313–326. doi:10.1111/j.1467-1770.1987.tb00573.x

Field, J. (2005). Intelligibility and the listener: the role of lexical stress. *TESOL Q.* 39 (3), 399–423. doi:10.2307/3588487

Flege, J. E., Bohn, O.-S., and Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *J. Phonetics* 25, 437–470. doi:10.1006/jpho.1997.0052

Freed, B. F., Segalowitz, N., and Dewey, D. P. (2004). Context of learning and second language fluency in French: comparing regular classroom, study abroad, and intensive domestic immersion programs. *Stud. Second Lang. Acquis.* 26 (2), 275–301. doi:10.1017/s0272263104262064

Højen, A. (2019). "Improvement in young adults' second-language pronunciation after short-term immersion," in *A Sound Approach to Language Matters*. Editors Jespersen, A. B., Hejná, M., and Sørensen, M. H. AU Library Scholarly Publishing Services, 543–560.

Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *Can. Mod. Lang. Rev.* 64 (4), 555–580. doi:10.3138/cmlr.64.4.555

Isaacs, T., and Trofimovich, P. (2012). Deconstructing comprehensibility. *Stud. Second Lang. Acquis* 34 (3), 475–505. doi:10.1017/s0272263112000150

Iwashita, N., Brown, A., McNamara, T., and O'Hagan, S. (2008). Assessed levels of second languagespeaking proficiency: how distinct? *Appl. Linguistics* 29, 24–49.

Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *IRAL: Int. Rev. Appl. Linguist. Lang. Teach.* 28 (2), 99.

Kang, O. (2012). Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Lang. Assessm. Quart.* 9 (3), 249–269.

Kang, O., Ahn, H., Yaw, K., and Chung, S.-Y. (2021). Investigation of relationship among learner background, linguistic progression, and score gain on IELTS. *IELTS Research Reports Online Series, No. 1*. British Council, Cambridge Assessment English and IDP: IELTS Australia. Available at: https://www.ielts.org/teaching-and-research/research-reports.

Kang, O., and Ghanem, R. (2016). Learners' self-perception of target language study in overseas immersion. *Jltr* 7 (5), 819–828. doi:10.17507/jltr.0705.01

Kang, O., and Moran, M. (2014). Functional loads of pronunciation features in nonnative speakers' oral assessment. *Tesol Q.* 48 (1), 176–187. doi:10.1002/tesq.152

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* 38, 301–315.

Kang, O., Rubin, D., and Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Mod. Lang. J.* 94 (4), 554–566. doi:10.1111/j.1540-4781.2010.01091.x

Kang, O., Thomson, R. I., and Moran, M. (2020). Which features of accent affect understanding? exploring the intelligibility threshold of diverse accent varieties. *Appl. Linguistics* 41 (4), 453–480. doi:10.1093/applin/amy053

Lafford, B. (2004). The effect of context of learning on the use of communication strategies by learners of Spanish as a second language. *Stud. Second Lang. Acquis.*, 26, 201–226.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Q.* 39 (3), 369–377. doi:10.2307/3588485

Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign Lang. Ann.* 43 (3), 488–503. doi:10.1111/j.1944-9720.2010.01094.x

MacIntyre, P. D. (2007). Willingness to communicate in the second language: understanding the decision to speak as a volitional process. *Mod. Lang. J.* 91 (4), 564–576. doi:10.1111/j.1540-4781.2007.00623.x

Martinsen, Rob A. (2008). Short-term study abroad: predicting changes in oral skills. *Foreign Lang. Ann.* 43 (3), 504–530.

Muñoz, C., and Llanes, À. (2014). Study abroad and changes in degree of foreign accent in children and adults. *Mod. Lang. J.* 98 (1), 432–449. doi:10.1111/j.1540-4781.2014.12059.x

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 (4), 451–468. doi:10.1017/s0272263101004016

Munro, M. J., and Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: an exploratory study. *System* 34 (4), 520–531. doi:10.1016/j.system.2006.09.004

Ortega-Llebaria, M., and Colantoni, L. (2014). L2 English intonation. *Stud. Second Lang. Acquis.* 36 (2), 331–353. doi:10.1017/s0272263114000011

Pérez-Vidal, C., Juan-Garau, M., and Mora, J. C. (2011). "The effects of formal instruction and study abroad contexts on foreign language development: the SALA project," in *Implicit and Explicit Language Learning: Conditions, Processes and Knowledge in SLA and Bilingualism*. Editors R. Leow and C. Sanz (Washington, DC: Georgetown University Press), 115–138.

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Q.* 35, 233–255. doi:10.2307/3587647

Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English Specif. Purp.* 23 (1), 19–43. doi:10.1016/s0889-4906(03)00020-6

Plonsky, L., and Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Lang. Learn.* 64 (4), 878–912. doi:10.1111/lang.12079

R Core Team (2020). *R: a Language and environment for statistical computing*. Vienna, Austria. Available at: https://www.R-project.org/ (Accessed August 25, 2020).

Saito, K., Macmillan, K., Mai, T., Suzukida, Y., Sun, H., Magne, V., et al. (2020). Developing, analyzing and sharing multivariate datasets: individual differences in L2 learning revisited. *Ann. Rev. Appl. Linguist.* 40, 9–25. doi:10.1017/s0267190520000045

Segalowitz, N., and Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition. *Stud. Second Lang. Acquis.* 26, 173–199. doi:10.1017/s0272263104262027

Stevens, J. J. (2002). The acquisition of L2 Spanish pronunciation in a study abroad context. [Unpublished Doctoral Dissertation]. Los Angeles (CA): University of Southern California.

Suzukida, Y., and Saito, K. (2019). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Lang. Teach. Res.* doi:10.1177/1362168819858246

Tavakoli, P., and Skehan, P. (2005). *9. Planning and task performance in a second language*. Amsterdam, Netherlands: John Benjamins.

Towell, R. (2002). Relative degrees of fluency: a comparative case study of advanced learners of French. *Int. Rev. Appl. Linguist.* 40 (2), 117–150.

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. Acquis.* 28 (1), 1–30. doi:10.1017/s0272263106060013

Wennerstrom, A. (2000). "The role of intonation in second language fluency," in *Perspectives on Fluency*. Editor H. Riggenbach (Ann Arbor, MI: University of Michigan Press), 102–127.

Conflict of Interest: 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.

frontiers in Communication

GENERAL COMMENTARY
published: 22 November 2021
doi: 10.3389/fcomm.2021.788177

Commentary: Evidence-Based Principles for Pronunciation Teaching & ESL Immersion and Pronunciation Development

*Sinem Sonsaat Hegelheimer**

Department of English, Iowa State University, Ames, IA, United States

Keywords: pronunciation teaching, functional load, immersion context, suprasegmentals, Spanish, English

A Commentary on

Evidence-Based Principles for Pronunciation Teaching & ESL Immersion and Pronunciation Development
by Colantoni L., Escudero P., Marrero-Aguiar V., and Steele J. (2021). Front. Commun. 6:639889. doi: 10.3389/fcomm.2021.639889

OPEN ACCESS

Edited by:
Antonio Benítez-Burraco, Sevilla University, Spain

Reviewed by:
Andrew Sewell, Lingnan University, Hong Kong SAR, China

****Correspondence:***
Sinem Sonsaat Hegelheimer
sonsaat@iastate.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *01 October 2021*
Accepted: *25 October 2021*
Published: *22 November 2021*

Citation:
Sonsaat Hegelheimer S (2021) Commentary: Evidence-Based Principles for Pronunciation Teaching & ESL Immersion and Pronunciation Development. Front. Commun. 6:788177. doi: 10.3389/fcomm.2021.788177

"EVIDENCE-BASED DESIGN PRINCIPLES FOR SPANISH PRONUNCIATION TEACHING" (COLANTONI, ESCUDERO, MARRERO-AGUIAR, AND STEELE)

There is a plethora of research exploring how English pronunciation should be taught or what features should be prioritized in teaching to help improve learners' intelligibility. However, there is not much published on the same issue in other languages. Because each language has its own phonetic inventory, phonological idiosyncrasies and variations, research on issues related to teaching priorities in English pronunciation may be helpful but not fully applicable to other languages. This conceptual analysis on five research-informed design principles discussing the priorities and considerations for Spanish pronunciation teaching is important and informative for researchers, practitioners, and materials developers.

Of the five principles, "a focus on features with high functional load" is especially valuable. In this section, the authors suggest a frequency-based method to reach conclusions about which sounds are more important to address at different levels of L2 Spanish learning. For teachers who do not feel very confident about what sounds to prioritize in pronunciation instruction, the authors provide a list of the most frequent sounds in Spanish and guide readers about the importance of minimal pairs in determining the importance of sound contrasts. In a further study, the authors could potentially create a rank ordering of Spanish sound pairs (for the variety of their choice) that is likely to cause intelligibility problems by following a method such as the one in Brown (1988). The authors criticize the functional load principle for not addressing suprasegmental features and suggest ways of incorporating lexical stress and sentence-type intonation into the functional load principle. Functional load, as we usually understand it, is built primarily on numbers of minimal pairs for a given contrast. But this feature is not easily applicable to suprasegmentals. Perhaps we need a different measure altogether. Such a measure does exist in the form of the information-theoretic approach to functional load, outlined by Hockett (1967) and elaborated by Surendran and Niyogi (2003); see also Sewell, 2021). The work of Surendran and Niyogi (2003) focused on the relative contributions of vowels, consonants, and lexical stress etc. for English, Dutch, German, and

Mandarin, but there is still need for research focusing on Spanish. This might be a potential future exploration both for the authors of this paper and other researchers.

Another helpful section of this paper is the one on 'targeting features and segments shared by the majority of the varieties of the target language since it brings attention to the commonly shared features in most Spanish varieties. The authors propose that commonly shared sounds in most Spanish varieties should be prioritized in pronunciation instruction to support mutual intelligibility and inform about the most appropriate developmental stage (i.e., proficiency level) to address particular dialectal features. In this section, the authors increase the visibility of research done in Spanish varieties by not only reviewing the studies published in English but also the ones in Spanish.

One of the contributions of this paper is to highlight the role of prosody in Spanish, that is, marking inflectional features through lexical stress and sentence structure through intonation (i.e., declaratives versus questions). Having called attention to the distinct role of prosody in Spanish, the paper might be strengthened by reviewing the studies addressing prosody in Spanish. Most studies cited in the current paper focus on prosody in English.

In conclusion, the conceptual analysis in this paper provides a foundation for the further development of Spanish pronunciation teaching. Particularly important is its consideration of the importance of perceptual training, the role of prosodic components in Spanish, the use of contextualized activities, the importance of extending research on functional load to Spanish, and the benefits of teaching pronunciation features that are shared across Spanish varieties.

"THE EFFECTS OF ESL IMMERSION AND PROFICIENCY ON LEARNERS' PRONUNCIATION DEVELOPMENT" (KOSTROMITINA AND KANG)

The development of segmental and suprasegmental pronunciation features in L2 English is a commonly studied topic in pronunciation instruction studies in laboratory or classroom settings (for example, see Lee et al., 2015). In longitudinal studies, however, it is mostly global measures of speech such as comprehensibility, accentedness, and fluency that are explored. It is less common to see studies in which the development of multiple segmental and suprasegmental features is reported over an extended period of time. This exploratory study differs by investigating ESL speakers' development of segmentals and suprasegmentals (i.e., fluency, prominence, and intonation) in an immersion context without any classroom instruction.

The study has a number of features that enable it to capture ESL speakers' pronunciation development in a more holistic way. First, it explores pronunciation issues using elicited free speech data (extended samples of monologic speech), which is challenging while looking at segmentals since speakers may not produce sufficient tokens for some sounds. Second, rather than focusing on a selected set of sounds, the authors map out the developmental trajectory of all segmentals by employing a functional load-based analysis. Third, fluency, prominence, and intonation are analyzed through acoustic measures rather than listener-based judgements to give a more fine-detailed picture of learners' development. Lastly, this study compares the development of beginner, intermediate and advanced speakers to investigate the relationship between proficiency level and developmental patterns, providing information about whether there is an optimal time to be immersed into the second language environment.

The study reports no significant improvement for segmental features across proficiency levels. In fact, the proficiency levels of the speakers and the immersion experience could explain only about 2% of the change in L2 speakers' speech and the authors report that most of the variance was explained by individual differences among the participants. For suprasegmentals, the study reports significant effects of immersion experience with medium to large effect sizes only for some features (fluency and prominence) but not across all proficiency levels. As for intonation, there was no significant effect of immersion but only of proficiency level. Indeed, for all features in the study, individual differences were influential and the study as a whole points to the importance of both external factors (such as age of onset) and internal factors (such as motivation) in explaining the development of L2 pronunciation. The authors also reported substantial influence of random factors (i.e., participants and their L1 background) based on their Linear Mixed Effects Models analyses. This study is important in terms of showing the importance of individual differences since "participants" was the random factor accounting for so much difference in speakers' performances. The study is also important in emphasizing the need for explicit training to support the improvement of productive speech and pronunciation skills of L2 speakers. Lastly, the results of this exploratory study bring up the question of the optimal proficiency level that is required for L2 speakers to benefit from an immersion experience. This study reinforces the findings of previous research showing the importance of both segmentals and suprasegmentals (Pickering, 2001; Hahn, 2004; Field, 2005; Kang, 2010; Kang et al. 2010; Kang and Moran, 2014) for L2 pronunciation performance and how they may follow different developmental trajectories for speakers who started their immersion experience at different levels (Collentine, 2009; Davidson, 2010; Kang and Ghanem, 2016). It also indicates that some pronunciation features may be quite challenging to acquire, at least in a short time frame.

The way this study lies at the crossroads of multiple research topics (i.e., functional load, segmentals and suprasegmentals, immersion contexts) reflects the ongoing research agenda of Okim Kang and her collaborators, whose work on functional load and the role of suprasegmentals in L2 pronunciation has pushed the field forward in connecting acoustic measurements and listener-based ratings (Kang et al., 2010; Kang, 2012; Kang and Moran, 2014; Kang et al., 2020).

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Brown, A. (1988). Functional Load and the teaching of pronunciation. *TESOL Q.* 22 (4), 593–606. doi:10.2307/3587258

Collentine, J. G. (2009). "Study abroad research: Findings, implications, and future directions," in *The handbook of language teaching*. Editors M. H. Long and C. J. Catherine (UK: Wiley), 218–233.

Davidson, D. E. (2010). Study abroad: When, how Long, and with what results? New data from the Russian front. *Foreign Lang. Ann.* 43 (1), 6–26.

Field, J. (2005). Intelligibility and the Listener: The role of Lexical stress. *TESOL Q.* 39 (3), 399–423.

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Q.* 38 (2), 201–223. doi:10.2307/3588378

Hockett, C. (1967). The quantification of functional Load. *Word* 23, 320–339. doi:10.1080/00437956.1967.11435484

Kang, O., and Ghanem, R. (2016). Learners' self-perception of target Language study in overseas immersion. *J. Lang. Teach. Res.* 7 (5), 819–828.

Kang, O. (2012). Impact of Rater Characteristics and Prosodic Features of Speaker Accentedness on Ratings of International Teaching Assistants' Oral Performance. *Lang. Assess. Q.* 9 (3), 249–269. doi:10.1080/15434303.2011.642631

Kang, O., and Moran, M. (2014). Functional Loads of pronunciation features in nonnative speakers' oral assessment. *TESOL Q.* 48 (1), 176–187.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* 38 (2), 301–315.

Kang, O., Rubin, D., and Pickering, L. (2010). Suprasegmental Measures of Accentedness and Judgments of Language Learner Proficiency in Oral English. *Mod. Lang. J.* 94 (4), 554–566. doi:10.1111/j.1540-4781.2010.01091.x

Kang, O., Thomson, R. I., and Moran, M. (2020). Which features of accent affect understanding? Exploring the intelligibility threshold of diverse accent varieties. *Appl. Linguistics* 41 (4), 453–480.

Lee, J., Jang, J., and Plonsky, L. (2015). The effectiveness of second Language pronunciation instruction: A meta-analysis. *Appl. Linguistics* 36 (3), 345–366. doi:10.1093/applin/amu040

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Q.* 35 (2), 233–255.

Sewell, A. (2021). Functional Load and the teaching-Learning relationship in L2 pronunciation. *Front. Commun.* 6, 1–6. doi:10.3389/fcomm.2021.627378

Surendran, D., and Niyogi, P. (2003). *Measuring the functional load of phonological contrastsTech. Rep. No. TR-2003-12)*. Chicago: Department of Computer Science, University of Chicago.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

MINI REVIEW
published: 24 May 2021
doi: 10.3389/fcomm.2021.627378

Functional Load and the Teaching-Learning Relationship in L2 Pronunciation

Andrew Sewell *

Department of English, Lingnan University, Hong Kong, China

Though frequent recourse has been made to the functional load (or FL) principle in establishing priorities for L2 pronunciation teaching, it remains an under-theorized and relatively under-utilized concept. This is despite the existence of empirical evidence pointing to correlations between the FL ranking of phonemic contrasts and a) the effect that the absence of particular contrasts has on the comprehensibility of speech, and b) their occurrence at different levels of proficiency. Previous studies have found that errors involving high FL sound contrasts are linked with educed comprehensibility, and have also found that high FL errors are less common in learners at higher proficiency levels. Taken together, these findings suggest that language learners tend to pay more attention to high FL contrasts and incorporate them into their repertoires more readily than low FL contrasts, possibly because the high FL contrasts are more salient in terms of contrastive potential and frequency of occurrence. The concept of FL therefore appears to be relevant in considering the relative ease (or difficulty) of learning and teaching particular features, and in understanding the relationship between learning and teaching. Frequent calls have been made for FL considerations to inform the setting of priorities in L2 pronunciation teaching, for example. In this mini-review I will explore and re-evaluate the concept of FL in terms of both theoretical formulation and empirical application, aiming to identify both its contributions and its limitations.

Keywords: L2 pronunciation, functional load, second language acquisition and development, L2 pronunciation teaching, L2 phonology

OPEN ACCESS

Edited by:
Sidarta Ribeiro,
Federal University of Rio Grande do Norte, Brazil

Reviewed by:
Mark McAndrews,
Concordia University, Canada
Yui Suzukida,
University College London, United Kingdom
Peter Avery,
York University, Canada

****Correspondence:***
Andrew Sewell
asewell@ln.edu.hk

Specialty section:
This article was submitted to Culture and Communication, a section of the journal Frontiers in Communication

Received: *09 November 2020*
Accepted: *10 May 2021*
Published: *24 May 2021*

Citation:
Sewell A (2021) Functional Load and the Teaching-Learning Relationship in L2 Pronunciation.
Front. Commun. 6:627378.
doi: 10.3389/fcomm.2021.627378

INTRODUCTION

The concept of functional load (hereafter, FL) is approaching its centenary. From its first mention in the discussions of the Prague School linguists (e.g., Jakobson, 1931), a line of influence can be traced through postwar structural linguistics (e.g., Martinet, 1952; Hockett, 1967) to the application of FL to language teaching (e.g., Catford, 1987; Brown, 1991). In recent years there has been a resurgence of interest in FL in the field of L2 pronunciation (e.g., Munro and Derwing, 2006; Suzukida and Saito, 2019) and assessment (e.g., Kang and Moran, 2014).

How can the enduring appeal of FL be explained? It appears to hold out the promise of a parsimonious explanation for various linguistic phenomena, ranging from diachronic sound change to the effects of different sound contrasts on the perceived comprehensibility of spoken language. But there is no agreed-upon definition of FL, and studies in the field of L2 pronunciation still rely on lists of minimal pairs drawn up in the pre-computer age. In this mini-review article I have two main objectives: firstly to critically examine the concept of FL itself, and secondly to review its deployment in research, aiming to identify both its contributions and its limitations. In doing so I will address the central concern of this special issue, namely the relationship between ease of acquisition and ease of teaching in L2 speech. I will argue that FL can inform our understanding of this relationship and help to answer the question of why it is that certain phenomena are more difficult to learn and teach. However, I also argue against the mechanistic application of FL and call instead for an increased awareness of what lies behind the concept and its measurements. Suitably reconceptualized, the FL concept can maintain its usefulness in an era of international communication and dynamic language practices.

THE ORIGINS AND NATURE OF FL

The durability of the FL concept reflects an enduring interest in the relationship between the structure of linguistic systems and the functional roles of their components. The appeal of FL as a way of measuring the functional or informational value of these components was, and still is, an "intuitively attractive idea" (Wedel et al., 2013a: 397). The first applications of FL to language teaching appeared in the work of Catford (1987) and Brown (1991), both of whom were concerned with identifying priorities for L2 pronunciation teaching. The ranked lists of English phonemic contrasts prepared by these scholars are still used in present-day studies (e.g., Derwing and Munro, 2006, which looked at the relationship between FL and the perceived accentedness and comprehensibility of L2 English speech).

The simplest definition of FL, and the one with which most linguists are familiar, is that of the amount of "work" performed by the phonemic contrasts found in a given language. The simplest measurement of FL—what Martinet (1955: 54) called the most "naïve" measurement - is that of the number of minimal pairs a particular contrast serves to differentiate. This was the basis of the lists prepared by Catford (1987). The lists of Brown (1991) adopt a more sophisticated approach, taking account of factors such as the relative frequency with which the constituents of minimal pairs occur and the number of pairs that have the same part of speech (such as *live/leave*). Nevertheless, despite the slightly different approach to measurement there is broad agreement between the two lists. Consonantal contrasts such as /l, r/ and vowel contrasts such as /ɔː, əʊ/ have a high FL in both. Indeed, in comparing the effects of using the Catford and Brown lists for their study, Munro and Derwing (2006) observed that were no conflicts between them.

THE APPLICATION OF FL IN L2 PRONUNCIATION RESEARCH

Despite its intuitive appeal, the complexity of FL turns out to be daunting. Beyond the core principle of "amount of contrastive work" and its measurement by minimal pair counts, there is no agreed-upon way to define or measure FL. This probably explains why the lists of Catford (1987) and Brown (1991) still serve as the go-to resource for researchers (e.g., Munro and Derwing, 2006; Kang and Moran, 2014; Suzukida and Saito, 2019). In this *Introduction* will briefly review these studies to illustrate the application of FL and begin to identify its contributions to L2 pronunciation research and its overall significance.

In their exploratory investigation, Munro and Derwing (2006) found preliminary confirmation of the "functional load hypothesis", namely that high FL errors (such as substituting /n/ for /l/) have a greater impact on listeners" perceptions of the accentedness and comprehensibility of L2 speech than low FL errors (p. 529). The study of Suzukida and Saito (2019) compared the effects of vowel and consonant substitutions with different FL values. It lent further support to the FL hypothesis by discovering that consonant substitutions were more detrimental to comprehensibility, and concluded that it was "only high FL consonant substitutions (e.g., mispronunciation of /l/ as /r/ or /v/ as /b/) that negatively impacted on native listeners" comprehensibility judgments" (p. 1). Taking a slightly different approach, Kang and Moran (2014) studied the patterning of high and low FL errors within speech samples taken from different levels of the Cambridge ESOL exam suite. The study found that the percentage of high FL errors was inversely related to proficiency level (p. 185), providing indirect evidence of a developmental trend in which learners progressively learn to avoid, or correct, high FL errors.

The studies of Munro and Derwing (2006) and Suzukida and Saito (2019) were both concerned with the dimensions of comprehensibility (i.e., perceived ease of understanding) and accentedness. I have elsewhere proposed that FL can also help to explain the findings of certain studies focusing on intelligibility (i.e., actual understanding in terms of word recognition; see Sewell, 2010; Sewell, 2017). The studies of Jenkins (2000) and Deterding (2013) were both concerned with international intelligibility among non-native speakers of English, and involved collecting corpora of misunderstandings. FL considerations can largely explain the hierarchies of error significance found in these studies, even though the researchers did not make explicit reference to FL. For example, Jenkins's proposed Lingua Franca Core (LFC) of intelligibility-preserving features is far more prescriptive for consonants than it is for vowels, echoing the findings of Suzukida and Saito (2019). The only vowel contrast given priority treatment in Jenkins's LFC is /ɪ, i/, which is a high FL contrast. Consonantal substitutions were also found to be more problematic in Deterding's study, and within this category the most problematic were the substitution of /n/ with /l/ and of /l/ with /r/—again, these are high FL contrasts in the Catford and Brown rankings.

FL therefore appears to be a promising explanatory factor in studies of the three dimensions of accentedness (Munro and Derwing, 2006), comprehensibility (Munro and Derwing, 2006; Suzukida and Saito, 2019) and intelligibility (Jenkins, 2000; Deterding, 2013). It may also be relevant in explaining the order in which sound contrasts are typically learned (Kang and Moran, 2014). However, both the concept and its application need to be placed on a firmer footing. The continuing use of the Catford and Brown lists is both reassuring and troubling. It is reassuring because they are mostly consistent with each other and are able to generate fairly consistent results when used in statistical analyses. It is troubling because the lists are over 30 years old, and because neither made their assumptions and procedures particularly clear (see Levis and Cortes, 2008: 200). The underlying theoretical basis of FL is therefore also unsatisfactory. Although minimal pair counts and rankings are necessary for statistical analysis, we must also be concerned what lies behind the concept: what do the statistics and rankings represent, and what do we mean by "communicative value"?

FL 2.0: A BROADER VIEW

Putting FL on a firmer theoretical footing should help to connect the various areas of interest in this special issue and address the central question of why it is that certain phenomena are more difficult to learn and teach. In an earlier work on the subject I suggested that a useful distinction could be made between narrow and broad senses of FL (Sewell, 2017). This was intended to widen the scope of FL beyond minimal pair counts. The broad sense of FL—also called the information-theoretic, entropy or global approach (Oh et al., 2015)—allows for FL to be measured and compared not only at the level of phoneme contrasts, but also with regard to individual phonemes and higher-level categories. Applying such an approach, Oh et al. (2015) demonstrated that consonants had a higher FL than vowels, not only in English but in all nine languages they studied. Taking a broader view not only extends the scope of FL beyond minimal pairs, but also increases its theoretical coherence by providing a psycholinguistic perspective and clarifying what it is that FL is actually measuring. For the purposes of this special issue I will illustrate this by explaining the relevance of FL across three interrelated temporal frames, those of interaction, learning and language change. Within and across these frames, FL measurements retain their essential character of indicating how, and where, the scope for variation in language use is constrained by the need to retain information value.

The temporal frame of interaction represents the here-and-now of communicative activity, and L2 pronunciation research approaches it *via* the concepts of intelligibility and comprehensibility, among others. In researching this temporal frame, FL provides an indication of which features are relied upon by participants to secure mutual understanding. Taking a longer-term view, the second temporal frame is that of learning. Participants can be seen as bringing their habitual ways of speaking and their implicit or explicit knowledge of language (e.g., phoneme and word frequencies, collocational patterns, and sound/spelling correspondences) to the frame of interaction. Of course, their habits and knowledge are the cumulative result of interaction, and interaction provides further opportunities for learning; the two frames are in fact interdependent. Learning is partly open-ended, but is also shaped by the demands of interaction. In the context of L1 learning, Bybee (2001) characterizes the process of phonological acquisition as one of increasing fluency and automation, but one which is "constrained by the need to retain information value" (p. 15). Also writing from a functionalist perspective, Croft (2000) notes that communities of language users deal with the "problem" of communication by converging on "a regular solution to a recurring co-ordination problem" (p. 97). The co-ordination problem of how to distinguish between similar words is solved by relying on phonemic contrasts, and FL can be seen as indicating the relative usefulness of particular contrasts in maintaining information value. The finding of Kang and Moran (2014), namely that high FL errors are less common at higher proficiency levels, provides indirect evidence of an L2 learning process that is also shaped by emerging knowledge of information value, among other factors.

It is important to consider how this L2 learning process may take place, with reference to FL. It appears that high FL errors are gradually eliminated, but that low FL errors often remain, perhaps as more or less permanent accent features. The gradual elimination of high FL errors may be a result of actual communication breakdown or other kinds of negative feedback. A related possibility is that high FL errors tend to be more salient, because they occur more frequently or because they involve contrasts that are relatively easy for users to distinguish, and are thus more likely to be the target of monitoring by self or others. Similarly, the persistence of low FL errors, noted by Kang and Moran (2014) even in high-proficiency samples, may be due to their relative insignificance in terms of triggering communication breakdown. If the absence of an L2 contrast does not lead to problems in the temporal frame of interaction, the contrast is less likely to be incorporated into learner's repertoires.

To a certain extent, then, FL lends support to naturalistic methods of language learning: if learners are exposed to sufficient input and have opportunities for meaningful interaction, they will automatically learn which features are most important (Krashen, 1981). However, and to move back into the temporal frame of interaction, an awareness of FL may help instructors to provide high-quality feedback in the form of awareness-raising activities. With same-L1 classes I have found it useful to play recordings of local speakers as a dictation exercise. The words that are difficult (or impossible) to transcribe are often found to contain high FL errors, which can then be brought to the learners' attention. Taking a broader view of FL, such intelligibility problems are often associated with phonological contexts (such as word-initial position) that enhance the information value of contrasts. This was visible in the intelligibility study of Deterding (2013), which required listeners from different L1 backgrounds to transcribe extracts of L2 English conversations. Substitutions in word-initial position were a prevalent cause of intelligibility problems; the substitution of [n] for /l/ (e.g., "noisy" pronounced as "loisy") was particularly problematic, as would be expected for a high FL contrast.

The "noisy/loisy" example raises the question of the relevance of minimal pairs in FL, and also starts to illuminate the psycholinguistic basis of the concept. It would not appear on traditional minimal pair lists, as "loisy" is not a currently-existing word. However, the issue at stake is not merely the confusability of minimal pairs—the activation of non-words such as "loisy" can also be distracting for the L2 listener, who will often be unsure as to whether it is a real word or not. Weber and Cutler (2004) contend that much of the difficulty of listening to and understanding L2 speech arises from the activation of "spurious competitor words," which can take the form of L2 minimal pair members (when "still" is heard for "steel"), near-matches (as when "belly" is heard for "balance") or non-words. The significance of the /r, l/ contrast for Japanese learners, for example, resides not only in the potential confusability of pairs such as "belly" and "berry" but also in the unwanted activation of words like "barrow" and "barren" when the target word is "balance" (Weber and Cutler, 2004: 3). The measurement, and perhaps more importantly the theorization, of FL needs to take account of this. It may be that measurements based on minimal pair counts are no more than indirect reflections of overall "information value."

In addition to interaction and learning, the third temporal frame relevant to FL is that of language change. Indeed, the original meaning of the "functional load hypothesis" was the question of whether "sound change is biased toward selective maintenance of those phonemes that contribute more to distinguishing existing lexical items in usage" (Wedel et al., 2013a: 396). Recent corpus-based studies have lent support to this version of the FL hypothesis (Wedel et al., 2013a; Wedel et al., 2013b). While it is possible to speculate on the likely future direction of phoneme merger (or equally, contrast preservation) based on FL considerations, the focus in this special issue is on the temporal frames of interaction and learning. The reason for considering all three frames is that this gives the concept of FL greater theoretical coherence—at least, as long as the assumptions of generative phonology are passed over in favor of models that have usage, rather than abstract systems, as their guiding principle. FL originated as a functionalist concept, and it continues to influence what Wedel et al. (2013a: 396) categorize as "VUE" (variationist, usage-based, evolutionary) models of language. According to Wedel et al., a central feature of such models is "the assumption of a causal chain linking properties of individual usage events to long-term change in the abstract, linguistic category system of a speech community" (2013a: 410; see also Bybee 2001; Bybee and Hopper, 2001).

The key characteristic of FL across all three frames is that it attempts to measure the prevailing information value of linguistic features. Language is inherently variable, and change is always in progress, but the need to retain information value provides a centripetal brake on these centrifugal forces. High FL, in other words, indicates that a feature or contrast is heavily relied upon to make distinctions of meaning. These features are, by and large, automatically prioritized by language learners and language users; from a longer-term perspective their information value is a consequence of usage, in turn influenced by technological affordances (such as writing) and societal trends (such as literacy).

By reconceptualizing FL as "FL 2.0" I am not arguing for the abandonment of minimal pair counts. It may turn out that these offer a shorthand approach to the complexities of FL. However, regardless of how FL is measured, there needs to be a greater awareness of what the concept and the counts represent. As Suzukida and Saito (2019) point out, there are more users of English as an L2 than there are native speakers. This introduces greater scope for variation and for alternative solutions to the co-ordination problems of communication. If FL and its measurement continue to be based on the narrow notion of minimal pairs (the lists of which may also be outdated), and if its indications are treated as language universals, its ability to inform research and language education will be compromised. Although research studies and teaching guides prefer to operate with ranked lists of features, the lists represent the effects of complex forces operating across different timescales.

CONCLUSION

What, then, does an expanded view of FL have to offer with regard to the central question of this issue, namely the relationship between ease of acquisition and ease of teaching in an L2? One aspect of the relationship between the two is illuminated by Levis"s observation that "certain features are not acquirable in the long run, no matter what we do to teach effectively or no matter how much effort learners put into learning" (2018: 213). Many of the accent features that are retained by advanced learners have a low FL (Kang and Moran, 2014) and may not need to be prioritized in either teaching or testing. An FL perspective also suggests, by implication, that the process of acquisition involves mastering features and contrasts whose FL is high, or to put it another way, learning to avoid high FL errors. The traditional contribution of FL has been to help predict what these features might be and to indicate targets for teaching and assessment, even though many of these features are likely to be difficult to acquire (e.g., /i/ and /ɪ/ for Cantonese speakers).

FL has several limitations, however. Paradoxically, the more the theoretical foundations of FL are buttressed by taking it out of the narrow realm of minimal pairs, the more it becomes apparent that it is one of the many factors that shape the contours of interaction, learning and language change. Behind the statistics, the underlying principle of FL is simply that information value plays an important role in determining the nature and scope of variation in human language, both at relatively shorter timescales (such as those of interaction and learning) and across longer timescales of language change. That it does not have a determining role is shown by the many phenomena that appear to run counter to the FL principle. For example, it has been observed that the absence of contrast between /i/ and /ɪ/ is a feature of many L2 English accents around the world (see, e.g., Lim, 2004, on Singapore English; Deterding et al. (2008) and Hung, 2000 on Hong Kong English). It is noticeable in the speech of advanced learners (i.e., it is associated with the temporal frames of interaction and learning), and may represent language change in progress, at a local level.[1] Yet this contrast is given a high FL ranking in both the Catford and Brown lists. Unless this is due to the relatively "narrow" measures of FL represented by these lists,

[1] The lack of data on variation within these varieties makes it difficult to draw firm conclusions. Sarmah et al. (2009) found that the contrast also existed in Thai English, but only for early-stage learners or "new speakers". There is historical evidence for the instability of this contrast, and Jones (2012) concludes that it probably did not exist in the phonology of eighteenth-century "Late Modern English" (p.828).

we must consider the possible reasons for this and other exceptions before making pedagogical recommendations.

The learners' L1 is a major reason for these exceptions. For example, the /i/ and /ɪ/ contrast is allophonic in Cantonese, and many other well-known and recurring "errors"—such as Japanese learners' difficulties with the /l, r/ contrast—are related to the L1, illustrating the psycholinguistic principle that "the hardest second-language contrasts to learn are those which are ignored in the native language because each of the contrasting sounds is a permissible token of a single native category" (Best, 1995, in Weber and Cutler, 2004: 2). It may simply be, therefore, that the FL_derived "information value" of these contrasts is outweighed by the difficulty of acquiring them. The local adaptation of abstract language "systems" is precisely what VUE models would suggest. To the extent that language involves a "system," it is not one so delicate as to be functionally compromised by the loss of certain contrasts. Rather, the system is adaptive and resilient, not 'rigid, homogeneous, self-contained or finely "balanced"' (Croft 2000: 231). For both of these reasons—the particularities of learners' L1s and the resilience of language systems—we may therefore be mistaken if we automatically prioritize high FL "errors" in teaching and assessment. As English is decolonized and appropriated by different cultures, the frequency with which words occur and the ways they are realized in phonological terms will show local patterns of variation. Although the FL principle would predict substantial continuity, assuming that the centripetal forces of written language and literacy remain in place, it is important to avoid treating FL measures as language universals. The FL-informed approach taken by some pronunciation teaching handbooks (e.g., Celce-Murcia et al., 2010) needs to be tempered by an awareness of local patterns of variation, and of the limitations of existing measures of FL. It may even be that the proper application of FL lies more in *post hoc* explanation than in the prediction of difficulties or the formulation of teaching priorities.

There is an urgent need to conduct studies of FL in different contexts around the world. These should take advantage of corpus data and statistical modeling, to avoid the continuing reliance on lists drawn up over four decades ago. Such studies also need to grapple with the theoretical complexities of FL, and decide how to model such factors as frequency, as Brown (1991) attempted to do. Suitably reconceptualized and remodeled, the century-old concept of FL can continue to serve as a useful heuristic in assessing questions of language acquisition and language teaching, and relating them in turn to language usage and language change.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Best, C. (1995). "A Direct Realist View of Cross-Language Speech Perception: Standing at the Crossroads," in *Speech Perception and Linguistic Experience: Issues in Cross- Language Research*. Editor W. Strange (Baltimore: York Press), 171–204.

Brown, A. (1991). *Pronunciation Models*. Singapore: Singapore University Press.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press. doi:10.1017/cbo9780511612886

Catford, J. C. (1987). "Phonetics and the Teaching of Pronunciation: A Systematic Description of English Phonology," in *Current Perspectives on Pronunciation: Practices Anchored in Theory*. Editor J. Morley (Washington, DC: TESOL), 87–100.

Celce-Murcia, M., Brinton, D. M., and Goodwin, J. M. (2010). *Teaching Pronunciation: A Course Book and Reference Guide*. 2nd Edition. New York: Cambridge University Press.

Croft, W. (2000). *Explaining language change: an evolutionary approach*. Harlow: Longman.

Deterding, D. (2013). *Misunderstandings in English as a Lingua Franca: An Analysis of ELF Interactions in South-East Asia*. Berlin: Walter de Gruyter. doi:10.1515/9783110288599

Deterding, D., Wong, J., and Kirkpatrick, A. (2008). The Pronunciation of Hong Kong English. *English World-wide* 29 (2), 148–175. doi:10.1075/eww.29.2.03det

Hockett, C. (1967). The Quantification of Functional Load. *Word* 23, 320–339. doi:10.1080/00437956.1967.11435484

Hung, T. (2000). Towards a Phonology of Hong Kong English. *World Englishes* 19 (3), 337–356. doi:10.1111/1467-971X.00183

Jakobson, R. (1931). *Prinzipien der historischen Phonologie*, 4. Prague: Travaux du cercle linguistique de Prague, 246–267.

J. Bybee and P. Hopper (2001). in *Frequency and the Emergence of Linguistic Structure* (Amsterdam: John Benjamins).

Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.

Jones, C. (2012). "Phonology,". *English Historical Linguistics*. Editors A. Bergs and L. J. Brinton (Boston & Berlin: Mouton de Gruyter), Vol. 1, 827–841.

Kang, O., and Moran, M. (2014). Functional Loads of Pronunciation Features in Nonnative Speakers' Oral Assessment. *Tesol Q*. 48, 176–187. doi:10.1002/tesq.152

Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.

Levis, J. M., and Cortes, V. (2008). "Minimal Pairs in Spoken Corpora: Implications for Pronunciation Assessment and Teaching," in *Towards Adaptive CALL: Natural Language Processing for Diagnostic Language Assessment*. Editors C. A. Chapelle, Y.-R. Chung, and J. Xu (Ames, IA: Iowa State University), 197.–208

Levis, J. M. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. Cambridge: Cambridge University Press. doi:10.1017/9781108241564

Lim, L. (2004). "2. Souding Singaporean," in *Singapore English: A Grammatical Description*. Editor L. Lim (Amsterdam: John Benjamins), 19–56. doi:10.1075/veaw.g33.04lim

Martinet, A. (1952). Function, Structure, and Sound Change. *Word* 8, 1–32. doi:10.1080/00437956.1952.11659416

Munro, M. J., and Derwing, T. M. (2006). The Functional Load Principle in ESL Pronunciation Instruction: An Exploratory Study. *System* 34 (4), 520–531. doi:10.1016/j.system.2006.09.004

Oh, Y. M., Coupé, C., Marsico, E., and Pellegrino, F. (2015). Bridging Phonological System and Lexicon: Insights from a Corpus Study of Functional Load. *J. Phonetics* 53, 153–176. doi:10.1016/j.wocn.2015.08.003

Sarmah, P., Gogoi, D. V., and Wiltshire, C. (2009). Thai English: Rhythm and Vowels. *English World-wide* 30, 196–217. doi:10.1075/eww.30.2.05sar

Sewell, A. (2017). Functional Load Revisited. *Jslp* 3, 57–79. doi:10.1075/jslp.3.1.03sew

Sewell, A. (2010). Research Methods and Intelligibility Studies. *World Englishes* 29, 257–269. doi:10.1111/j.1467-971x.2010.01641.x

Suzukida, Y., and Saito, K. (2019). Which Segmental Features Matter for Successful L2 Comprehensibility? Revisiting and Generalizing the Pedagogical Value of the Functional Load Principle. *Lang. Teach. Res.*, 1–20. doi:10.1177/1362168819858246

Weber, A., and Cutler, A. (2004). Lexical Competition in Non-native Spoken-word Recognition. *J. Mem. Lang.* 50 (1), 1–15. doi:10.1016/S0749-596X(03)00105-0

Wedel, A., Jackson, S., and Kaplan, A. (2013a). Functional Load and the Lexicon: Evidence that Syntactic Category and Frequency Relationships in Minimal Lemma Pairs Predict the Loss of Phoneme Contrasts in Language Change. *Lang. Speech* 56 (3), 395–417. doi:10.1177/0023830913489096

Wedel, A., Kaplan, A., and Jackson, S. (2013b). High Functional Load Inhibits Phonological Contrast Loss: A Corpus Study. *Cognition* 128 (2), 179–186. doi:10.1016/j.cognition.2013.03.002

Conflict of Interest: 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.

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fcomm.2021.662934

Developmental Sequences in Second Language Phonology: Effects of Instruction on the Acquisition of Foreign sC Onsets

Walcir Cardoso[1]*, *Laura Collins*[2] *and Walciléa Cardoso*[2]

[1]Concordia University and Centre for the Study of Learning and Performance, Montreal, QC, Canada, [2]Centro de Educação de Jovens e Adultos and Escola Municipal Manuela Freitas, Belém, Brazil

This study investigates the effects of three types of instruction on the acquisition of foreign /s/-initial onset clusters (/sl/, /sn/, and /st/-sC clusters), a process characterized by a developmental sequence in which /sl/ is acquired before /sn/ and /st/. 118 native speakers of Brazilian Portuguese participated in a 4-week language course to learn a set of vocabulary and associated pronunciation of Taki, a self-constructed miniature linguistic system. The participants were divided into three groups, each corresponding to one of the hypotheses that characterize three types of explicit second language instruction: Teachability, Projection of Markedness (Projection), and a combination of the two (Mixed). A mixed analysis of variance (ANOVA) analysis of the participants' production of sC clusters revealed that, in one of the tasks employed (read aloud), the group that focused exclusively on the more marked/st/ (Projection Group) had the best overall performance in the acquisition of the three clusters: this group was able to generalize their knowledge to the sC clusters they were not taught. In general, the results support Zobl's Projection Model of Markedness for explaining the phonological development of syllable structure, wherein the instructional effect of a focus on the most marked /st/ projects to knowledge of the less marked structures.

Keywords: developmental sequences, second language phonology, syllable structure, miniature phonology, sC clusters, artificial language

OPEN ACCESS

Edited by:
John Archibald,
University of Victoria, Canada

Reviewed by:
Carrie Pichan,
University of Illinois at Chicago,
United States
Raquel Santos,
University of São Paulo, Brazil
Robert Carlisle,
California State University, Bakersfield,
United States

****Correspondence:***
Walcir Cardoso
walcir.cardoso@concordia.ca

Specialty section:
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: *01 February 2021*
Accepted: *06 May 2021*
Published: *28 May 2021*

Citation:
Cardoso W, Collins L and Cardoso W
(2021) Developmental Sequences in
Second Language Phonology: Effects
of Instruction on the Acquisition of
Foreign sC Onsets.
Front. Commun. 6:662934.
doi: 10.3389/fcomm.2021.662934

INTRODUCTION

The literature on phonological acquisition is replete with studies showing that the oral production of/ s/-initial onset clusters (sC; e.g., /st/op, /sl/eep, /sn/ow) is particularly problematic for first language acquirers (L1; Goad and Rose, 2004; Yavaş & Barlow, 2006) and second/foreign language learners (L2; Carlisle, 2006; Major, 1996).[1] Some of these studies also reveal that the acquisition of sC follows a "natural order of acquisition" or developmental sequence in which the /s/ + sonorant sequences /sl/

[1]English is used and studied in Brazil as a *foreign* language. We use the term *second* language in this paper in its broader sense of the learning of an additional language.

and /sn/ tend to appear before /st/ (Carlisle, 2006; Yavaş & Barlow, 2006; Boudaoud, 2008; Cardoso & Liakin, 2009).

From a pedagogical perspective, these studies raise an interesting question with regards to instructional intervention: Will the effects of a focus on the form that is acquired late (and assumed to be difficult) project to the forms that are usually acquired early (and assumed to be easy)? Or will the reverse lead to a more successful mastery of developmental sequences, as is often implied in the design of L2 instructional materials in which these sequences are introduced starting from the easy end of the hierarchy?

Research investigating this issue has focused on the instruction of morphosyntax and has yielded mixed findings (reviewed below). The pedagogical implementation of these ideas from an L2 phonological perspective includes suggestions that problematic L2 sounds be taught first (e.g., Eckman & Iverson, 1997; Doughty & Williams, 1999, p. 21), or *via* tasks that progress in stages from easy to more challenging (Pennington, 1999). However, to our knowledge, there has been no published research that has tested these claims (but see a pilot study by Cardoso, 2010). One of the goals of this study to address the issue from a *phonological* perspective and thus lay groundwork for future research on the acquisition of L2 phonological developmental sequences.

LITERATURE REVIEW

Natural Order of Acquisition or Developmental Sequences

The concept of a natural order of acquisition or developmental sequence characterized some of the "morpheme order" studies that propagated in the 1970s and 1980s. Influenced by Chomsky (1970) claims of a language acquisition device or Universal Grammar, many linguists were concerned with uncovering the innate knowledge that guided language acquisition by showing that learners mastered linguistic structures in the same order, regardless of the quality and quantity of input in the ambient language, or the first language in the case of L2 learning. For L1 acquisition, two of the most prominent studies are those of Brown (1983) and de Villier and de Villiers (1977), who investigated the development of English morphosyntactic features in children. They observed that the present progressive-*ing* form was the first morpheme to be acquired, followed by a set of other morphosyntactic elements (e.g., articles *the*/*a* followed by the possessive'*s*), and culminating with the mastery of contractible auxiliary forms (e.g., 's in *Daddy's eating*).

In the context of L2 acquisition, the studies by Bailey et al. (1974), Larsen-Freeman (1975), and Rosansky (1976) reported similar results confirming most of the abovementioned findings, *mutatis mutandis*. In these studies, speakers of a variety of L1s (Arabic, Farsi, Japanese and Spanish) acquired English morphosyntax obeying a developmental sequence that initiated with the present progressive-*ing*, advanced towards the articles *the*/*a*, and concluded with the possessive'*s*. Inspired by Brown (1983) hypothesis for L1 acquisition, suggestions by Corder (1967) and empirical evidence by Dulay and Burt (1978), the concept of a natural order for L2 acquisition was later formalized by Krashen (1981) in the form of the Natural Order Hypothesis (see also Ellis, 1997; Lightbown, 1980; Spada and Lightbown, 1999; Wode, 1976 for actual studies documenting developmental sequences in L2 acquisition, and Kwon, 2005 for a review of the literature on natural order morphemes).

Although not as widely investigated as in morphosyntax, developmental sequences have also been observed in L1 phonological acquisition, particularly involving the development of segments and syllable structure. For segments, for example, it is common knowledge that vowels are acquired before consonants (e.g., Jakobson, 1968; Davis and MacNeilage, 1990), and that stops are mastered before fricatives and liquids, in that order (e.g., Bernhardt and Stemberger, 1998). With regard to syllable structure, onsets have a universal tendency to be acquired before codas (e.g., Smith and Stoel-Gammon, 1983; Vihman, 1996) and, pertinent to this study, /s/ plus liquid onset clusters (/sl/) are usually mastered before /s/ plus stop sequences (e.g., Smith, 1973; Gerrits and Zumach, 2006; Yavaş and Barlow, 2006, and Hefter and Cardoso, 2010).

There is also evidence that L2 phonological development follows similar natural order patterns. For example, in the development of L2 syllable structures, learners tend to produce more errors in word-final codas than in word-initial onsets (e.g., Flege and Davidian, 1984; Cardoso, 2007). For sC onsets, as mentioned earlier, acquisition follows a developmental sequence in which /s/ plus sonorant sequences (e.g., /sl/ and /sn/) are acquired before their /s/ plus plosives counterparts such as /st/ and /sk/ (Tropf, 1987; Carlisle, 1991, 2006; Rauber, 2006; Yavaş and Barlow, 2006; Boudaoud, 2008; Cardoso & Liakin, 2009).

In sum, there is convincing empirical evidence to substantiate the claim that acquisition of certain linguistic items occurs in predictable orders. What is unclear is the extent to which instruction can affect the acquisition of items that comprise a given developmental sequence: Will tutored learners acquire structures in the order in which they are presented in instructional settings? If so, which end of the sequence should pronunciation teaching emphasize in order to become more effective?

Instruction and Developmental Sequences in Second Language Acquisition

One hypothesis is that instruction should follow the known order of acquisition; that is the design of L2 instructional materials in which these sequences are introduced from the easy end of the hierarchy. A second hypothesis postulates that a more effective use of instructional time is to target the later acquired (and thus more difficult to learn) form, based on the assumption that they will project to the forms that are usually acquired early. In this section, we review the empirical evidence in support of these two positions, all in the realm of morphosyntax.

The first hypothesis can be subsumed under the Teachability Hypothesis, proposed by Pienemann (1984, 1989, 1998), and later revised as Processability Theory (Pienemann et al., 2005;

Pienemann, 2007).[2] This hypothesis predicts that a novel linguistic form can only be acquired when learners are developmentally ready, when they "can produce and comprehend only those L2 linguistic forms that the current state of the language processor can handle" (Pienemann, 2007, p. 137), and when they are able to process the structures that will lead them to the next developmental stage (Meisel et al., 1981). Mackey and Goo (2007) demonstrated that when adult English learners from a variety of L1 backgrounds acquire question formation, only those who are developmentally ready (e.g., have acquired auxiliary inversion in yes/no questions-Stage 4) are able to make greater gains in higher level structures (i.e., acquire Wh-questions and copula inversion-Stage 5). Mackey and Philp concluded that, "If learners are not at the correct developmental level they will not acquire the structure; it is supposedly unlearnable, unteachable, and untreatable" (p. 340). L2 acquisition studies that also support this hypothesis include Bardovi-Harlig (1995-English pluperfect), Ellis (1984-English Wh-questions in children; 1989-German word order), Felix (1981-English negation, interrogation, and other morphosyntactic structures), Jensen et al. (2003) and Pienemann et al. (1988). Based on typological explanations and first language acquisition research, a common denominator among these studies is that they demonstrate that learners acquire the most common and most basic structures first, and the rare and more complex ones last, if at all.

A second proposal for the investigation of the effects of instruction in relation to the order of acquisition of particular grammatical structures is the Projection Model of Markedness, proposed by Zobl (1983, 1985). Contrary to the Teachability Hypothesis, this proposal advocates an instructional focus on more advanced or more marked structures. The prediction is that an instructional focus on these more complex forms might lead (project) to the learning of basic or less marked structures. In a study investigating the effects of different types of instruction on the L2 acquisition of Japanese relative clauses, Yabuki-Soh (2007) showed that an instructional focus on a more marked relative clause [e.g., the Japanese equivalent of "The person (whom I had a fight with)"] facilitated the learning of less marked relativization [e.g., "The person (who gave me a book)"]: learners who were taught the more marked relative clause were able to "project" that knowledge to lower-level structures and thus generalize relativization rules to simpler contexts. This hypothesis is supported by several studies, mostly involving the acquisition of relative clauses in English (Gass, 1982; Doughty, 1988, 1991; Eckman et al., 1988), French (Mitchell, 2001), and Japanese (Yabuki-Soh, 2007), or possessive determiners in English (Zobl, 1985). Interestingly, the hypothesis has also been observed in speech pathology, with research showing that the treatment of marked fricatives enhances the learning of unmarked stops (Dinnsen and Elbert, 1984), while a focus on consonant clusters leads to an overall improvement in the production of less marked singletons (Gierut, 1999; Gierut and Champion, 2001).

Finally, a third hypothesis to which we will refer as "the Mixed Approach" questions the efficacy of step-by-step teaching of specific items or features following a given developmental sequence (Shirai, 1997; Lightbown, 1998; Spada and Lightbown, 1999). Based on the scarcity of empirical evidence (i.e., all involving the acquisition of morphosyntactic features, and most with English, French or German as the target languages) and the inconclusiveness of the available studies favouring either one of the hypotheses, proponents of the Mixed Approach for teaching developmental sequences suggest that *both* more complex and less complex structures should be emphasized in instruction. This view is shared by Ammar and Lightbown (2004), who investigated the acquisition of English relative clauses by Arabic speakers and found that, regardless of the form emphasized in teaching, learners were able to generalize to the opposite end of the developmental hierarchy. These findings led the authors to conclude that combining different types of relative clauses in instruction can be as effective as starting at either end of the developmental sequence. Similarly, Shirai (1997) recommended that the instruction of natural order phenomena be conducted in a way that emphasizes exposure to both marked and unmarked structures.

Aside from a pilot study conducted by Cardoso (2010), we are not aware of any other published study that examines the effects of teaching (instructional intervention, as defined earlier) on the acquisition of developmental sequences from a phonological perspective. One of the goals of this study is to contribute to our understanding of the nature of pedagogical interventions and their effects on learning and, more importantly, to address this gap in the literature by examining, in an instructional setting, the L2 acquisition of a phonological developmental sequence: foreign onset sC clusters.

The L2 Acquisition of sC Onsets

The oral production of foreign sC onsets is notoriously difficult for learners whose first languages disallow such sequences (Major, 1996; e.g., Japanese, Portuguese, Spanish, Turkish). In the context of Brazilian Portuguese (BP) speakers learning an "sC language" such as English, French or German, for instance, learners variably syllabify the cluster *via* a prothetic (i) (i-epenthesis), thus triggering the resyllabification of the original onset into a nucleus-coda sequence (e.g., /st/op → [is. t]op). While i-epenthesis can be usually attributed to an L1 effect (sC onsets are non-existent in BP and i-epenthesis is the most commonly-employed strategy for syllabifying illicit structures; see also Cristófaro Silva and Freitas (2020), who found that sC clusters are phonetically represented as either [is. C] for native words, or [sC] for loanwords), the phenomenon is rather complex and is motivated by a variety of linguistic and extralinguistic factors (Cardoso and Liakin, 2009).

[2] Note that Processability Theory (or the Teachability Hypothesis) is a theory of L2 development that was proposed to analyse morphosyntactic structure, not phonological phenomena such as the one addressed in this study. However, the general premises of the Theory can be easily extended to the analysis of other linguistic components: 1) It involves a developmental trajectory; 2) processing components operate automatically, i.e., they are not consciously controlled; 3) processing is incremental; 4) the output of the processor is linear; and 5) processing has access to a temporary memory store that can hold grammatical information (Pienemann, 1998, 2007). The patterns observed in sC acquisition clearly satisfy these requirements, with respective differences taken into consideration.

One of these factors include the concept of sonority, defined *via* a combination of features such as amplitude or intensity (Ladefoged, 1993), acoustic energy (Goldsmith, 1989), and propensity for voicing (Kenstowicz, 1994). Together, and focusing exclusively on the set of relevant segments, these features determine a sonority hierarchy that ranges from the least sonorous stop/t/ to the more sonorous liquid /l/: /t/ </s/ </n/ </l/ (where "<" indicates "less sonorous than"). In order to constitute legitimate onset clusters, the sC combination should follow a pattern in which sonority progressively rises towards the nucleus of the syllable (Sonority Sequencing Principle-SSP, Selkirk, 1984). While both the /sl/ and /sn/ satisfy the SSP requirement because sonority rises from /s/ to the following segment, /st/ constitutes an SSP violation because sonority sequencing decreases in the second consonant. In addition, onset cluster syllabification tends to favour sequences that have a "maximal and most evenly-distributed rise in sonority" (Minimal Sonority Distance-Clements, 1990, p. 303). This preference favours the sequence /sl/, which has a wider sonority distance than /sn/. Appealing to the concept of markedness (e.g., de Lacy, 2006), one may then assume that these three sequences constitute a hierarchy in which /sl/ is the least marked, followed by /sn/ and then the most marked /st/: /sl/ </sn/ </st/ (where "<" indicates "less marked than"). The implication of this generalization based on markedness is that learners will have less difficulty in acquiring the least marked /sl/ than the more marked /sn/ and /st/ clusters.[3] Not surprisingly, this is exactly what is found in a number of studies that investigate L2 sC acquisition, as will be discussed next.

The majority of the literature on L2 sC acquisition indicates that the path to sC development initially favours unmarked segments such as /s/ + liquids, which are acquired earlier and with less difficulty than their more marked counterparts (e.g., Tropf, 1987; Carlisle, 1991, 2006; Rauber, 2006; Boudaoud, 2008; Cardoso and Liakin, 2009). In a study involving the same community of BP speakers examined in this investigation, Cardoso and Liakin (2009) found a pattern of sC development that reflects the tendency found in earlier studies. Their participants produced /sl/ and /sn/ (the difference between these two SSP-abiding clusters was not statistically significant) more accurately than the more marked /st/, similar to what is observed in L1 acquisition (e.g., Hefter and Cardoso, 2010; Yavaş and Barlow, 2006), and in the development of L2 phonologies, as discussed above.[4] Studies that contradict this order of acquisition may be explained by methodological limitations: the low number of participants (e.g., Abrahamson, 1999-one participant; Major, 1996-four participants) and the inclusion of a large number of heterorganic and complex sC clusters such as /sp/, /skr/, and /spr/ (Escartin, 2005). Two additional factors that may also play a role include L1 transfer phenomena (e.g., devoicing of epenthetic [i] in Brazilian Portuguese, which leads to a misinterpretation of [isC] as target-like [sC]-Major, 1996); 2); and frequency effects in the L2 input (e.g., Escartin, 2005 maintains that the unexpected low performance in /s/ + liquid sequences was possibly due to the low frequency of these cluster types in English). Both of these factors have recently received research attention in the literature on the development of morphosyntax (e.g., Luk and Shirai, 2009 on L1 influence on the L2 acquisition order of grammatical morphemes; Collins et al., 2009 on the interaction between input frequency and tense-aspect acquisition).

The insights provided from markedness theory on sC syllabification and the empirical evidence just discussed allow us to substantiate the claim that the acquisition of these clusters is characterized by the following developmental sequence (where ">>" indicates "acquired before" and "(>>)" suggests an inconclusive but well-motivated pattern based on markedness and some previous studies): /sl/ (>>) /sn/ >> /st/.

A review of the literature also reveals some confounding factors that may interfere in the acquisition of sC. Firstly, some of these studies suggest that the heterorganicity of the onset constituents might have an effect on the production of sC (Boudaoud, 2008; Cardoso and Liakin, 2009). For instance, while the /s/ + nasal /sn/ and /sm/ clusters are equally marked with respect to sonority sequencing, as discussed earlier, they differ in place of articulation (while /sn/ is comprised of two coronal segments, /sm/ contains the coronal and labial articulators). Considering Clements' (1990) Sequential Markedness Principle ("For any two segments A and B and any given context X_Y, if A is simpler than B, then X$\underline{A}$Y is simpler than X$\underline{B}$Y"; p. 313) and the fact that the coronal /n/ is less marked than the labial /m/ (Prince and Smolensky, 2004), it follows that the /sm/ sequence is the

[3] Using L1 data from West Germanic languages and based on an analysis that considers/s/ in sC clusters an Appendix constituent (i.e., the segment is directly linked to the syllable node, thus overpassing the Onset), Goad and Rose (2004) conclude that these sequences correspond to what UG provides as unmarked, thus contradicting our assumption of a markedness relationship between the three clusters. The authors acknowledge, however, that the Appendix and consequently the unmarked analysis for sC clusters "is often not well accepted" (p. 123), and this is particularly the case in the L2 literature (e.g., Carlisle, 1991, 2006; Major, 1996, 2001; Cardoso and Liakin, 2009). Other less orthodox proposals for sC representation include the assignment of/s/ as the first member of a complex segment (e.g., Selkirk, 1982) or as an adjunct to the syllable (Barlow, 2001). Along the lines of Boyd (2006) and based on robust empirical evidence from L1 and L2 acquisition studies, we assume the standard view that no structural distinction exists between sC and other complex onset cluster and, accordingly, that there exists a markedness relationship between the segments that comprise the sC set, as established in this paper.

[4] One could also argue that the frequency of epenthesized forms of sC in the L1 Brazilian Portuguese could also affect its acquisition (e.g., the high frequency of [is.t] vis-à-vis [is.n] and [is.l] in BP could lead learners to have more difficulty in acquiring the target [st] sequence). While this seems to be a valid hypothesis, we believe that the issue is more complex than what is implied here. For instance, while the described L1 effect could be argued for production (Cardoso and Liakin, 2009; i.e., the [ist] cluster is indeed more frequent in BP, which might hinder the acquisition of the target [st] form), it would not hold for perception, since the target [st] has a higher propensity to be perceived accurately than the other clusters (Cardoso et al., 2009). The effects of the L1 on different types of sC instruction are being addressed by a larger program of research by some of the authors.

most marked of the two clusters and, consequently, more likely to be acquired last.[5] Motivated by the Sequential Markedness Principle and the allegation that heterorganicity may affect sC development, this study will focus exclusively on the acquisition of the homorganic sC clusters /sl/, /sn/ and /st/.

Secondly, some of these studies allude to the fact that the frequency with which the sC structures occurs in the L2 input might also affect their acquisition (Escartin, 2005; but see Cardoso and Liakin, 2009 for evidence to the contrary). This analysis is based on the assumption that the productivity of a pattern is determined by its type and/or lexical frequency: "the more items (are) encompassed by a schema, the stronger it is, and the more available it is for application to new items" (Bybee, 2001, p. 13; see also Archibald and Libben, 1995 and Trofimovich et al., 2007 for similar claims).

Finally, it is possible that some of the divergences found in previous research could be attributed to the type of instruction that the participants received when learning the target language. The studies consulted, for instance, provide no information about the characteristics of the teaching environment, particularly on whether the production and perception of sC received any instructional focus in the classroom and, if that was the case, on how pronunciation was taught. One could presume, for instance, that a given group of participants produced the most difficult /s/ + stop sequences more accurately simply because these learners took part in pronunciation activities that targeted this form. Being the most frequent sC structure in English (/st/ occurs in over 87% of all sC forms found in a student-directed teacher talk corpus-Cardoso and Liakin, 2009), an imbalanced focus on /st/ is unavoidable in any pronunciation activity containing the cluster. Information about the type of instructional exposure is also important because techniques such as corrective feedback can enhance L2 learning (Lyster and Saito, 2010).

In sum, any study investigating the effects of types of instruction on the L2 acquisition of sC should take into consideration the confounding factors of heterorganicity within the cluster, the frequency distribution of the relevant structures in the target L2, and the types of instruction to which learners have been exposed. While the place of articulation confound can be easily addressed by confining the set of learnable targets to coronal plus coronal sequences such as /sl/, /sn/, and /st/, the target L2 input and the type (and quality) of previously experienced pedagogical interventions cannot be reliably controlled in a standard language classroom. To address these limitations, this study adopts a miniature linguistic system (MLS), Taki, an artificial language designed to allow for control of the input participants encounter.

Adopting a Miniature Linguistic System

In research conducted with natural language, variables that are difficult to control and/or determine include the distribution of target forms in the input, participants' previous experience with the L2 and their level of proficiency, and the quality and quantity of previous instruction). A number of SLA scholars have outlined the potential contributions of an MLS to control and manipulate key variables (Cook, 1988; VanPatten, 1990; DeKeyser, 1995; Ellis and Schmidt, 1997; Hulstijn, 1997; see also the articles in a 1997 special issue of *Studies in Second Language Acquisition* on laboratory research methods). We designed Taki (described below) to investigate sC learning, as it allowed us 1) to control the language to be learned so that the target input could be easily manipulated (quantitatively and qualitatively) to accommodate the demands of our study; 2) to guarantee that no participants in the experiment would have advanced knowledge of the target structures; and 3) to ensure that all participants received the same type of instruction in which only the order of sC presentation is manipulated, as predicted by our research questions. The adoption of an MLS thus increases the reliability of our study due to the control over the target language and the instructional environment. However, we are aware that the use of an MLS does raise ecological validity issues, and the interpretation of the findings must also consider potential limitations to the teaching and learning of natural languages. We return to these points in the discussion of the results.

This Study

The purpose of the present study is to explore the effects of three types of exposure (to which we also refer as instruction) on the development of foreign sC onsets and, at the same time, to observe how these same clusters are acquired at the end of a series of instructional interventions. This study's research question is thus formulated as follows:

- How do the three types of exposure (instructional treatment) affect the L2 acquisition of sC clusters in production? Specifically, which type of instruction is more effective for the teaching of sC clusters?

It remains difficult to predict outcomes based on the existing literature on L2 instruction and phonological developmental sequences, as it is non-existent. In addiction, the findings from the studies of morphosyntax have yielded contradictory findings.

Table 1 summarizes the three hypotheses entertained by the current study and their respective rationale, accompanied by the teaching order that they advocate (where ">" indicates "should be taught before" and "=" means "should be taught together with"):

METHOD

Participants and Experimental Groups

Hundred-eighty seven participants were initially recruited to participate in the study. However, due to previous formal learning experience with an sC language (e.g., English, French, German) or familiarity with the target structure (detected *via* the

[5] Obviously, the homorganic sC clusters violate a phonotactic constraint against homorganicity (the Obligatory Contour Principle for Place: Adjacent identical place features are prohibited; Carlisle, 2006; Barlow, 2001; Goad and Rose, 2004), which is strongly operative in English as the following unattested sequences illustrate: *dl, *tl, *pw, *fw. The three target sC clusters are equally marked regarding this constraint, with sonority being the only variable feature.

TABLE 1 | Developmental sequences and teaching: Hypotheses and advocated teaching order.

Hypothesis	Rationale	Teaching order advocated
1. Teachability	Easy evolves to hard	sl > sn > st
2. Projection model of markedness	Hard projects to easy	st
3. Mixed	Contra step-by-step	st = sn = sl

pretest), 18 participants were excluded from the pool of potential participants. For personal reasons, 51 other participants left the experiment without completing one or more of the teaching sessions or tests. The remaining 118 were primary and secondary students with ages ranging from 11 to 22 (Mean = 14.4 years old), enrolled in two public school in the city of Belém (Brazil), a community of primarily monolingual Portuguese speakers. The participants were all monolingual native speakers of Portuguese, without any previous oral experience with an sC language. The 118 participants were randomly assigned to one of the three experimental groups, each constituting an intact class: The Projection of Markedness Group (P Group; *N* = 38), the Teachability Group (T Group; *N* = 38) and the Mixed Group (M Group; *N* = 42).

As part of the school curriculum in Brazil, students have compulsory weekly 1.5-h English classes that focus exclusively on the acquisition of morphosyntax (grammar) and translation skills (Lopes, 1996; Izidro, 2007), and sometimes on receptive (written) vocabulary and reading comprehension in order to fulfill the requirements for high-stakes exams such as the “vestibular” (a competitive nationwide examination that selects students for entry into a university program). In these typically very large classes (40–60 students), oral interactions in English are rare (Lima et al., 2014). Participants had thus had very limited exposure to spoken English.[6]

Research Design

The study employed a quasi-experimental, within groups pretest/posttest design. To examine the development of sC acquisition, it included a pretest (to measure the participants' initial knowledge of sC in oral production), an immediate posttest (week 4) and a delayed posttest (conducted 1 week after the last pedagogical intervention which, according to standards from Form-Focused Instruction research, could be classified as a “short-delayed posttest”; Spada and Tomita, 2010). The experiment lasted 4 weeks (not including the delayed posttest) and consisted of three teaching sessions (the grey area in **Figure 1**), each designed according to the three types of instruction considered in the study: While P Group was taught exclusively /st/-initial words, T Group was taught one sC per session (/sl/ > /sn/ > /st/), following their natural order of acquisition. Finally, M Group was taught all three sC sequences throughout the duration of the Taki course. The research design adopted in this study is illustrated in **Figure 1**.

Procedure

Materials: Target of Instruction and Tests

The teaching and testing materials used in this study consisted of 162 target sC-initial words and 48 distractors, equally divided among the clusters, created using WordGenerator 1.9 (http://billposer.org/Software/WordGenerator.html), a computer application that generates hypothetical words based on designated specifications such as segmental content, syllable structure and word size. To ensure that the only the learnable structure was the target sC cluster and that all remaining structures were of easy articulation, a number of phonetic, phonological, orthographic, morphological and semantic criteria were obeyed in the design of the Taki words (in the examples, “.” defines syllable boundaries and “'” indicates stress):

1. Word size: words were all disyllabic (e.g. ['sle.gak] “hat” ['sta.mik] “dress”), based on the observation that dissyllabic structures are unmarked and, consequently, more easily acquired (Broselow, Chen, & Wang, 1998).
2. Foot structure (stress): trochaic, stressed on the leftmost syllable as is the case for English sC-initial words (e.g. ['snu.pak] “bird” ['sti,kab] “tie”), and following the “trochaic bias” proposed for language acquisition (Allen and Hawkins, 1978; Adam and Bat-El, 2009; but see Rose and Champdoizeau, 2007 for an opposing view on this bias).
3. Skeletal (syllabic) structure: sCV.CVC [e.g. ('sta.nud) “train” ('sla.pid) “watch”], due to the requirements of the experiment. Word-final consonants were included for another study targeting coda acquisition and, accordingly, they will not be discussed in this paper.
4. Segmental content: five vowels [(a e i o u)] and thirteen consonants [(p t k g b d f v s z m n l)] were utilized, all considered segments of easy articulation in the participants' L1.
5. Morphology: words were devoid of superfluous inflectional or derivational morphology (they were simple, uninflected words).
6. Orthography: words followed strict one-to-one grapheme-to-phoneme associations (no digraphs and diacritics were employed), based on the effects that L2 sound-to-spelling mismatches may have on vocabulary acquisition (Ludwig, 1984; Nation, 1990).

[6] In Brazil, it is widely accepted that “[private] language courses seem to constitute the only environment where one is likely to learn the English language” (Gasparini, 2005, p. 159; translation from Portuguese; see also Izidro, 2007 and Lopes, 1996 for similar claims). While this is especially true of the public-school system, it is also a characteristic of most private schools. The reasons for this situation include: pedagogical goals that are unattainable (e.g., teaching the four language skills; Lopes, 1996), large class sizes (Gasparini, 2005), limited class time dedicated to language teaching (Izidro, 2007), the teachers' low proficiency in English and limited knowledge of current L2 pedagogy (Gasparini, 2005), and possibly the pressure of lobby groups in a country where language courses are considered excellent business investments.

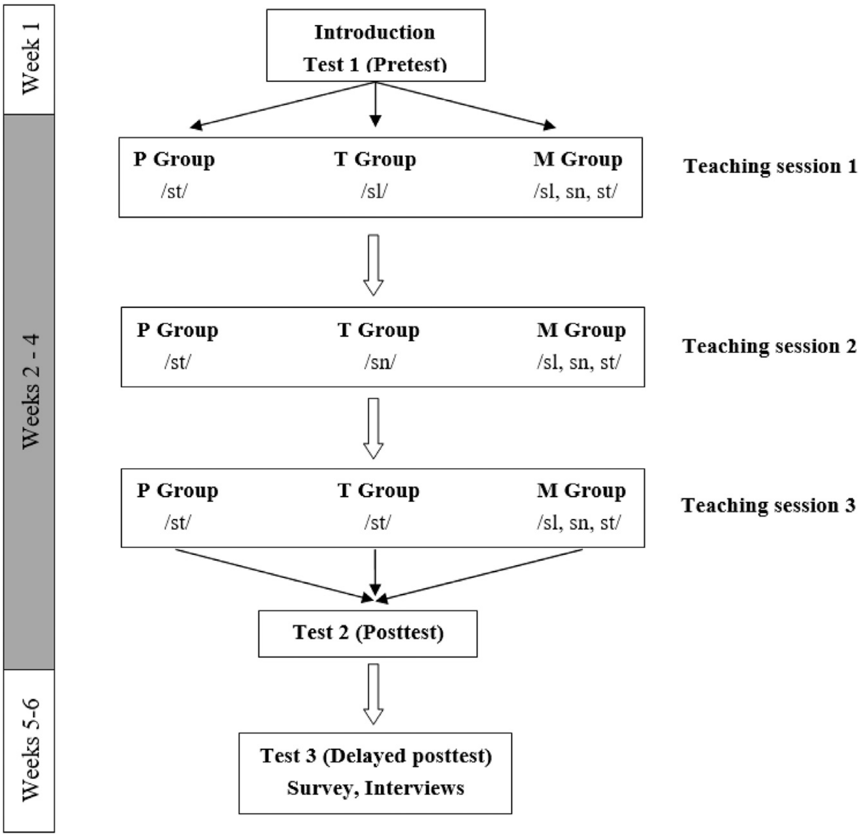


FIGURE 1 | Research design.

7. Semantic content: words were assigned to specific meanings randomly and consisted of concrete and unambiguous nouns. This decision was based on Thornbury (2002) claim that concrete meanings are more easily acquired and less susceptible to forgetting (see also Nation, 1990 and de Groot and Keijzer (2000) for similar claims). It is also in agreement with MacWhinney (1983) and Moeser and Bregman (1973) findings suggesting that words in miniature linguistic systems are better learned when the communicative context is maximized by explicit referential content.

In sum, the abovementioned criteria were motivated by attempts to manipulate only the relevant foreign sC onset and to minimize the influence of potential extraneous factors such as semantic or morphological complexity. For a complete list of the Taki words used in the three tests administered, see **Supplementary Appendix A**.

Treatment

The Taki teaching sessions were taught by the first author, a speaker of Brazilian Portuguese who has native-like oral fluency in the language (e.g., he can pronounce the target sC clusters). On the first day of class, the participants were told that, within the period of a month, they were going to learn some words from Taki, “a language especially designed to answer some questions about learning foreign languages.” The weekly instructional sessions were conducted in a standard classroom in the school premises and lasted 45 min each, for a total of approximately 2 h and 15 min of Taki instruction. While this proposed learning phase seems short, it is considerably longer than most MLS-based studies (usually lasting between a few minutes and 1 h-Hulstijn, 1997, p. 138), and it falls within Norris and Ortega (2000) classification as a treatment of “medium” duration (see also Lyster and Saito, 2010).[7]

The teaching sessions, conducted in the participants’native language (Portuguese), focused exclusively on the learning of vocabulary and related pronunciation. During the introduction session (week 1) and before each class, participants were

[7] In a case study investigating the effects of a computer-based perceptual training on the acquisition of a large set of /s/ plus consonant onsets, Bettoni and Koerich (2009) found that in approximately 7 h of perceptual training, their participant was able to significantly improve her perception and production of sC for trained and unfamiliar words, in both immediate and delayed posttests.

reminded that they had to pay attention to what each new word means (semantics), how it is spelled (orthography), how it sounds (aural perception), and how it is pronounced (oral production). Briefly, the sessions consisted of the following teaching strategies (based on Thornbury, 2002 recommendations for teaching vocabulary and associated word knowledge):

1. Introduction of the word *via* an Open Office Impress slide projected on a screen with a picture and its associated sC word (See **Supplementary Appendix B** for a sample).
2. Pronunciation practice *via* the listening and oral production of the target word (either by orally imitating the instructor or reading it from the screen), followed by choral repetitions.
3. Personalized practice, so that learners could relate the word being learned to personal experiences (e.g., "[stovap], I use it to listen to music"). Whenever necessary (e.g., when the instructor heard students mispronouncing a given word), the instructor provided general explicit feedback by, for instance, repeating the target form (for the rationale, see Lyster and Saito, 2010).
4. Word retrieval activities such as picture naming ("what do you see in the picture?"), fill-in the blanks ("write what you see"), oral translations ("what's the word in Taki for "hand"?"), cross-word puzzles ("translate the Portuguese words into Taki to complete the puzzle"), and bingos ("listen and mark the word you heard on the scorecard").

Whenever possible, an attempt was made to present and discuss the sC words in preceding pausal environments (e.g., "[slovab], this is the word for hand in Taki") to prevent them from being lost *via* resyllabification [e.g., in natural speech, /a slovab/ will be produced as [as.lo.vab] "the hand", where the original /sl/ onset sequence re-syllabifies as a coda-onset cluster]. Finally, to reduce the effects of type/token frequency distribution in the L2 input, the amount of oral and visual exposure to each sC word was carefully monitored so that the participants received the same quantity and quality of treatment across the three experimental groups.

Measures: Data Collection and Assessing sC Production

To test the participants' developing "proficiency" in Taki, the study included two oral production tests, both targeting words in a context-free, pause-initial environment to ensure that the sC sequence remains intact and to mitigate preceding environment effects (see Carlisle, 1991 and Major, 1996 for evidence that the preceding word-final segment influences the frequency of prothesis before sC onsets). The two production tests were: 1) *Read aloud*, in which participants were presented a Taki word with its corresponding image on a Impress slide (e.g., snumid "violin") and asked to read it aloud; and 2) *Listen and Repeat* in a carrier phrase, in which the participants listened to the researcher and repeated what they heard in a pause-initial carrier phrase "____, mi vedu" ("____, I see"). Each test consisted of 18 target sC-initial words (six of each sC type) and six non sC-initial (the latter were used as distractors; see **Supplementary Appendix A**). The production tests were audio recorded *en masse*, with each participant holding a mobile, hand-held audio-recorder (Sony ICD-CDUX522). The data collected were then transcribed and coded independently by two research assistants, both native speakers of an sC language.

Accurate sC production was calculated by examining only the relevant sC form for each participant and, consequently, the remaining syllabic structures and segmental content were ignored. The target sC forms were scored as either correct (e.g., if the participant produced the cluster in a target-like manner, without a preceding or following epenthetic vowel [i] ['slu.mid] "violin") or incorrect (e.g., if the target form was produced preceded or followed by [i]-epenthesis, *[is.'lu.mid] or *[si.'lu.mid] respectively, or deleted *['lu-mid]). Cases of L1-influenced /s/ palatalization (e.g [iʃ.'ti.mut]) or /t/ affrication (e.g [is.'tʃi.mut] [iʃ.'tʃi.mut]) were deemed incorrect (all instances of /s/ palatalization or /t/ affrication consisted of /i/-epenthesized sC forms). Other incorrect (but possibly developmental) variants of the target pronunciation were coded as incorrect, but due to their relative infrequency in the corpus, they were noted in the transcription for further qualitative analyses in future research (e.g., /s/ lengthening [s:. 'lu.mid] and/or /i/-devoicing; sC substitution [ka.'kub]). In case of disparity of analysis among the two research assistants (1,210 out of 6,372 tokens = 19%, or 81% inter-rater reliability), accuracy was determined by one of the researchers, sometimes with the aid of spectrographic analyses *via* Praat (Boersma and Weenink, 2019).

The total score for each participant was determined by a calculation of the number of correct production in each test, for each of the type of sC clusters considered in the study (i.e., /st/, /sn/ and /sl). In sum, all instances of sC clusters were coded according to their accuracy in production, as well as the following independent variables: type of sC cluster (/sl/, /sn/, /st/), Test (1, 2, 3), and Instructional Group (P = Projection, T = Teachability, M = Mixed). A mixed between-within subjects analysis of variance (ANOVA) was used to calculate differences between the independent variables included in the investigation.

RESULTS

The general descriptive statistics of the analysis appears in **Tables 2**, **3**, with some of the relevant results plotted in **Figures 2**, **3**. The tables present the mean scores of the 118 participants' accurate sC production in two tests (*Read Aloud*-**Table 2**; *Listen and Repeat*-**Table 3**), across the three experimental groups: P (*n* = 38), T (*n* = 38) and M (*n* = 42), assessed at three points in time (pretest, posttest, delayed posttest). As indicated earlier, six tokens of each target sC cluster were produced per participant per test. Considering the number of dependent variables, six mixed ANOVA tests were required, with a Bonferroni adjustment to alpha = 0.0083. To verify the homogeneity and sphericity of the data, Levene's, Box's and Mauchly's tests were used.

The results revealed a Time*Group interaction for each sC cluster, indicating that the experimental groups had an overall

TABLE 2 | Descriptive statistics for sC production over time, across three experimental groups (Mean scores) in Read Aloud test.

Test	P group						T group						M group					
	sl (*n* = 6)		sn (*n* = 6)		st (*n* = 6)		sl (*n* = 6)		sn (*n* = 6)		st (*n* = 6)		sl (*n* = 6)		sn (*n* = 6)		st (*n* = 6)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	0.68	1.23	0.32	0.784	0.82	1.52	0.34	0.63	0.08	0.27	0.66	1.07	0.36	1.08	0.12	0.39	0.62	1.06
2	3.61	2.07	2.84	1.95	3.97	1.96	1.32	1.78	1.34	1.59	2.24	1.64	1.36	1.80	1.36	1.86	2.50	2.09
3	3.34	2.16	3.11	2.20	3.92	2.00	1.61	1.64	1.53	1.84	2.42	1.98	2.50	2.09	1.98	1.96	2.40	1.85

TABLE 3 | Descriptive statistics for sC production over time, across three experimental groups (Mean scores) in Listen and Repeat test.

Test	P group						T group						M group					
	sl (*n* = 6)		sn (*n* = 6)		st (*n* = 6)		sl (*n* = 6)		sn (*n* = 6)		st (*n* = 6)		sl (*n* = 6)		sn (*n* = 6)		st (*n* = 6)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	0.61	0.88	0.58	0.92	1.13	1.63	0.53	1.03	0.47	0.76	1.63	1.65	0.67	1.30	0.60	1.08	1.31	1.92
2	2.37	1.84	2.39	2.05	2.92	1.84	1.84	1.70	1.37	1.60	3.08	2.15	1.48	2.01	1.33	1.92	2.50	2.19
3	1.89	1.74	1.95	1.96	3.11	2.09	1.92	1.87	1.92	1.75	2.37	2.10	2.38	1.75	1.95	1.85	2.98	2.19

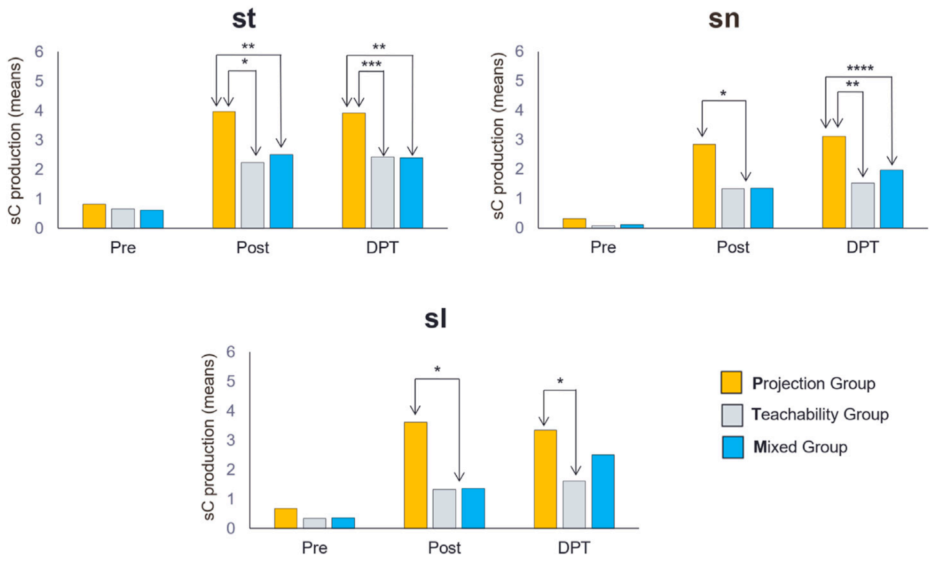


FIGURE 2 | Group Means on Read Aloud Test Over Time. Note. Significance: * p < 0.001; ** p = 0.002; *** p = 0.003; **** p = 0.040.

positive impact in the teaching of the target clusters. There were no between-group differences within sC clusters for both *Read Aloud* and *Listen and Repeat*, $F(2, 115) = 1.36$, $p > 0.05$, signaling that all clusters behaved in a similar manner across the two tests; e.g., the participants' performance improved over time across the tests, from pretest ($M = 0.44$, $SD = 0.91$) to posttest ($M = 2.26$, $SD = 2.03$) and delayed posttest ($M = 2.52$, $SD = 2.07$). Mauchly's test specified that the assumption of sphericity was violated for T Group; χ^2 (2) = 12.20, $p = 0.002$, so to gain a valid F-value, the Greenhouse-Geisser correction test was applied ($p = 0.004$). For p Group, however, the assumption of sphericity was met; χ^2 (2) = 0.56, $p > 0.05$, and AT; χ^2 (2) = 5.95, $p = 0.05$, suggesting that there were differences among the sC forms within P Group and M Group at the pretest. Because of these differences, the statistical analyses that follow will focus on Time*Group interactions for each sC cluster.

Results: *Read Aloud* Test

/st/

For /st/, a statistically significant interaction was found between the three groups and time of testing, $F(4, 230) = 4.60$, $p = 0.001$, partial $\eta^2 = 0.074$. The results of the between-subjects effects

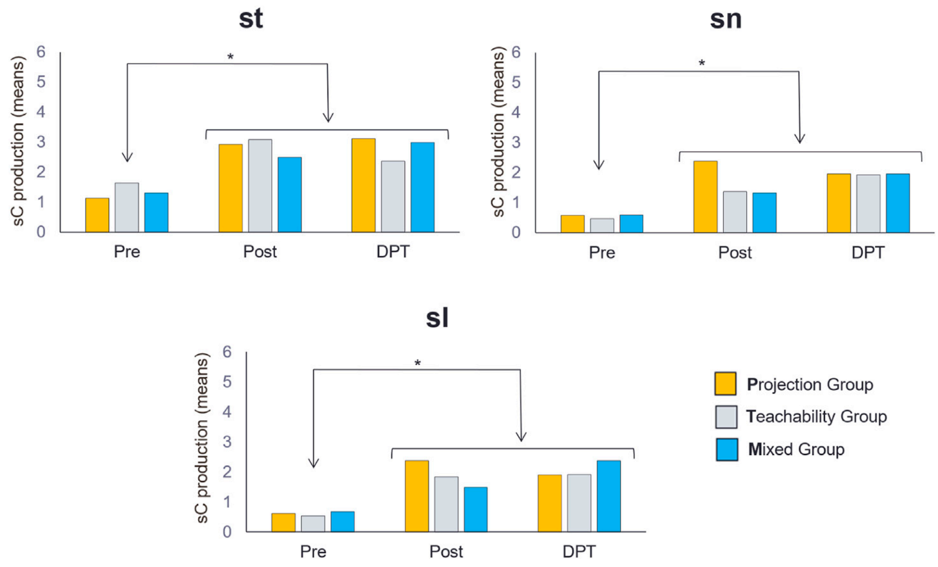


FIGURE 3 | Group Means on Listen and Repeat Test Over Time. Note. Significance: * p < 0.001.

ANOVA indicated that there was significant variation in /st/ production between the groups, $F(2, 115) = 8.77, p < 0.001$, partial $\eta^2 = 0.132$. At pretest, the Bonferroni *post-hoc* analysis indicated that there was no statistically significant effect of group between the pairwise comparisons ($p > 0.05$). However, at the posttest and delayed posttest, a statistically significant effect was observed for P Group ($p < 0.005$), in comparison with the other two groups. Pairwise comparisons of T Group and M Group were not significant at the post-and delayed post-test ($p > 0.05$).

/sn/

There was a statistically significant interaction between the three groups and time of testing for /sn/, $F(4, 230) = 4.09, p = 0.003$, partial $\eta^2 = 0.067$. The results of the between-subjects effects ANOVA indicated that there was a significant difference in /sn/ production between the three groups, $F(2, 115) = 9.71, p < 0.001$, partial $\eta^2 = 0.144$. According to the Bonferroni *post-hoc* analysis, there was no statistically significant group effect between the pairwise comparisons at the pre-test ($p > 0.05$). At the posttest and delayed posttest, however, P Group significantly outperformed both T Group ($p = 0.001$) and M Group ($p < 0.05$). Pairwise comparisons of T Group and M Group were not significant at both the post-and delayed posttest ($p > 0.05$).

/sl/

A statistically significant interaction was observed between the groups and the tests for this cluster, $F(3.883, 223.246) = 7.91, p < 0.001$, partial $\eta^2 = 0.121$. According to the analysis of between-subjects effects ANOVA, a significant difference was also observed between the three groups, $F(2, 115) = 13.99, p < 0.001$, partial $\eta^2 = 0.196$. At the pretest, Bonferroni *post-hoc* analysis revealed no significant differences in /sl/ production between the groups ($p > 0.05$). However, at the posttest, P Group outperformed both T Group and M Group (both $p < 0.001$). At the delayed posttest, P Group outperformed T Group ($p < 0.05$), but not M Group ($p = 0.184$). Pairwise comparisons of T Group and M Group were not statistically significant at the post-and delayed posttest ($p > 0.05$).

Results: *Listen and Repeat* Tests

/st/

A statistically significant interaction between the three groups and the tests was observed, $F(4, 230) = 2.86, p = 0.024$, partial $\eta^2 = 0.047$. However, the results of the between-subjects effects ANOVA revealed no significant difference between the participants' /st/ production and the experimental groups, $F(2, 115) = 0.066, p = 0.936$, partial $\eta^2 = 0.001$.

/sn/

For /sn/, there was a statistically significant interaction between the three groups and tests, $F(4, 230) = 2.91, p = 0.022$, partial $\eta^2 = 0.048$. The results of the between-subjects effects ANOVA, however, suggested that there was no significant difference between the three groups, $F(2, 115) = 1.108, p = 0.334$, partial $\eta^2 = 0.019$.

/sl/

Finally, we observed a statistically significant interaction between the experimental groups and tests, $F(4, 230) = 3.40, p = 0.010$, partial $\eta^2 = 0.056$. Based on the result of the between-subjects effects ANOVA, however, no significant difference between the three groups was observed, $F(2, 115) = 0.221, p = 0.802$, partial $\eta^2 = 0.004$.

Results: Summary

Overall, participants in all groups improved their production of sC clusters, and the observed improvement was maintained over time. Regarding the effects of different types of instruction, the findings can be divided into two categories, based on the tests adopted to assess learning.

In *Read Aloud*, the group that received instruction on the most marked /st/ (P Group) had the best overall performance in comparison with the other experimental groups (T Group and M Groups), on all sC clusters, as predicted by the Projection Model of Markedness Hypothesis. A visual representation of these findings is shown in **Figure 2**.

In *Listen and Repeat*, on the other hand, no group effects were observed, indicating that all groups equally improved on all sC clusters over time. Interestingly, the group that was taught only one single sC (/st/, P Group) had similar performance to the other groups on the two clusters that they were never taught or had a chance to practice. **Figure 2** illustrates the results obtained in the *Listen and Repeat* test, in which the three experimental groups performed similarly.

Overall, these findings indicate that, for the phonological developmental sequence under consideration, instruction that emphasizes the most marked, harder-to-acquire form is more effective for the teaching of sC clusters: the group (P) that was taught the most marked /st/ form was able to generalize the acquired structure to other less marked forms, /sn/ and /sl/. In addition, they outperformed the other two groups on the Read-Aloud task.

DISCUSSION

The goal of this study was to examine the effects of three types of instruction on the development of foreign homorganic sC onsets (/st/, /sn/, /sl/), considering three hypotheses for the teaching of items characterized by a "natural order of acquisition." Overall, the findings reported yield support for Zobl's (1993) Projection Model of Markedness inasmuch as the group that was taught the more marked /st/ (P Group) achieved the best overall results in the production of the three sC clusters in one of the tests, *Read Aloud*, and performed as well as the other groups on clusters they were not taught. These patterns are exactly what the Projection Model of Markedness predicts for linguistic items that are implicationally related in acquisition: the instructional effects of mastering the most marked /st/ cluster projects to the acquisition of the less marked forms /sl/ and /sn/. Despite the dearth of evidence demonstrating similar effects on phonological developmental sequences, our findings have parallels in the morphosyntactic literature, as will be discussed next.

The majority of the studies that corroborate the Projection Model of Markedness hypothesis involve the acquisition of either relative clauses (e.g., Doughty, 1981, 1991; Eckman et al., 1988; Gass, 1997; Mitchell, 2001, Yabuki-Soh, 2007-but see Ammar & Lightbown, 2004 for mixed results, as discussed earlier) or possessive determiners (e.g., Zobl, 1985). In contrast, the studies supporting the Teachability Hypothesis typically focus on the acquisition of distinct morphosyntactic features such as question formation (e.g., Felix, 1981; Ellis, 1984), tense and aspect (e.g., Bardovi-Harlig, 1995), and word-order (e.g., Ellis, 1984). Comparing the sets of morphosyntactic features selected to substantiate the claims of each hypothesis, it seems reasonable to assume that different teaching strategies may be necessary for the teaching of different morphosyntactic features. Whether the patterns observed here for morphosyntax applies to different aspects of phonological development remains an empirical question that should be addressed with the investigation of a wider selection of L2 phonological phenomena.

The study inquired about the effects of type of instruction on the development of each sC structure. Based on previous studies involving morphosyntax (e.g., Zobl, 1985; Yabuki-Soh, 2007), it was hypothesized that there would be a correlation between these two linguistic fields, phonology and morphosyntax, with results favouring groups that are taught the most marked form of a given developmental hierarchy. While there are no phonological studies with which these results could be compared (except for a pilot study reported in Cardoso, 2010), the findings presented here are consistent with those observed in previous studies involving the acquisition of morphosyntactic features (e.g., the acquisition of relative clauses; Eckman et al., 1988; Roberts, 2000; Ammar and Lightbown, 2004; Yabuki-Soh, 2007).

Could it be that a piecemeal (rather than an all-at-once) approach to teaching could partially explain the results? Although not originally conceived to address phonological difficulty, one possible theoretical explanation as to why a piecemeal approach could be beneficial for sC acquisition might come from Cognitive Load Theory (Chandler and Sweller, 1991). Briefly, Chandler and Sweller propose a framework for instructional designers and teachers that allows them to control the conditions for learning by reducing extraneous cognitive load. They assume that there is a single and limited cognitive resource to acquire new knowledge, so the addition of more items to the process may limit the amount of resources available for learning. In the current study, the reduction of learnable sC structure in the P Group to one item (the most marked /st/) per teaching session might have triggered a focus toward that single unit, thus rendering the overall learning process more effective. Regardless of the reasons why learners performed better in a piecemeal instructional environment, from a pedagogical standpoint, these results suggest that gradually providing learners with the elements that constitute a given developmental sequencing could be conducive to better L2 speech production. Clearly, this topic is worthy of further investigation, particularly for developmental sequence phenomena.

A tangential but interesting finding uncovered by this research was the "task effect" observed between the *Read Aloud* and *Listen and Repeat* tests, with only the former displaying significant differences in performance (see Cardoso et al., 2009 and Saito and Plonsky, 2019 for similar findings in L2 pronunciation studies). One possible explanation for this test-related disparity may be due to the cognitive characteristics of the two tests: while *Read Aloud* relied heavily on the participants' ability to recall and orally produce the target sC forms (e.g., with the aid of grapheme-to-phoneme associations, which characterize the act of reading aloud), *Listen and Repeat* depended on the participants' ability to imitate speech, considered a less cognitively demanding activity (e.g., Alós-Ferrer and Schlag, 2009). In fact, reading aloud has been recognized as a by-product of the development of accuracy and automaticity in reading, a cognitively demanding task that includes both phonological and orthographic processes to access, process and orally produce written forms (Hudson et al., 2008). Another possible explanation relates to the level of attention paid to speech, a concept that is often operationalized as level of formality or style. As has been attested in the sociolinguistic literature (e.g., Major, 2004), more formal tasks (styles) such as reading aloud have greater propensity for target-like forms in comparison with less formal activities such as imitating one's speech. Despite these task effects, the results obtained in *Listen and Repeat* confirm that a focus on the most marked /st/ is the most effective way of teaching sC clusters, as the group that was taught exclusively this form was still able to project the acquired knowledge to other clusters, even without any exposure to them, in both tests.

Finally, this study indirectly investigated the effects of sonority within the sC cluster based on the markedness relations motivated by the principles of Sonority Sequencing and Minimal Sonority Distance. Findings related to this question, observed at the last stage of sC development covered by this study (delayed posttest), revealed that, *a priori*, the development of sC clusters across the three experimental groups does not fully conform to what is predicted by sonority and its markedness effects and some of the current literature on sC acquisition. Consider, for example, the results for the *Read Aloud* test, an activity that we claim to represent aspects of the participants' phonological knowledge (i.e., in comparison with imitative *Listen & Repeat*) because it involves the processing of grapheme-to-phoneme rules and it does not rely on the participants' ability to mimic speech (see preceding paragraph). In this task, the most marked /st/ was found to be the more easily acquired in two of the experimental groups: Groups P and T, but not M. These findings seem to corroborate Cristófaro-Silva and Freitas (2020) analysis for Mineiro, a regional BP variety; they found that two phonetic representation co-exist for sC, depending on whether the word is native to Portuguese (is.C) or a borrowing (sC) (see also Collischonn and Schwindt, 2005 for an analysis in which the underlying representation for fricative plus consonant sequences do not contain a vowel).

However, we would exert caution when interpreting these results. First, this study was not designed to examine the acquisition of sC onsets (see Cardoso and Liakin, 2009 for a similar study conducted in a naturalistic, ecologically valid setting). Instead, it was designed to investigate the effects of three types of instruction on the development of these clusters, using a highly controlled (and consequently unnatural) methodology. Second, the Groups that displayed higher rates of /st/ production were the ones in which the participants received instruction on /st/ last, possibly indicating an immediate post-treatment effect. Finally, the only group that behaved in a predictable manner was Group M, in which the least marked /sl/ cluster was more prevalent. Interestingly, this is the group that most resembles a natural learning environment in which the three sC forms are taught in tandem. The results pertaining to a developmental sequence between /sl/ and /sn/, as predicted by Clements' (1990) Minimal Sonority Distance (MSD) and some of the previous studies (e.g., Carlisle, 2006), was not borne out, since the difference in performance between the two clusters was not significant across the three instructional groups. This suggests that although the L2 learners who participated in this study are sensitive to sonority sequencing and its markedness effects on syllabifying foreign sC clusters, they remain oblivious to other principles such as the MSD and, consequently, they process these clusters in a bipartite way in which /s/ + sonorants (/sl/ and /sn/) pattern together as a set in opposition to the most marked /st/ cluster. In a study conducted in a natural (i.e., not MLS-based) language learning classroom setting, Cardoso and Liakin (2009) examined the acquisition of English sC by the same speech community of L2 learners investigated in this study and found identical patterns of acquisition. Not being an idiosyncrasy of BP L1 speakers, similar patterns have also been found for other language backgrounds in both L1 (e.g., Yavaş and Barlow, 2006; Hefter and Cardoso, 2010) and L2 acquisition (e.g., Rauber, 2006; Boudaoud, 2008).

To conclude, although sonority-based markedness is a good predictor of the order in which sC clusters are acquired, this study has also shown that instruction can somehow alter patterns of acquisition, as has been shown by the works of Ammar and Lightbown (2004), Eckman et al. (1988), Roberts (2000), and Yabuki-Soh (2007). Contra Felix (1981), this study has also shown that formal instruction followed by explicit feedback can indeed minimize the impact of the processes that "constitute man's natural ability to acquire language" (p. 87).

CONCLUDING REMARKS

This study's goal was to examine the effects of three types of instruction on the development of foreign sC clusters to determine which is more pedagogically effective. The results have shown that instruction plays an important role in the acquisition of sC clusters. Specifically, our findings show that a piecemeal introduction to novel sC forms may lead to better oral performance, particularly if the introduction starts from the more difficult and marked end of the developmental hierarchy. As such, this study constitutes the empirical evidence needed to substantiate Eckman and Iverson (1997) and Doughty and Williams (1999) pedagogical recommendations that problematic L2 sounds be taught first in the most difficult and latest acquired (marked) environment, as discussed at the outset of this paper.

Although this study has revealed some interesting findings about the effects of instruction on the teaching of a phonological developmental sequence (sC), it has uncovered some limitations that need to be acknowledged and addressed in future research. Primarily, among its major shortcomings is the adoption of an artificial language, Taki, to test the hypotheses regarding the teaching of developmental sequences. As indicated earlier, while this allowed us to tightly control several confounding aspects commonly found in the standard language classroom (e.g., the students' previous experience with the target form, the quantity and quality of the L2 input, the quality of the pedagogical intervention), one can certainly question its ecological validity: The teaching environment simulated in this study reflects only a fraction of the richness that characterize the L2 classroom reality. Accordingly, we must exercise caution in generalizing our findings because, despite being high on reliability (Hulstijn, 1997), studies with low ecological validity cannot be generalized beyond the settings where they were carried out. Despite this limitation inherent to the laboratory approach adopted, we hope that future research will address similar questions in a more ecologically valid environment, with a natural target language and in authentic language learning settings.

Another important limitation is that the participants in the Teachability Group were not tested at different stages of the experiment to confirm if they had indeed acquired the sC cluster targeted in each instructional session. As discussed earlier, the premise of the hypothesis represented by this group presupposes that a novel linguistic form can only be acquired when learners are ready for the next developmental stage (Pienemann, 1998; Pienemann et al., 2005). Without clear evidence that learning took place at each instructional stage, it is difficult to confirm whether the learners were in fact developmentally ready to learn a more advanced structure, particularly for sC items that were introduced early in the experiment (i.e., /sl/ and /sn/). However, based on the immediate posttest results obtained for the last item introduced to this group (i.e., /st/), we are confident that some learning took place during the teaching sessions. For instance, in both *Read Aloud* and *Listen and Repeat* tests, the participants significantly improved in /st/ production from pretest to immediate posttest from 0.66 to 2.24 (*Read Aloud*) and 1.63 to 3.08 (*Listen and Repeat*), respectively. The issue remains a critical limitation of this study, which we plan to address in future research.

Certain aspects of the research design also limit the implications and generalizability of the results reported. First, due to time constraints and the participants' fatigue after having already participated in three tests (in addition to surveys and interviews), only the effects of the "short-delayed" posttest were investigated, which did not allow us to observe long-term effects of the different types of pedagogical interventions adopted in the study. While the adoption of a "short-delayed" posttest was included as a compromise to slightly delay the posttest and at the same time reduce the number of tests, the literature provides strong evidence that L2 instruction decrease gradually between immediate and delayed posttest (Norris and Ortega, 2000; but see Mackey and Goo, 2007 and Lyster and Saito, 2010 for contradictory findings). Secondly, the pedagogical interventions adopted relied exclusively on the teaching of a small set of phonological features (sC) in a vocabulary acquisition context. A natural question that arises from such a limited pronunciation focus is whether learners would have performed differently had they been provided with opportunities to transfer the newly acquired knowledge to high-level communicative tasks as spontaneous speech (see Couper, 2006 and Derwing and Rosita, 2003 for evidence of how phonological gains obtained *via* instruction can be transferable to other lexical environments or spontaneous speech).

Despite these limitations, some pedagogical implications can be derived from our findings. Assuming that they can be extrapolated to "real-life" teaching, the most obvious recommendation would be for L2 teachers of sC languages such as English, French, and German to start instruction from the hard-to-acquire end of the developmental hierarchy, /st/ (possibly including other /s/ plus stop clusters such as /sp/ and /sk/). Due to the frequency distribution of sC forms in these languages, this is a relatively easy move, considering that /st/ comprises the vast majority of sC forms in most languages. In English, for instance, /st/ constitutes 87.4% of all sC forms in student-directed teacher speech and 87.9% in the Brown Corpus (Cardoso and Liakin, 2009). Following Celce-Murcia et al. (2010) framework for teaching pronunciation, instructors could initially engage students in sound awareness and listening discrimination activities so that learners can hear and discriminate the differences between the L1-influenced form (e.g., *[i]stop, *s[u] top) and the target /st/ (e.g., /st/op). This could be followed by controlled practice in which there is a focus on sC articulation accompanied by teacher-or peer-based explicit feedback (e.g., using tongue twisters such as "Stella, stop selling stocks," or the exaggerated/lengthened pronunciation of /s/ so that it sounds like a separate syllable, taking advantage of this segment's continuant characteristic: /s:.t/op-see Cardoso and Liakin, 2009 for the rationale behind this recommendation). Finally, students could practice the newly-acquired forms in spontaneous, less controlled activities such as preparing and orally presenting a set of suggestions on how to succeed as a language student (e.g., /st/udy every day, /st/ay cool when you make mistakes). Based on the findings reported here, following these recommendations, the effects of an instructional focus on the hard-to-acquire /st/ will likely project to the forms that are assumed to be more easily acquired, /sn/ and /sl/.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Concordia University's Ethics Committee (Office of Research). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

We would like to thank the people who contributed to the successful completion of this project. First, lots of *obrigados* go to the two schools where the research took place (in Belém, Brazil) and more importantly, their directors: Sra. Miriam Augusta de Oliveira e Sr Augusto Trindade (Centro de Educação de Jovens e Adultos-CEEJA Belém), and Sra. Helena Cristina Moura (Escola Municipal Manuela Freitas). The next round of *obrigados* go to our amazing team of research assistants, particularly Carol Johnson, Danial Mehdipour-Kolour, Anne-Marie Sénécal, George Smith, and Ross Sundberg; and to our statistician, Randall Halter. We would also like to extend our gratitude to John Archibald and the three reviewers for their constructive feedback and suggestions. Finally, we acknowledge financial support from the Social Sciences and Humanities Research Council of Canada (Cardoso and Collins; SSHRC 435-2012-1668).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.662934/full#supplementary-material

REFERENCES

Adam, G., and Bat-El, O. (2009). When Do Universal Preferences Emerge in Language Development? the Acquisition of Hebrew Stress. *Brill's Annu. Afroasiatic Languages Linguistics* 1 (1), 255–282. doi:10.1163/187666309x12491131130468

Allen, G., and Hawkins, S. (1978). "The Development of Phonology Rhythm," in *Syllables and Segments*. Editors A. Bell and J. Hooper (North-Holland Publishing Company), 173–185.

Alós-Ferrer, C., and Schlag, K. (2009). "Imitation and Learning," in *The Handbook of Rational and Social Choice*. Editors P. Anand, P. Pattanaik, and C. Puppe (Oxford University Press).

Ammar, A., and Lightbown, P. (2004). "Teaching Marked Linguistic Structures: More about the Acquisition of Relative Clauses by Arab Learners of English," in *Investigations in Instructed Second Language Learning*. Editors A. Housen and M. Pierrard (Berlin: Mouton de Gruyter), 167–198.

Archibald, J., and Libben, G. (1995). *Research Perspectives on Second Language Acquisition*. Mississauga, ON: Copp Clark.

Bailey, N., Madden, C., and Krashen, S. D. (1974). Is There a "Natural Sequence" in Adult Second Language Learning?. *Lang. Learn.* 24 (2), 235–243. doi:10.1111/j.1467-1770.1974.tb00505.x

Bardovi-Harlig, K. (1995). "The Interaction of Pedagogy and Natural Sequences in the Acquisition of Tense and Aspect," in *Second Language Acquisition Theory and Pedagogy*. Editors F. Eckman, D. Highland, P. Lee, J. Mileham, and R. Weber (Lawrence Erlbaum), 151–168.

Bernhardt, B., and Stemberger, J. (1998). *Handbook of Phonological Development: From the Perspective of Constraint-Based Nonlinear Phonology*. Academic Press.

Bettoni, M., and Koerich, R. (2009). "Perceptual Training in the Pronunciation of/s/-Clusters in Brazilian Portuguese/English Interphonology," in *Recent Research in Second Language Phonetics/phonology: Perception and Production*. Editors B. Baptista, A. Rauber, and M. Watkins (Cambridge Scholars), 154–173.

Boersma, P., and Weenink, D. (2019). *Praat: Doing Phonetics by Computer (Version 6.0.56)*. [Computer program]. Retrieved from http://www.praat.org.

Boudaoud, M. (2008). The Variable Development of/s/ + Consonant Onset Clusters in the Interlanguage Of Farsi ESL Learners *[Unpublished Master's Thesis]*. Montreal: Concordia University. doi:10.1109/noms.2008.4575271

Broselow, E., Chen, S.-I., and Wang, C. (1998). The Emergence of the Unmarked in Second Language Phonology. *Stud. Second Lang. Acquis* 20 (2), 261–280. doi:10.1017/s0272263198002071

Brown, J. (1983). "An Exploration of Morpheme-Group Interactions," in *Second Language Acquisition Studies*. Editors K. Bailey, M. Long, and S. Peck (Rowley, MA: Newbury House), 25–40.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge University Press. doi:10.1017/cbo9780511612886

Cardoso, W., John, P., and French, L. (2009). "The Perception of sC Onset Clusters in Second Language Phonology: A Variationist Perspective," in *Recent Research in Second Language Phonetics/phonology: Perception and Production*. Editors B. Baptista, A. Rauber, and M. Watkins (Cambridge Scholars), 203–233.

Cardoso, W., and Liakin, D. (2009). "When Input Frequency Patterns Fail to Drive Learning: Evidence from Brazilian Portuguese English," in *Recent Research in Second Language Phonetics/phonology: Perception and Production*. Editors B. Baptista, A. Rauber, and M. Watkins (Cambridge Scholars), 174–202.

Cardoso, W. (2010). "Teaching Foreign sC Onset Clusters," in *Achievements and Perspectives in Second Language Acquisition of Speech*. Editors K. Dziubalska-Kolaczyk, M. Wrembel, and M. Kul (Peter Lang), 29–40.

Cardoso, W. (2007). The Variable Development of English Word-Final Stops by Brazilian Portuguese Speakers: A Stochastic Optimality Theoretic Account. *Lang. Variation Change* 19 (3), 218–249. doi:10.1017/s0954394507000142

Carlisle, R. S. (1991). The Influence of Environment on Vowel Epenthesis in Spanish/English Interphonology. *Appl. Linguistics* 12 (1), 76–95. doi:10.1093/applin/12.1.76

Carlisle, R. S. (2006). "The Sonority Cycle and the Acquisition of Complex Onsets," in *English with a Latin Beat – Studies in Portuguese/Spanish English Interphonology*. Editors B. Baptista and M. Watkins (Johns Benjamins), 105–137. doi:10.1075/sibil.31.09car

Celce-Murcia, M., Brinton, D., Goodwin, J., and Griner, B. (2010). *Teaching Pronunciation: A Course Book and Reference Guide*. 2nd ed. Cambridge University Press.

Chandler, P., and Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cogn. Instruction* 8 (4), 293–332. doi:10.1207/s1532690xci0804_2

Chomsky, N. (1970). *Current Issues in Linguistic Theory*. Berlin: Mouton de Gruyter.

Collins, L., Trofimovich, P., White, J., Cardoso, W., and Horst, M. (2009). Some Input on the Easy/difficult Grammar Question: An Empirical Question. *Mod. Lang. J.* 99 (2), 336–353. doi:10.1111/j.1540-4781.2009.00894.x

Collischonn, G., and Schwindt, L. (2005). Considerações sobre a sequencia/sC/inicial em Português Brasileiro. *Lingua(gem)* 2 (2), 249–266.

Cook, V. J. (1988). Language Learners' Extrapolation of Word Order in Micro-artificial Languages*. *Lang. Learn.* 38 (4), 497–529. doi:10.1111/j.1467-1770.1988.tb00165.x

Corder, S. (1967). The Significance of Learners' Errors. *Int. Rev. Appl. Linguistics* 5 (4), 161–170. doi:10.1515/iral.1967.5.1-4.161

Couper, G. (2006). The Short and Long-Term Effects of Pronunciation Instruction. *Prospect* 21 (1), 46–66.

Cristófaro Silva, T., and Freitas, M. (2020). sC-Clusters in Brazilian Portuguese. *J. Portuguese Linguistics* 19 (14), 1–17. doi:10.5334/jpl.231

Davis, B. L., and MacNeilage, P. F. (1990). Acquisition of Correct Vowel Production. *J. Speech Lang. Hear. Res.* 33 (1), 16–27. doi:10.1044/jshr.3301.16

de Groot, A. M. B., and Keijzer, R. (2000). What Is Hard to Learn Is Easy to Forget: The Roles of Word Concreteness, Cognate Status, and Word Frequency in Foreign-Language Vocabulary Learning and Forgetting. *Lang. Learn.* 50 (1), 1–56. doi:10.1111/0023-8333.00110

de Lacy, P. (2006). *Markedness: Reduction and Preservation in Phonology*. Cambridge University Press. doi:10.1017/cbo9780511486388

de Villier, J., and de Villiers, P. (1977). A Cross-Sectional Study of the Acquisition of Grammatical Morphemes in Child Speech. *J. Psycholinguistic Res.* 2, 267–278.

DeKeyser, R. M. (1995). Learning Second Language Grammar Rules. *Stud. Sec Lang. Acq* 17 (3), 379–410. doi:10.1017/s027226310001425x

Derwing, T., and Rossiter, M. (2003). The Effects of Pronunciation Instruction on the Accuracy, Fluency, and Complexity of L2 Accented Speech. *Appl. Lang. Learn.* 13 (1), 1–17.

Dinnsen, D., and Elbert, M. (1984). “On the Relationship between Phonology and Learning,” in *Phonological Theory and the Misarticulating Child*. Editors M. Elbert, D. A. Dinnsen, and G. Weismer (American Speech-Language-Hearing Association monographs), 59–68.

Doughty, C. (1988). *The Effect of Instruction on the Acquisition of Relativization in English as a Second Language*. Department of Educational Linguistics, University of Pennsylvania. [Unpublished doctoral dissertation].

Doughty, C. (1991). Second Language Instruction Does Make a Difference. *Stud. Sec Lang. Acq* 13 (4), 431–469. doi:10.1017/s0272263100010287

Doughty, C., and Williams, J. (1999). “Pedagogical Choices in Focus on Form,” in *Focus on Form in Classroom Second Language Acquisition*. Editors C. Doughty and J. Williams (Cambridge University Press), 197–262.

Dulay, H., and Burt, M. (1978). “Some Remarks on Creativity in Language Acquisition,” in *Second Language Acquisition Research: Issues and Implications*. Editor W. Ritchie (Academic Press), 65–89.

Eckman, F., and Iverson, G. (1997). “Structure Preservation in Interlanguage Phonology,” in *Language Acquisition & Language Disorders: Focus on Phonological Acquisition*. Editors S. Hannahs and M. Young-Scholten (John Benjamins), 143–164.

Eckman, F. R., Bell, L., and Nelson, D. (1988). On the Generalization of Relative Clause Instruction in the Acquisition of English as a Second Language1. *Appl. Linguistics* 9 (1), 1–20. doi:10.1093/applin/9.1.1

Ellis, N. C., and Schmidt, R. (1997). Morphology and Longer Distance Dependencies. *Stud. Second Lang. Acquis* 19 (2), 145–171. doi:10.1017/s0272263197002027

Ellis, R. (1984). Can Syntax Be Taught? A Study of the Effects of Formal Instruction on the Acquisition of WH Questions by Children. *Appl. Linguistics* 5 (2), 138–155. doi:10.1093/applin/5.2.138

Ellis, R. (1997). *Second Language Acquisition*. Oxford University Press.

Escartin, C. (2005). *The Development of sC Onset Clusters in Spanish English*. Montreal, Canada: Concordia University. [Unpublished master’s thesis].

Felix, S. W. (1981). The Effect of Formal Instruction on Second Language Acquisition1. *Lang. Learn.* 31 (1), 87–112. doi:10.1111/j.1467-1770.1981.tb01374.x

Flege, J. E., and Davidian, R. D. (1984). Transfer and Developmental Processes in Adult Foreign Language Speech Production. *Appl. Psycholinguistics* 5, 323–347. doi:10.1017/s014271640000521x

Freeman, D. E. L. (1975). The Acquisition of Grammatical Morphemes by Adult ESL Students. *TESOL Q.* 9 (4), 409–419. doi:10.2307/3585625

Gasparini, E. (2005). Sentidos de ensinar e aprender ingles na escolar de ensino médio e fundamental – Uma análise discursiva. *Polifonia* 10 (10), 159–175.

Gass, S. (1982). “From Theory to Practice,” in *On TESOL ’81: Selected Papers from the Fifteenth Annual Conference of Teachers of English to Speakers of Other Languages*. Editors M. Hines and W. Rutherford (Washington, DC: TESOL), 129–139.

Gass, S. (1997). *Input, Interaction, and the Second Language Learner*. Erlbaum.

Gerrits, E., and Zumach, A. (2006). The Acquisition of #sC-Clusters in Dutch. *J. Multilingual Commun. Disord.* 4 (3), 218–230. doi:10.1080/14769670601110549

Gierut, J. A., and Champion, A. H. (2001). Syllable Onsets II. *J. Speech Lang. Hear. Res.* 44 (4), 886–904. doi:10.1044/1092-4388(2001/071)

Gierut, J. A. (1999). Syllable Onsets. *J. Speech Lang. Hear. Res.* 42 (3), 708–726. doi:10.1044/jslhr.4203.708

Hefter, H., and Cardoso, W. (2010). *The L1 Acquisition of sC Onset Clusters: Comparing The Effects of Markedness and Input Frequency*. Toronto, ON: Generative Approaches to Language Acquisition in North America (GALANA) Conference. [Paper presentation].

Hudson, R. F., Pullen, P. C., Lane, H. B., and Torgesen, J. K. (2008). The Complex Nature of reading Fluency: A Multidimensional View. *Reading Writing Q.* 25 (1), 4–32. doi:10.1080/10573560802491208

Hulstijn, J. H. (1997). Second Language Acquisition Research in the Laboratory. *Stud. Second Lang. Acquis* 19 (2), 131–143. doi:10.1017/s0272263197002015

Izidro, R. (2007). O ensino de língua inglesa nas escolas públicas contribui para uma transformação social?. *Recanto das Letras*, 412944. http://www.recantodasletras.com.br/artigos/412944.

Jakobson, R. (1968). *Child Language, Aphasia, and Phonological Universals*. Mouton. doi:10.1515/9783111353562

Jensen, C., Contantinides, J., and Hanson, K. (2003). Natural Order of Acquisition of German Syntactic Structures vs. Order of Presentation in Elementary German Textbooks in the U.S. *Teach. German* 16 (2), 199–211.

Krashen, S. (1981). The “Fundamental Pedagogical Principle” in Second Language Teaching. *Studia Linguistica* 35 (1-2), 50–70. doi:10.1111/j.1467-9582.1981.tb00701.x

Kwon, E.-Y. (2005). The "natural Order" of Morpheme Acquisition: A Historical Survey and Discussion of the Putative Determinants. *Columbia Univ. Working Pap. TESOL Appl. Linguistics* 5 (1), 1–21. doi:10.5089/9781451862041.001

Lightbown, P. (1980). “The Acquisition and Use of Questions by French L2 Learners,” in *Second Language Development: Trends and Issues*. Editor S. Felix (New York: Gunter Narr), 151–175.

Lightbown, P. (1998). “The Importance of Timing in Focus on Form,” in *Focus on Form in Classroom SLA*. Editors C. Doughty and J. Williams (Cambridge University Press), 177–196.

Lima, L., Souza, S., and Luquetti, E. (2014). O ensino da habilidade oral da língua inglesa nas escolas públicas. *Caderno do CNLF* 18 (10), 86–103.

Lopes, L. (1996). *Oficina de lingüística aplicada*. São Paulo: Mercado de Letras.

Ludwig, J. (1984). Vocabulary Acquisition as a Function of Word Characteristics. *Can. Mod. Lang. Rev.* 40 (5), 522–562. doi:10.3138/cmlr.40.4.552

Luk, Z. P.-s., and Shirai, Y. (2009). Is the Acquisition Order of Grammatical Morphemes Impervious to L1 Knowledge? Evidence from the Acquisition of Plural -s, Articles, and Possessive â€™s. *Lang. Learn.* 59 (4), 721–754. doi:10.1111/j.1467-9922.2009.00524.x

Lyster, R., and Saito, K. (2010). Oral Feedback in Classroom Sla. *Stud. Second Lang. Acquis* 32, 265–302. doi:10.1017/s0272263109990520

Mackey, A., and Goo, J. (2007). “Interaction Research in SLA: A Meta-Analysis and Research Synthesis,” in *Conversational Interaction in Second Language Acquisition: A Collection of Empirical Studies*. Editor A. Mackey (Oxford University Press), 407–452.

MacWhinney, B. (1983). Miniature Linguistic Systems as Tests of the Use of Universal Operating Principles in Second-Language Learning by Children and Adults. *J. Psycholinguist Res.* 12 (5), 467–478. doi:10.1007/bf01068027

Major, R. C. (1996). “Markedness in Second Language Acquisition of Consonant Clusters,” in *Variation and Second Language Acquisition*. Editors R. Bayley and D. Preston (Benjamins), 75–96. doi:10.1075/sibil.10.04maj

Major, R. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Erlbaum. doi:10.4324/9781410604293

Major, R. (2004). Gender and Stylistic Variation in Second Language Phonology. *Lang. Variation Change* 16, 169–188. doi:10.1017/s0954394504163059

Meisel, J. M., Clahsen, H., and Pienemann, M. (1981). On Determining Developmental Stages in Natural Second Language Acquisition. *Stud. Sec Lang. Acq* 3 (2), 109–135. doi:10.1017/s0272263100004137

Moeser, S. D., and Bregman, A. S. (1973). Imagery and Language Acquisition. *J. Verbal Learn. Verbal Behav.* 12 (1), 91–98. doi:10.1016/s0022-5371(73)80064-7

Norris, J. M., and Ortega, L. (2000). Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-Analysis. *Lang. Learn.* 50 (3), 417–528. doi:10.1111/0023-8333.00136

Pennington, M. (1999). Computer-aided Pronunciation Pedagogy: Promise, Limitations, Directions. *Comp. Assist. Lang. Learn.* 12 (50), 427–440. doi:10.1076/call.12.5.427.5693

Pienemann, M., Di Biase, B., and Kawaguchi, S. (2005). “7. Extending Processability Theory,” in *Cross-linguistic Aspects of Processability Theory*. Editor M. Pienemann (Benjamins), 199–251. doi:10.1075/sibil.30.09pie

Pienemann, M. (1989). Is Language Teachable? Psycholinguistic Experiments and Hypotheses. *Appl. Linguistics* 10 (1), 52–79. doi:10.1093/applin/10.1.52

Pienemann, M., Johnston, M., and Brindley, G. (1988). Constructing an Acquisition-Based Procedure for Second Language Assessment. *Stud. Sec Lang. Acq* 10 (2), 217–243. doi:10.1017/s0272263100007324

Pienemann, M. (1998). *Language Processing and Second Language Development: Processability Theory*. Benjamins. doi:10.1075/sibil.15

Pienemann, M. (2007). “Processability Theory,” in *Theories in Second Language Acquisition – an Introduction*. Editors B. VanPatten and J. Williams (Benjamins), 137–154.

Pienemann, M. (1984). Psychological Constraints on the Teachability of Languages. *Stud. Sec Lang. Acq* 6 (2), 186–214. doi:10.1017/s0272263100005015

Prince, A., and Smolensky, P. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Cambridge: Blackwell.

Roberts, M. (2000). *Implicational Markedness and the Acquisition of Relativization by Adult Learners of Japanese as a Foreign Language*. Honolulu: University of Hawai'I at Manoa. [Unpublished doctoral dissertation].

Rosansky, E. J. (1976). Methods and Morphemes in Second Language Acquisition Research1. *Lang. Learn.* 26 (2), 409–425. doi:10.1111/j.1467-1770.1976.tb00284.x

Rose, Y., and Champdoizeau, C. (2007). Debunking the Trochaic Bias Myth: Evidence from Phonological Development. *Bls* 33 (1), 323–334. doi:10.3765/bls.v33i1.3537

Saito, K., and Plonsky, L. (2019). Effects of Second Language Pronunciation Teaching Revisited: A Proposed Measurement Framework and Meta-Analysis. *Lang. Learn.* 69 (2), 652–708. doi:10.1111/lang.12345

Shirai, Y. (1997). "Linguistic Theory & Research: Implications for Second Language Teaching," in *The Encyclopedia of Language and Education: Second Language Education*. Editors G. Tucker and D. Corson (Kluwer Academic), 1–9. doi:10.1007/978-94-011-4419-3_1

Smith, B. L., and Stoel-Gammon, C. (1983). A Longitudinal Study of the Development of Stop Consonant Production in Normal and Down's Syndrome Children. *J. Speech Hear. Disord.* 48 (2), 114–118. doi:10.1044/jshd.4802.114

Smith, N. (1973). *The Acquisition of Phonology: A Case Study*. Cambridge University Press.

Spada, N., and Lightbown, P. M. (1999). Instruction, First Language Influence, and Developmental Readiness in Second Language Acquisition. *Mod. Lang. J* 83 (1), 1–22. doi:10.1111/0026-7902.00002

Spada, N., and Tomita, Y. (2010). Interactions between Type of Instruction and Type of Language Feature: A Meta-Analysis. *Lang. Learn.* 60 (2), 263–308. doi:10.1111/j.1467-9922.2010.00562.x

Thornbury, S. (2002). *How to Teach Vocabulary*. Essex: Longman.

Trofimovich, P., Gatbonton, E., and Segalowitz, N. (2007). A Dynamic Look at L2 Phonological Learning: Seeking Psycholinguistic Explanations for Implicational Phenomena. *Stud. Second Lang. Acquisition* 29 (3), 407–448. doi:10.1017/s027226310707026x

VanPatten, B. (1990). Attending to Form and Content in the Input. *Stud. Sec Lang. Acq* 12 (3), 287–301. doi:10.1017/s0272263100009177

Vihman, M. (1996). *Phonological Development: The Origins of Language in the Child*. Cambridge, MA: Blackwell.

Wode, H. (1976). Developmental Sequences in Naturalistic L2 Acquisition. *Working Pap. Bilingualism* 2, 1–13.

Yabuki-Soh, N. (2007). Teaching Relative Clauses in Japanese. *Stud. Second Lang. Acquisition* 29 (2), 219–252. doi:10.1017/s027226310707012x

Yavaş, M., and Barlow, J. (2006). Acquisition of #sC Clusters in Spanish-English Bilingual Children. *J. Multilingual Commun. Disord.* 4 (3), 182–193.

Zobl, H. (1985). "Grammars in Search of Input and Intake," in *Input in Second Language Acquisition*. Editors S. Gass and C. Madden (Rowley, MA: Newbury House), 329–344.

Zobl, H. (1983). Markedness and the Projection Problem. *Lang. Learn.* 33 (3), 293–313. doi:10.1111/j.1467-1770.1983.tb00543.x

Conflict of Interest: 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.

frontiers in Communication

ORIGINAL RESEARCH
published: 10 June 2021
doi: 10.3389/fcomm.2021.660917

When the Easy Becomes Difficult: Factors Affecting the Acquisition of the English /iː/-/ɪ/ Contrast

Juli Cebrian[1]*, *Celia Gorba*[1] *and Núria Gavaldà*[2]

[1]*Departament de Filologia Anglesa i Germanística, Facultat de Filosofia i Lletres, Universitat Autònoma de Barcelona, Barcelona, Spain,* [2]*Facultad de Educación, Universidad Internacional de la Rioja, Logroño, Spain*

The degree of similarity between the sounds of a speaker's first and second language (L1 and L2) is believed to determine the likelihood of accurate perception and production of the L2 sounds. This paper explores the relationship between cross-linguistic similarity and the perception and production of a subset of English vowels, including the highly productive /iː/-/ɪ/ contrast (as in "beat" vs. "bit"), by a group of Spanish/Catalan native speakers learning English as an L2. The learners' ability to identify, discriminate and produce the English vowels accurately was contrasted with their cross-linguistic perceived similarity judgements. The results showed that L2 perception and production accuracy was not always predicted from patterns of cross-language similarity, particularly regarding the difficulty distinguishing /iː/ and /ɪ/. Possible explanations may involve the way the L2 /iː/ and /ɪ/ categories interact, the effect of non-native acoustic cue reliance, and the roles of orthography and language instruction.

Keywords: vowel contrast, L2 perception, L2 production, cross-linguistic similarity, individual variation

OPEN ACCESS

Edited by:
John Archibald,
University of Victoria, Canada

Reviewed by:
Philip J. Monahan,
University of Toronto, Canada
Anabela Rato,
University of Toronto, Canada

****Correspondence:***
Juli Cebrian
juli.cebrian@uab.cat

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *29 January 2021*
Accepted: *04 May 2021*
Published: *10 June 2021*

Citation:
Cebrian J, Gorba C and Gavaldà N (2021) When the Easy Becomes Difficult: Factors Affecting the Acquisition of the English /iː/-/ɪ/ Contrast.
Front. Commun. 6:660917.
doi: 10.3389/fcomm.2021.660917

INTRODUCTION

Learning a second or foreign language (L2) after childhood is a difficult task for a variety of reasons. In addition to learner-related factors such as age of learning and amount of first and second language use (Piske et al., 2001; Flege et al., 2003, among others), the existence of a first language (L1) phonological system in part accounts for the fact that adult L2 speakers rarely sound like native speakers (Trubetzkoy, 1939; Rochet, 1995; Strange, 1995, among others). Variability in L2 performance has been linked to the degree of similarity between first and second language sound systems (Lado, 1957; Best, 1995; Flege, 1995; Kuhl and Iverson, 1995, among others). Models of L2 speech relate the likelihood of accurate perception and production of target language phones to the degree of similarity between L1 and L2 phones. Early models of L2 speech, such as the contrastive analysis hypothesis (Lado, 1957), suggested that the closer a target language phone was to a L1 category, the easier it was to perceive and produce it accurately. Escudero's (2005) Second Language Linguistic Perception Model (L2LP) posits that L2 phones that have a similar L1 counterpart are easier to learn than new phones with no clear counterpart in the L1. The L2LP claims that the L2 is, originally, a copy of the L1, and this copy may then evolve toward more target-like values. Thus, learning a similar sound involves an adjustment of the category boundaries and their acoustic properties, whereas acquiring a new sound requires establishing a new L2 to L1 mapping prior to the process of phonetic adjustment. Other recent theories link successful L2 category creation to the ability to distinguish between L1 and L2 sound categories. Best and colleagues' Perceptual Assimilation Model (PAM, Best, 1995), and its adaptation to L2 speech learning PAM-L2 (Best and Tyler, 2007), propose that L2 learners' ability to perceive L2 contrasts depends on the degree to which L2 phones are heard as, or assimilated to, an exemplar of an L1 category. The model describes several ways in which non-native or L2 sound categories can be assimilated to L1 categories and predicts discrimination ability of non-native (PAM) or L2 phones (PAM-L2), accordingly. Thus, for instance, if two non-native phones are perceptually mapped onto two separate L1 categories (two category assimilation type), their discrimination will be easier than if the two phones are perceived as exemplars of the same L1 category (Best, 1995). In turn, if the two phones perceived as belonging to the same L1 category differ in how closely they match the L1 category (category goodness assimilation type), discrimination will be better than if the two phones are perceived as equal matches to the L1 category (single category assimilation type). Other assimilation types involve combinations of categorized and uncategorized sounds, whose discrimination accuracy may vary depending on the degree of cross-language assimilation overlap (Levy, 2009a), or perceived phonological overlap (Faris et al., 2018), that is, the overlap in the categorizations to native phones (see Tyler, 2021 for a description of all types and their predictions).

The Speech Learning Model (SLM, Flege, 1995; Flege, 2003; see Flege and Bohn, 2021 for its recently revised version SLM-r) proposes that the greater the perceived dissimilarity between L1 and L2 phones, the greater the likelihood that learners establish target-like categories. Like the PAM, the SLM proposes that the ability to discern between L1 and L2 categories is expected to improve with increased exposure to target-language input. Therefore, phones that are perceived as different from any L1 phone may eventually be categorized more accurately than those that are readily mapped onto L1 categories and are consequently perceived and produced in terms of that L1 category (Flege, 1995; Flege and Bohn, 2021). Still, some target language phones may in fact be sufficiently close to L1 counterparts that their perception and production in terms of the L1 category may be undetected by native listeners (Flege, 1992). These can be referred to as identical or near identical L1-L2 pairs and would not require a separate L2 category (e.g., Spanish and English [f]).

The prediction that target phones with no clear match in the L1 will be learned more accurately than target phones with a counterpart in the L1 is supported by some studies. Busà (1992) examined the production of English vowels by Italians living in the United States and found that most of them produced English /uː/, classified as similar to Italian /u/, with lower F2 values than native English speakers did (that is, closer to Italian /u/). By contrast, a larger number of Italians produced the new L2 vowel /ʊ/ more accurately than the similar vowel /u/. Similarly, research has found that Japanese learners of English show greater improvement in their perception and production of English /r/ than of English /l/, where the former is more dissimilar from Japanese /r/ than the latter (Bradlow et al., 1997; Aoyama et al., 2004). Regarding the effect of L2 experience on the categorization of dissimilar phones, Jun and Cowie (1994) observed that experienced Korean learners of English with a length of residence of 16–31 years in the United States produced the new vowel /ɪ/ more accurately than inexperienced learners with a length of residence of 1.3–5.3 years. Further, Bohn and Flege (1992) also found that increased experience (i.e., longer length of residence in the US) resulted in a more accurate production of the new vowel /æ/ by German speakers of L2 English, while the more similar English vowels /iː, ɪ, ɛ/ were equally accurately produced by all learners.

The studies reviewed above generally support the SLM's prediction that increased L2 experience, particularly when experience involves exposure to authentic L2 input and predominant L2 use (Flege and Bohn, 2021, SLM-r), results in more authentic categories for target sounds without a clear match in the L1. By contrast, L2 phones that are readily assimilated to L1 phones may be less accurately categorized, and may not show improvement over time. However, there are two types of cases that do not follow these predictions. First, the case of target phones that are similar to L1 phones and are nevertheless accurately produced or perceived. For instance, Fullana-Rivera and MacKay (2003) tested the effect of starting age of learning and years of foreign language instruction on Catalan/Spanish bilinguals' production of English vowels. They observed that English /iː/, with a clear counterpart in the L1, was always produced more accurately than the more dissimilar /ɪ/, although an improvement with /ɪ/ was observed with more years of learning, consistent with the SLM's expectations. Further, in a study involving 240 Italian immigrants in Canada, Munro et al. (1996) found that degree of L2 English accentedness increased with age of arrival to the target language country (AOA), which varied from 3.1 to 21.5 years. However, the English vowel /iː/, judged on the basis of an acoustic comparison to be similar to Italian /i/, was more accurately produced by the Italian L2 speakers than more dissimilar vowels like /ʌ/ and /ɝ/. The second type of outcome that does not follow from the predictions is the case of pairs of target phones involving a new and a similar phone that obtain comparable results. For example, Munro et al. (1996) also report that English /iː/ and /æ/, considered similar and dissimilar to L1 categories, respectively, were comparably accurately produced. Flege et al. (1997) tested the production and perception of L2 speakers from four L1 backgrounds and found that two groups of L1 Spanish L2 English speakers differing in length of residence in the United States (9 years vs. less than 6 months) produced the similar English vowel /ɛ/ more accurately than the new vowel /æ/ (91–99% vs. 70–73% correct identification of the intended vowel by native English listeners). In addition, their scores for English /iː/ and /ɪ/, classified on the basis of acoustic measurements as similar and new, respectively, were comparable (57–69% for /iː/ and 51–61% for /ɪ/). The question that remains is why similarly classified target phones (in terms of their similarity to L1 categories) are not always equally learned, while differently classified target phones sometimes show comparable results.

The inconsistencies found in some studies thus underscore the need for a consistent and reliable method of measuring cross-linguistic similarity. One of the most common approaches involves the use of perceptual cross-language mapping tasks (Best, 1995; Flege et al., 1997; Strange et al., 2009; Tyler et al., 2014, among many others). In these tasks, commonly referred to as categorization tasks or perceptual assimilation tasks (PAT), listeners are presented with non-native or L2 speech stimuli, and asked to 1) indicate to which L1 phonetic category each token is most similar, and 2) rate its "goodness of fit" as an exemplar of that category (see *Materials and Methods* section below). Tyler (2021) argues that while there are some limitations with these tasks, PATs are currently the most suitable method for measuring cross-linguistic similarity given that they evaluate different possible sources of information that listeners attend to when perceiving non-native sounds. This measure of perceived similarity has been found to be a good predictor of L2 learning difficulty (e.g., Best and Strange, 1992; Guion et al., 2000; Strange et al., 2011; Tyler et al., 2014, among others).

The current paper aims to explore the link between similarity relations and L2 performance further by examining the perception and production of the English /iː/-/ɪ/ contrast by a group of L1 Spanish/Catalan speakers. Previous studies have found that L1 Spanish/Catalan speakers have difficulty with the English /iː/-/ɪ/ contrast but have not been able to relate it to the degree to which each vowel is assimilated to a L1 category. Cebrian (2006) found that two groups of L2 English speakers (20 Catalan speakers living in Canada and 30 Catalan undergraduate students of English living in Spain) perceptually assimilated Canadian English /iː, ɛ, eɪ/ to Catalan /iː, ɛ, ei/, respectively, whereas English /ɪ/ obtained lower assimilation scores and goodness ratings as Catalan /e/. Perception of L2 sounds was examined with an identification test involving a synthetic continuum from /iː/ to /ɪ/ to /ɛ/ varying in vowel quality and vowel duration. The two groups performed like native speakers in their perception of English /ɛ/, but differed from native speakers in their perception of English /iː/ and /ɪ/, which showed a predominant reliance on temporal cues, unlike native Canadian English speakers who relied mostly on spectral differences. These results do not follow from the results of the PAT since two English vowels that were strongly assimilated to Catalan vowels, English /iː/ and /ɛ/, obtained different results, and the former patterned like the weakly assimilated vowel /ɪ/. The production of the target English vowels by the 30 Catalan undergraduate students of English was examined in a subsequent study (Cebrian, 2007). Data from eight native English speaker judges indicated that both English /eɪ/ and /ɛ/ were accurately produced (correctly identified 99 and 86% of the time, with goodness ratings reaching 5.5 and 5.2 out of 7, respectively). By contrast, English /iː/ and /ɪ/ obtained similarly lower rates of correct identification (73 and 71%, respectively), and /iː/ obtained higher goodness ratings than /ɪ/ (4.5 vs. 3.9) but still lower than /eɪ/ and /ɛ/. Acoustic measurements confirmed the native speaker rating data and showed that the Catalan learners produced English /ɛ/ distinctly from English /iː/ and /ɪ/ and produced the latter two with a large amount of spectral overlap in terms of first and second formant. The learners, however, produced a temporal difference between the tense and the lax vowel (/iː/: 243 ms, /ɪ/: 153 ms), resulting in a ratio (1.62) that falls within the values reported for native English speakers (Hillenbrand et al., 2000).

The inability to produce a spectral difference between English /iː/ and /ɪ/ thus seems to be linked at least in part to the fact that Catalan and Spanish learners of English tend to base the /iː/-/ɪ/ distinction on a temporal difference (e.g., Kondaurova and Francis, 2008; Cerviño and Mora, 2009). In fact, L2 English learners have been found to exploit temporal cues in their categorization of the English /iː/-/ɪ/ contrast to a greater extent than native English speakers, who appear to rely mostly on spectral cues (Hillenbrand et al., 2000; Escudero and Boersma, 2004; Cebrian, 2006). This is found with learners whose L1 has temporal contrasts (e.g., /iː/-/i/ contrast), such as Japanese and Finnish (Ylinen et al., 2009; Grenon et al., 2019), and importantly also with speakers whose L1 has no vowel duration contrast, such as Mandarin Chinese, Korean, Russian, Catalan, Portuguese and Spanish (Bohn, 1995; Flege et al., 1997; Wang and Munro, 1999; Escudero and Boersma, 2004; Cebrian, 2006; Mora and Fullana, 2007; Kondaurova and Francis, 2008; Morrison, 2008; Aliaga-García, 2011; Kivistö de Souza et al., 2017). These results lend support to Bohn's (1995) desensitization hypothesis, which claims that L2 learners may not be sensitive to L2 spectral distinctions that are not exploited in their L1, to which they have become desensitized. In these cases, duration may be used to differentiate L2 pairs with no parallel L1 spectral distinction. Further support comes from studies that show that learners exploit duration for specific contrasts only. Flege and Bohn (1989) found that Spanish learners exploited temporal cues to a greater extent than spectral cues in their identification of English /iː/ and /ɪ/, which do not have two clear L1 counterparts, but overreliance on duration was not evident with /ɛ/ and /æ/, which may be identified in terms of Spanish /e/ and /a/. Hence, duration is used to distinguish the pair of target vowels for which the L1 does not have a comparable spectral contrast. Kondaurova and Francis (2008) conclude that several factors play a role in the categorization of the /iː/-/ɪ/ contrast as a mostly temporal contrast, including the role of duration as a phonetic cue in the learners' L1 (e.g., as a cue to stress differences), experience with different target-language varieties, psychoacoustic explanations based on the desensitization to spectral differences not exploited in the L1, and a possible salience of duration and consequent ease of learning (Holt and Lotto, 2006; Hacquard et al., 2007; Kivistö-de Souza et al., 2017).

This study examines the perception and production of the English /iː/-/ɪ/ contrast by Spanish/Catalan bilinguals in relation to their perceived similarity to L1 categories. The English high front tense-lax pair /iː/-/ɪ/ has received a lot of attention in the literature both because it poses a problem to learners from different language backgrounds and also because it has a very high functional load in English (with more than 450 minimal pairs according to Higgins, 2018). The fact that a large number of these minimal pairs involve high-frequency words, and that many share the same grammatical category (e.g. feel-fill, leave-live), contribute to the high functional load of this contrast [Munro, 2021 (this volume)]. Thus, failure to distinguish these vowels can result in intelligibility problems. The perception and production of English /ɛ/ and non-rhoticized /ɜː/ (as in bed and bird in Southern British English, respectively) will also be analyzed for comparison purposes as these vowels represent a clear case of a similar and a dissimilar vowel. **Table 1** shows the assimilation scores (percent identification in terms of Spanish vowel categories) reported for English /iː, ɪ, ɛ, ɜː/ in a few previous studies involving Spanish speakers. Generally, English /iː/ and /ɛ/ seem to be strongly assimilated to an L1 category (/i/ and /e/, respectively), while /ɪ/ is more weakly assimilated to Spanish /e/. Cebrian (2019) also examined English /ɜː/, which patterned with

TABLE 1 | Percent assimilation of English /iː, ɪ, ɛ, ɜː/ to Spanish vowels reported in recent studies (Goodness of fit ratings are given in parentheses when available). SSBE, GA (General American) and Can. Eng. (Canadian English) indicate the English variety of the stimuli.

English stimuli	L1 Spanish response	Cebrian (2019),[a] SSBE	Baigorri et al. (2019),[b] GA	Morrison (2012),[c] Can. Eng.	Escudero and Chládková (2010),[d]		Flege (1991),[e] GA
					SSBE	GA	
/iː/	/i/	93 (5.9/7)	95–96 (8/9)	98–99	100	99	84–94
/ɪ/	/e/	66 (5.3/7)	50 (3–6/9)	95–96	35	90	19–39
	/i/	33 (5/7)	49	3	21	10	36–68
/ɛ/	/e/	97 (5.3/7)	87–94 (6/9)	95–96	77	96	44–81
/ɜː/	/e/	64 (3/7)					
	/a/	14 (3.4/7)					
	/o/	13 (3.9/7)					

[a] *Monolingual European Spanish speakers.*

[b] *Late and early L2 English learners.*

[c] *Monolingual European Spanish and Mexican Spanish speakers.*

[d] *Monolingual Peruvian Spanish speakers.*

[e] *L2 English speakers and monolingual Spanish speakers.*

/ɪ/ being weakly assimilated with Spanish /e/, with comparatively low goodness of fit ratings.[1]

Another reason for the lack of a consistent link between cross-linguistic perceived similarity and L2 performance may stem from the amount of individual variation potentially found in L2 performance. For instance, Munro et al. (1996) study involving 240 Italian L2 English speakers reported a lack of consistent results per vowel across L2 speakers as not all L2 speakers of similar experience produced the same vowels accurately. In a study on Cantonese speakers' production of English tense and lax high vowels, Munro (2021 (this volume)) also underscores the high degree of inter-speaker variability, which could not be related to specific linguistic (word, phonetic context) or individual (language use, length of residence) factors. Still, the majority of studies make predictions about L2 performance for a specific group of learners based on the average similarity judgements obtained for that group's performance (e.g., Cebrian, 2006; Rallo Fabra and Romero, 2012; Baigorri et al., 2019). This assumes that group tendencies are representative of individuals' perceptual judgements. Flege and Bohn's (2021) SLM-r claims that individual differences in L1 category precision may influence the discernment of cross-linguistic phonetic differences. In addition, Flege and Bohn caution that individuals may differ in the way they map L2 sounds onto L1 categories particularly in a PAT. This is because listeners' ratings may be affected by the interval between accessing long term memories of their native sounds in order to label the auditory stimulus and judging the degree of similarity between that stimulus and the selected native category. In fact, studies have shown evidence of variability among individuals sharing the same L1 in perceived similarity judgments (Strange et al., 2009; Gilichinskaya and Strange, 2010; Tyler et al., 2014; Faris et al., 2018). For example, an individual analysis of the perceptual assimilation data collected by Cebrian (2019) showed that 15 out of the 29 Spanish speakers mapped English /ɪ/ to Spanish /e/ 75% of the time or more, eight participants chose Spanish /i/ as the closest L1 vowel at a similar rate, and six participants divided their responses between Spanish /i/ and /e/ (67% or less for each vowel). Thus, a fair amount of variability was found among speakers of the same L1 background.

Therefore, another goal of the current study is to investigate the relationship between perceived similarity and likelihood of accurate L2 category creation at the individual level. The cross-linguistic perceived similarity judgements of each individual are compared to their ability to produce and perceive the two vowels distinctly. The study thus examines if previous findings that English /iː/ and /ɪ/ are perceived and produced with similar degrees of accuracy, despite differing in the extent to which they are mapped onto L1 vowels, are related to individual differences in cross-linguistic perception, or to other factors (e.g., the relevance of different acoustic cues, the role of language instruction or the influence of orthography). The focus of this paper is the English high front tense-lax vowel contrast (/iː/-/ɪ/), and two additional vowels are also examined, namely a vowel often reported to be strongly assimilated to an L1 vowel (English /ɛ/) and one that lacks a consistent counterpart in the L1 (English /ɜː/) in order to understand what is specific about the /iː/-/ɪ/ contrast and what is determined by perceived similarity relationships.

MATERIALS AND METHODS

A group of L2 English speakers performed a series of tests including a perceptual assimilation task, a vowel identification task, a vowel discrimination task and a picture naming production task. The data here presented are part of a larger study evaluating the perception and production of a number of English vowels. This paper focuses on the English /iː/-/ɪ/ and /ɛ/-/ɜː/ contrasts.

[1] Similar results have been obtained for similarity measurements involving Catalan and English despite differences in vowel inventory between the two languages (Catalan has seven vowel phonemes (/i e ɛ a ɔ o u/, plus [ə] in unstressed position) while Spanish has five (/i e a o u/). For instance, Cebrian (2006) found that English /iː/ was assimilated to Catalan /i/ (99%, GR: 6.2/7), while /ɪ/ was split between Catalan /e/ (66%, GR: 3.5), /ɛ/ (20%, GR: 3/7) and /i/ (14%, GR: 2.5). See also Cebrian (2021).

Participants

The participants were 43 Catalan/Spanish bilinguals (37 females, mean age 19.2, standard deviation 1 year) who were first-year English Studies undergraduate students at Universitat Autònoma de Barcelona (UAB), Barcelona, Spain. They reported having started learning English in school before the age of 12 (the average starting age of learning was six; average length of learning 13 years), and were at the time attending classes in English at university between 6 and 15 h per week. Most participants reported speaking English at university (with classmates, foreign students, and teachers) at least half the time, although they used English less often with family, friends and at work. The majority had not lived in an English-speaking country and those who had, had spent less than 6 months (an average of 2 weeks), generally as part of a summer stay. Participants reported having no hearing impairments. All 43 participants performed the three perception tasks, but five participants did not complete the production task and, thus, only the production of the remaining 38 participants was analyzed (33 females, mean age 19.4, standard deviation 0.9 years). The participants were all native speakers of Spanish although some spoke Catalan as well. Participants completed an online personal and language background questionnaire based on the Bilingual Language Profile (BLP; Birdsong et al., 2012). Among other information, the BLP renders a score that ranges from 218 to −218 representing the two monolingual endpoints, with values close to zero indicating balanced bilingualism. According to the questionnaire, 22 participants were Spanish-dominant, 11 appeared to be balanced in Spanish and Catalan, and 10 were more dominant in Catalan. This difference among participants need not be problematic as one of the goals of this paper is precisely to examine how individual variation in cross-linguistic perception may affect L2 performance. Nevertheless, an analysis of the Catalan-dominant bilinguals' results in the PAT revealed that their responses for the vowels under study did not differ significantly from the remaining Spanish speakers' performance (see *Perceptual Assimilation Task*). All tests took place at the Speech Laboratory at UAB on two different days. Participants performed the production task and the perceptual assimilation task in one session and the identification and discrimination tasks in a second session.

Cross-Language Perception

Stimuli

The stimuli used in the perceptual assimilation task (PAT) were a subset of the stimuli used in a previous study (Cebrian, 2019) and consisted of the Standard Southern British English (SSBE) vowels /iː, ɪ, ɛ, eɪ, aɪ, æ, ɑː, ɜː, ʌ/ produced in /bVt/ words.[2] These words were elicited from three monolingual male speakers of SSBE (mean age 35), who had spent most of their lives in the South of England and were living in London. The recordings were carried out in a soundproof booth at the speech laboratory at University College London, in London, United Kingdom, and were digitized at a 44.1 kHz-sampling rate and normalized for intensity [70 dB in sound pressure level (SPL)]. The best tokens per talker were selected, based on auditory judgements and spectrographic analysis. Neighboring sounds have been found to affect perceptual assimilation judgements (e.g., Bohn and Steinlen, 2003; Levy, 2009b). Hence stimuli were edited to include from the release of the /b/ until the closure of /t/, thus eliminating potential cross-linguistic and individual differences in stop production (prevoicing of /b/, /t/ release), while maintaining intact the cues to the vowel.

Perceptual Assimilation Task

In the perceptual assimilation tasks, listeners were required to identify each English vowel stimulus in terms of one of several possible L1 categories by clicking on one of the response options presented on a computer screen. Upon selecting an L1 category, listeners provided a goodness of fit rating on a 7-point scale, where 1 meant a poor example of the selected vowel and 7 meant a good, native-like example. The response options consisted of the most common and unequivocal spelling for each Spanish vowel and diphthong (<i, e, a, o, u, ai, ei, oi>) together with a monosyllabic word illustrating that vowel, namely di (for /i/), se (/e/), da (/a/), do (/o/), tu (/u/), hay (/ai̯/), rey (/ei̯/), hoy (/oi̯/), meaning say, self (reflexive pronoun), give, do (musical note), you, there is, and today, respectively. Every vowel appeared in 12 trials (3 talkers × 2 tokens × 2 repetitions). This paper reports the results for the vowels under study, namely /iː, ɪ, ɛ, ɜː/. See Cebrian (2019) for a study covering a near complete set of SSBE vowels.

Vowel Identification and Vowel Discrimination Tasks

Stimuli

The stimuli used in the identification and discrimination tasks consisted of high-frequency monosyllabic English words. Each word shared the initial and final consonants with at least one other word, thus resulting in a series of minimal pairs (e.g., bead, bid, bed, bird). Half the words ended in a voiced stop and the other half ended in a voiceless stop. The words were produced by two native speakers of SSBE, namely a 23 year-old female and a 33 year-old male, who were different from the speakers who produced the stimuli for the PAT. The speakers were recorded in a soundproof booth at the speech laboratory at University College London. The recordings were digitized at 44.1 kHz sampling rate and normalized for intensity (70 dB in SPL). The same stimuli were used in the identification task, presented individually, and the discrimination task, arranged in pairs (e.g., bead-bid, bead-bead), as explained below. These tasks were part of a larger study examining the perception of seven English vowels (/iː, ɪ, ɛ, æ, ɑː, ɜː, ʌ/).

Identification Task

Participants were asked to identify the vowel in the stimulus using one of seven response options appearing on a computer screen. The options consisted of a phonetic symbol together with two common words representing each sound, namely /æ/ ash, mass; /ʌ/

[2] Participants were exposed to different English varieties through music, TV and cinema, and the internet (according to the language background questionnaire, their average exposure to British and American English was 54 and 46%, respectively). Still, British English is the main variety taught in schools and language centers in Spain and it is the variety described in most textbooks used at UAB.

sun, thus; /ɪ/ fish, his; /iː/ cheese, leaf; /ɜː/ earth, first; /ɛ/ less, west; /ɑː/ arm, palm. There were eight stimuli per vowel (four words produced by two different speakers) and each stimulus appeared twice throughout the task, resulting in a total of 16 trials per vowel.

Discrimination Task

The discrimination task was a categorical AX same/different discrimination task, including several pairs of English vowels. The data presented in this paper focuses on the /iː/-/ɪ/ and /ɛ/-/ɜː/ pairs. For each pair, several stimuli were created by including the two possible speaker orders (speaker 1-speaker 2, speaker 2-speaker 1) and vowel orders (e.g., /iː/-/ɪ/ and /ɪ/-/iː/). Half the stimuli included the same vowel category (same-category trials, /iː/-/iː/), and half were different-category trials (/iː/-/ɪ/). This resulted in a total of 16 different-category trials per vowel pair, and 8 same-category trials per vowel. The interstimulus interval was 1.15 s so that participants used phonetic information stored in long term memory instead of relying on sensory memory (e.g. Højen and Flege, 2006). Participants responded by clicking on the words "same" or "different" displayed on a computer screen.

Production Task

Thirty-eight of the 43 participants who performed the PAT and the identification and discrimination tasks also completed a picture naming task in which they were asked to name 44 different pictures twice. The list included the four words containing each target vowel. Specifically, the words were bed, bet, head, pet for /ɛ/, bird, heard, hurt, dirt for /ɜː/, bid, bit, dip, lid for /ɪ/, feed, feet, league, leak, for /iː/. The words were presented in a random order, which was the same for all participants. The first four words were fillers used to familiarize participants with the task. For each word, a picture was displayed on a computer screen (MS PowerPoint was used). In order to ensure that the right word was produced, the target word was shown at the bottom of the screen in written form. The written word disappeared after two seconds, then participants were instructed to name the picture twice, prompted by the appearance of a dialogue balloon next to the picture. Participants were recorded in a soundproof chamber at the speech laboratory at UAB, Spain. In addition, the production of 13 native English speakers (seven female) was also collected using the same task to provide native English values to compare the L2 production to. These were native speakers of Southern British English in their twenties and early thirties. The group consisted of a mixture of international students and English teachers residing in Barcelona (for 2–10 years; reported weekly use of English: 75% of the time), who were recorded at the speech laboratory at UAB, and five undergraduate students at Queen Mary University, London (monolingual English speakers, three of whom had taken Spanish lessons but barely used Spanish outside the classroom), recorded in a sound attenuated room at that institution. Recordings were digitized at 44.1 kHz sampling rate and normalized for intensity (70 dB in SPL).

RESULTS AND INTERIM DISCUSSIONS

This section presents the results of the perceptual similarity task, followed by the L2 perception and L2 production results. The section ends with a correlational analysis contrasting the results of the different tasks. Each set of results is followed by a brief interim discussion.

Perceptual Assimilation Task

Perceptual Assimilation Tasks Results

The percentage of times participants identified each target vowel as one of the Spanish responses (i.e., assimilation scores) and the corresponding average goodness of fit ratings (GR) were calculated. In addition, following Guion et al. (2000), a fit index score (FI) was calculated by multiplying the identification proportion by the goodness of fit rating. This composite score is meant to distinguish between cases of comparable assimilation percentages that differ in GR, and was used by Guion et al. (2000) to determine the likelihood of accurate discrimination for pairs of L2 vowels. Further, given that individual differences in perceived similarity are often observed, the number of participants who selected a given response at a given level of consistency was tallied in order to assess the degree of agreement among listeners (Strange et al., 2009; Gilichinskaya and Strange, 2010; Cebrian, 2021). To that effect, the categorization thresholds used in some previous studies were used, i.e., 70% (Tyler et al., 2014) and 50% (Bundgaard-Nielsen et al., 2011; Faris et al., 2018). The results are presented in **Table 2**.

Recall from Section *Participants* that many of the participants spoke Catalan in addition to Spanish and participants differed in how strongly dominant in Spanish they were. An analysis was conducted to see if responses in the PAT were influenced by the bilinguals' language dominance. The results for the subgroup with a clear dominance in Spanish (n = 22) were compared to those who showed a balanced dominance (n = 11) or a more Catalan dominance (n = 10). The results for English /ɛ, iː, ɪ/ were practically identical across groups, while in the case of English /ɜː/, a greater number of Spanish /a/ responses were obtained as dominance in Catalan increased, reaching 20%. A series of Mann-Whitney U-tests yielded no significant effect of language dominance on assimilation scores, confirming that the general pattern of results was very comparable across dominance groups.

In terms of individual variation, out of the 43 participants who performed the PAT, 23 assimilated English /ɪ/ to Spanish /e/ at least 70% of the time, six learners chose Spanish /i/ as the best L1 match 70% of the time or more, and the remaining 14 learners chose either option between 25 and 67% of the time. This shows that, though the predominant pattern was to perceive English /ɪ/ as most similar to Spanish /e/, there was a certain degree of variability among participants. Responses for English /ɛ/ and /iː/ were very consistent, with all participants identifying English /ɛ/ with Spanish /e/ 70% of the time or more, and almost all participants (39) assimilating English /iː/ consistently to Spanish /i/, while four yielded responses split between Spanish /i/ and /ei/. Regarding /ɜː/, 28 participants chose Spanish /e/ as the closest vowel at least 70% of the time, three participants selected Spanish /a/, two other participants selected Spanish /o/ and the remaining 10 participants did not choose a specific L1 vowel more than 67% of the time.

TABLE 2 | Perceived similarity between nonnative (English) vowels and L1 (Spanish) vowels. (Num. of Ss assim. ≥70%/50% = number of subjects who selected a given response 70%/50% of the time or more). The total number of participants was 43.

English stimuli	L1 Spanish response	% Assimilation	Goodness Rating	Fit Index	Num. of Ss assimn. ≥70%	Num. of Ss assimn. ≥50%
/iː/	/i/	94	6.5	6.1	39	43
	/ei̯/	5.8	4.4	0.3		1
/ɪ/	/e/	67	5.9	4.0	23	32
	/i/	32	5.4	1.6	6	12
/ɛ/	/e/	99	5.3	5.2	43	43
/ɜː/	/e/	72	4.3	3.1	28	34
	/a/	15	2.8	0.4	3	4
	/o/	11	4.3	0.5	2	5

Discussion of Perceptual Assimilation Tasks Results

Assuming a categorization threshold of 70% (Tyler et al., 2014), English /iː/, /ɛ/ and /ɜː/ were categorized as Spanish /i/, /e/ and /e/, respectively. English /ɪ/ did not reach the categorization threshold of 70%, but was assimilated above chance to more than one L1 phone (Spanish /e/ and /i/) and thus illustrates an uncategorized-clustered type of assimilation (Faris et al., 2018). The current results are on the whole in agreement with the previous studies showing that Spanish speakers consistently assimilate English /iː/ and /ɛ/ to Spanish /i/ and /e/, and that English /ɪ/ displays a less consistent pattern (Escudero and Chládková, 2010; Morrison, 2012; Baigorri et al., 2019; Cebrian, 2019, see **Table 1** above). In fact, the results closely replicate those obtained by Cebrian (2019), which involved a near-complete set of SSBE and Spanish vowels and diphthongs and tested a group of Spanish speakers with little or no knowledge of English. The only difference is that English /ɜː/ fell short of a 70% categorization threshold in the previous study (64% as Spanish /e/ vs. 72% in the current study) and was thus classified as uncategorized-clustered. Setting the categorization threshold at 70% (or 50%) is an arbitrary decision, as there are no clear criteria supporting one or another cut-off (see Tyler, 2021, for discussion). In fact, in both studies English /ɪ/ obtained a higher GR as Spanish /e/ than English /ɜː/ did, indicating that the latter was perceived as a more dissimilar vowel. The lower GRs for /ɜː/ may also be related to the greater differences across English varieties regarding this vowel, which may have affected listeners' familiarity with SSBE /ɜː/.

Considering both the degree of categorization and the FI, what the data show is that English /ɛ/ and /iː/ are strongly assimilated to Spanish /e/ and /i/, while English /ɪ/ and /ɜː/ show weaker degrees of assimilation to Spanish /e/. The main focus of the current study is the English /iː/-/ɪ/ contrast. The PAT results show that the two English vowels differ notably in degree of assimilation to an L1 category (FIs of 6.1 vs. 4, respectively). The pair patterns as an uncategorized-categorized type of assimilation in PAM-L2's terms (as Spanish /i/ and /e/, respectively), with a cross-language assimilation or perceived phonological overlap (Levy, 2009a; Faris et al., 2018) of 32% as Spanish /i/. The English /ɛ/-/ɜː/ contrast patterns as a category-goodness assimilation type, with a considerable perceived phonological overlap (72%) and a notable difference in degree of assimilation to Spanish /e/, as English /ɛ/ is perceived as a better match for Spanish /e/ than English /ɜː/ (FIs: 5.3 vs. 3.1). PAM-L2 predicts that discrimination of these two types of contrasts by L2 speakers will be more difficult than between pairs that assimilate to two different non-native vowels (e.g., English /iː/ and /ɛ/) and their discrimination accuracy may depend on the degree of perceived phonological overlap. Thus, we may expect that English /iː/-/ɪ/ will be more accurately discriminated than English /ɛ/-/ɜː/.

In terms of the SLM (Flege, 1995; and its recently revised version, the SLM-r; Flege and Bohn, 2021), all four English vowels will be initially categorized in terms of the closest L1 phones. Exposure to authentic target language input may enhance the ability to differentiate the L2 from L1 sounds and thus establish separate categories for the L2 sounds. This is more likely for English /ɪ/ and /ɜː/, judged to be more different from L1 categories, than for English /ɛ/ and /iː/, which have a clear match in the L1. In fact, the latter two may pattern as near-identical to L1 categories. Flege (1992) suggests that in order to know if a non-native sound is perceived to be indistinguishable from a native category, this non-native phone should be undetectable when produced in a native context. In brief, English /iː/ and /ɛ/ pattern as highly assimilated to Spanish /i/ and /e/, while English /ɪ/ and /ɜː/ are weakly assimilated. The prediction is that the level of accuracy in the perception and production of English /iː/ and /ɛ/ by Spanish L2 speakers will be similarly high, while for /ɪ/ and /ɜː/ performance may be originally worse and accuracy will depend on the extent to which learners detect differences between L1 and L2 sounds thanks to continuous authentic input.

L2 Perception

Identification Results

The results of the identification task are presented in **Figure 1** and **Table 3**. **Figure 1** shows the median and distribution of the percent correct identification for each of the four target vowels. **Table 3** shows the average correct identification and misidentification responses (scores below 3% are omitted). As can be observed, English /ɛ/ was the most accurately identified, followed by /ɪ/, while identification was worse with /iː/ and /ɜː/. The identification results were submitted to a series of generalized linear mixed models (GLMM) with score (correctly or incorrectly identified) as the dependent variable. The best fitting model was a GLMM with Target vowel as fixed factor, and participant as

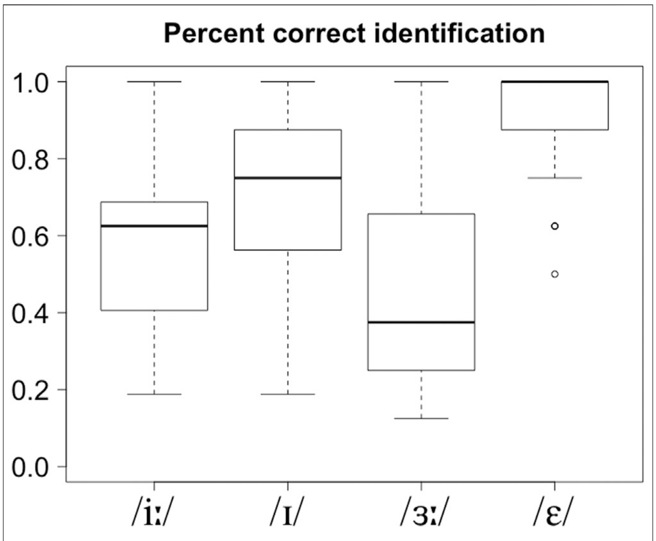


FIGURE 1 | Boxplots of identification accuracy for each of the four target English vowels. Boxplots represent the quartile ranges of the scores (top of box: 75th percentile, bottom of box; 25th percentile; line in the middle: median). The whiskers represent the highest and lowest values that are not outliers or extreme values, which are indicated by circles outside the whiskers.

TABLE 3 | Identification of the four English target vowels by 43 Spanish/Catalan learners of English. Mean correct identification percentages are highlighted in bold, misidentifications equal to or lower than 3% have been omitted.

	English target vowels			
Responses	/iː/	/ɪ/	/ɛ/	/ɜː/
/iː/	**58**	21		
/ɪ/	39	**72**		
/ɛ/		4	**90**	4
/ɜː/			7	**47**
/ʌ/				23
/ɑː/				19
/æ/				4

random intercept.[3] The results showed a significant effect of Target vowel (F (3, 2,444) = 65.3; $p < 0.001$). Bonferroni-adjusted pairwise comparisons showed that all between-vowel differences were significant ($p < 0.001$). Regarding the pattern of misidentifications, English /iː/ was most frequently misheard as /ɪ/ and vice versa, while /ɜː/ was most often misidentified as /ɑː/ or /ʌ/.

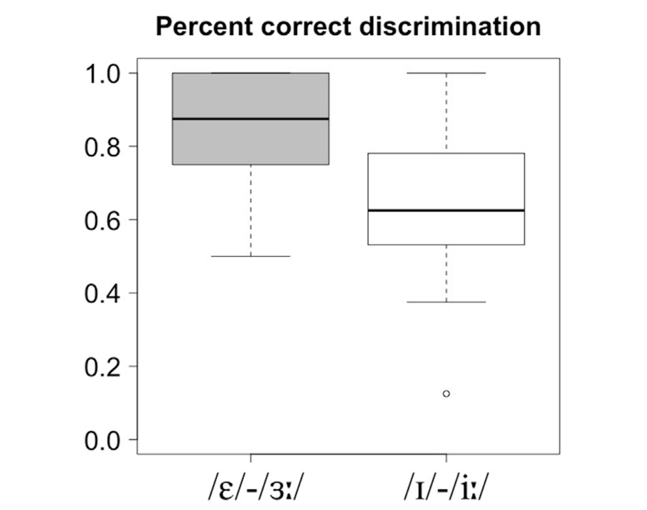


FIGURE 2 | Boxplots of discrimination accuracy for the two English vowel contrasts examined. Boxplots represent the quartile ranges of the scores (top of box: 75th percentile, bottom of box; 25th percentile; line in the middle: median). The whiskers represent the highest and lowest values that are not outliers or extreme values, which are indicated by circles outside the whiskers.

Discrimination Results

The results of the discrimination task are presented in **Figure 2**. The /ɛ/-/ɜː/ pair was discriminated more successfully than the /iː/-/ɪ/ pair, with mean percentages of correct discrimination reaching 87% for /ɛ/-/ɜː/ and 66% for /iː/-/ɪ/. The results of a GLMM with Vowel pair as fixed effect and a random slope for participant[3] confirmed that the difference between the two vowel pairs was significant (F (1, 1,030) = 38.15, $p < 0.001$).

Discussion of L2 Perception Results

The finding that /ɛ/ was very successfully identified and /ɜː/ was poorly identified can be explained in terms of the PAT results, as the former was strongly assimilated to an L1 category, while the latter barely reached a categorization threshold of 70% (99 and 72%, respectively, as Spanish /e/). Two studies testing the effect of high variability perceptual training on SSBE vowel identification by Catalan/Spanish learners of English (Carlet and Cebrian, 2019) and Spanish monolingual learners of English (Fouz-González and Mompean, 2020) also found that /ɜː/ was the most poorly identified vowel, but its identification tended to improve the most as a result of increased L2 experience (namely perceptual training), as expected for a dissimilar vowel (e.g., Flege, 1995). On the other hand, the fact that /ɪ/ was better identified than /iː/ (72 vs. 58%, respectively) cannot be equally explained by the PAT results. The expectation would be for /iː/ to pattern with /ɛ/ and be successfully identified given that both English vowels were strongly assimilated to L1 vowels. Still, better identification of /ɪ/ than of /iː/ has been reported for similar populations (Cebrian and Carlet, 2014; Carlet and Cebrian, 2019; Fouz-González and Mompean, 2020).

When misidentified, SSBE /iː/ was usually perceived as /ɪ/. The results of the PAT showed that /ɪ/ was assimilated to Spanish /i/

[3] For the statistical analyses on the L2 perception and production data, several generalized linear mixed models (GLMM) were considered, which included the independent variables in each case as the fixed effects, and different combinations of random intercepts and slopes for participant and stimulus. The selection of the best model was based on the lowest Akaike Information Criterion for a model that included participant as random intercept or random slope, given the theoretical relevance of participant variability in the current study. In fact, in every case, the results for the different models were very consistent and the levels of significance for each factor and pairwise comparison were very similar across models. IBM Corp (2017) software was used.

about a third of the time (32%), that is, there was a 32% overlap in the assimilation of English /iː/ and /ɪ/ to Spanish /i/. This overlap, and the relatively high goodness of fit ratings (GR) obtained for English /ɪ/ as Spanish /i/ (5.4/7), indicate a degree of perceptual confusion that may explain the misidentifications of English /iː/ as /ɪ/. On the other hand, the pattern of misidentifications of /ɜː/, mostly as /ɑː/ or /ʌ/ but hardly as /ɛ/, may be linked to the comparatively low GRs for /ɜː/ as Spanish /e/ (4.3). Thus, learners tended to identify as English /ɛ/ only stimuli that were clear exemplars of this vowel. In brief, the identification results are not expected from the PAT results, as the two strongly assimilated vowels (/iː/ and /ɛ/) obtained very different results, and the two weakly assimilated vowels (/ɪ/ and /ɜː/) also patterned very differently.

Regarding the discrimination task, the results seem to show more difficulty with the /iː/-/ɪ/ contrast than some previous studies, which reported discrimination accuracy rates close to or over 75% or A' scores denoting sensitivity to the contrast, i.e., over 0.6 (Mora and Fullana, 2007; Rallo Fabra and Romero, 2012; Baigorri et al., 2019). Comparisons across studies are complicated, however, by methodological differences such as the selection of English contrasts examined and the variety of English tested (SSBE vs. American English). In terms of the PAM, /iː/-/ɪ/ was classified as a categorized-uncategorized clustered type of assimilation, with partial perceptual overlap. The English pair /ɛ/-/ɜː/ was classified as a category-goodness difference type of assimilation (and as a categorized-uncategorized clustered type in a previous study, Cebrian, 2019). According to Faris et al. (2018), the two types of assimilation pose a similar degree of discrimination difficulty, and thus discrimination accuracy depends on the amount of phonological overlap. The /ɛ/-/ɜː/ pair, with a greater phonological overlap (72% as Spanish /e/), was expected to pose a greater difficulty than the /iː/-/ɪ/ contrast (32% overlap as /i/), but the results showed the opposite pattern.

L2 Production

Production Results

The vowel production of the 38 L2 speakers (recall that five participants only completed the perception tasks) and 13 native English speakers was analyzed acoustically in terms of first and second formant (F1 and F2) and vowel duration. Vowel formants were measured from a 25 ms window located manually at a steady-state portion between one third and two fourths into the vowel, using Praat (Boersma and Weenink, 2018). **Figure 3** displays the mean F1 and F2 values in Hertz per vowel for each of the L2 speakers, showing the data for female and male L2 speakers separately. Each individual is represented by a number. As can be observed from the distribution of the individuals' production, there is considerable spectral overlap between /iː/ and /ɪ/, while /ɛ/ and /ɜː/ appear to be distinguished more clearly. As expected, a fair amount of individual variation can be observed. For example, male speaker 15 produced practically indistinguishable /iː/ and /ɪ/ (see **Figure 3**). By contrast, male speaker 47 produced these two vowels with very different spectral values. The issue of whether this difference in performance can be predicted from individual differences in perceptual assimilation patterns is analyzed in *Cross-Task comparison. Correlational analysis.*

In order to compare the L2 data to native English data, and to group males and female speakers together, the formant values were normalized following the Lobanov method (Lobanov, 1971). A Lobanov transformation is a vowel extrinsic normalization method that has been found to render superior normalization outcomes to other methods, particularly if the data set examined includes a large set of vowels (Adank et al., 2004; Recasens and Espinosa, 2009). Recall that, although this paper focuses on four vowels, the current data is a subset of a larger set including a larger number of vowels, which were taken into account in the normalization procedure. Normalization was carried out using NORM (Thomas and Kendall, 2007). **Figure 4** displays the spectral characteristics of the native and L2 productions, averaged across the 38 L2 speakers (L2S; 33 females) and the 13 native English speakers (NES; seven females). As can be observed, as a group the L2 speakers seemed to produce English /ɜː/ and /ɛ/ as two separate vowels and with values that were close to those of NES's. By contrast, the learners' production of /iː/ and /ɪ/ were much closer than the respective productions by native speakers.

A series of GLMMs were conducted[3] with F1 and F2 (Lobanov-normalized values) as the dependent variables to assess if the L2 speakers produced the English vowels differently and if the L2 production differed from NES production. In both cases the best fitting model was a GLMM with Group (learners vs. NES) and Vowel (the four target vowels) as fixed effects, and random intercepts for participant and for word. Regarding F1 values, the model revealed a significant effect of Vowel (F (3, 1,801) = 238.02, $p < 0.001$) and a significant Group by Vowel interaction (F (3, 1,801) = 82.4, $p < 0.001$), but no significant effect of Group (F (1, 1,801) = 2.488, p = 0.115). Bonferroni-adjusted pairwise comparisons indicated that the L2 speakers differed from the native speakers with respect to /ɛ/, /iː/ and /ɪ/ ($p < 0.001$), but not to /ɜː/ (p = 0.15), which explains the significant interaction. NES produced significantly different F1 values for all four vowels ($p <$ 0.001), while L2S produced significant differences between all vowels ($p < 0.001$) except for the /iː/-/ɪ/ contrast (p = 0.335). With respect to F2, all main effects and the interaction reached significance (Vowel: F (3, 1,801) = 235.92, $p < 0.001$; Group: F (1, 1801) = 4.6, p = 0.031; Group x Vowel: F (3, 1,801) = 105.87, $p <$ 0.001). Pairwise comparisons in this case indicated that the L2 speakers differed from NES with respect to /ɜː/, /iː/ and /ɪ/ ($p <$ 0.001), and /ɛ/ ($p < 0.01$). The smaller level of significance for the group comparison involving /ɛ/ may explain the significant interaction. NES produced significantly different F2 values for all four vowels ($p < 0.001$). As was found for F1, L2S produced a significant F2 difference between all vowels ($p <$ 0.001) except for the /iː/-/ɪ/ contrast (p = 0.08). In brief, although L2 productions generally differed from NES's, the production of /ɛ/ and /ɜː/ by L2 speakers deviated to a lesser extent from NES's production. In addition, learners produced significant spectral differences between all vowels except for the /iː/ and /ɪ/ pair.

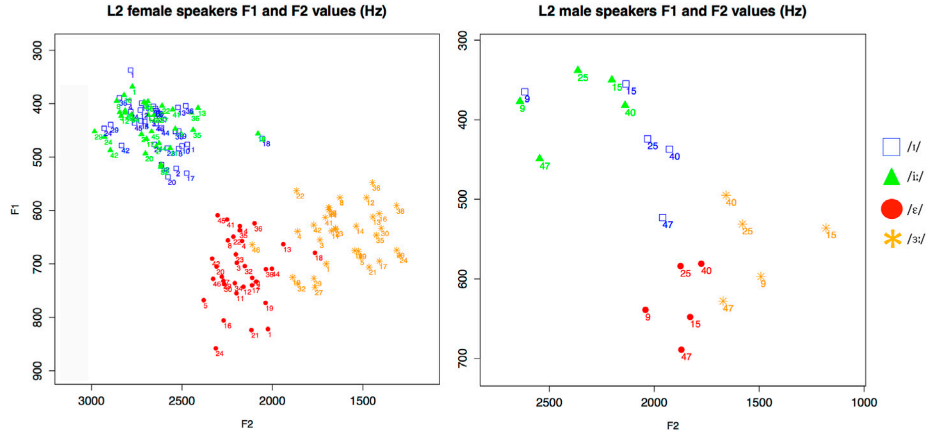


FIGURE 3 | F1 x F2 values (in Hertz) of the four English vowels produced by female (left, *N* = 33) and male (right; *N* = 5) L2 English speakers. Numbers represent individual participants.

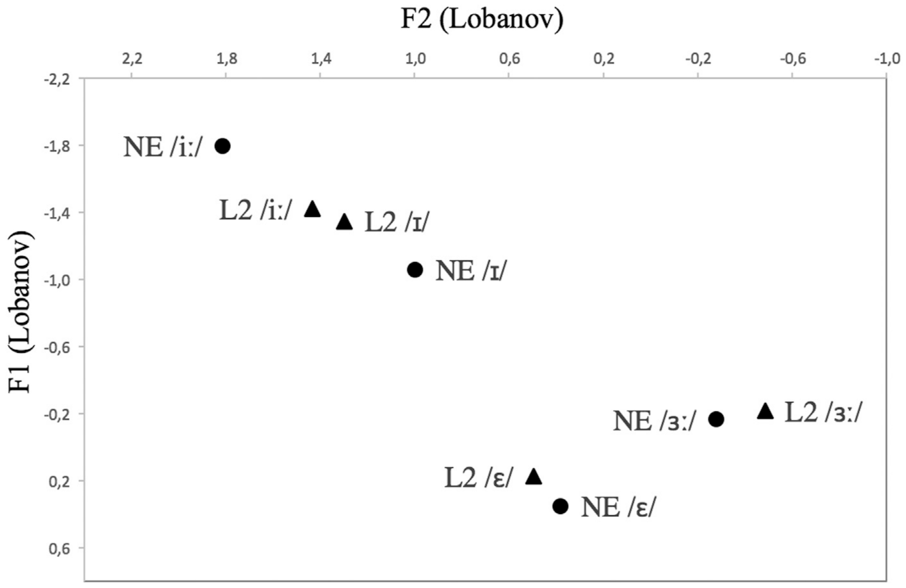


FIGURE 4 | Normalized F1 x F2 values of each of the four vowels averaged across the native English speakers (NE) and the L2 English speakers (L2). The Lobanov transformation has been used (see text).

Given previous findings about Spanish/Catalan speakers' reliance on temporal cues for the English /iː/-/ɪ/ contrast (Cebrian, 2006; Rallo Fabra and Romero, 2012), vowel duration was also examined. **Table 4** provides the means and standard deviations for each group and vowel, as well as the duration ratio for /iː/-/ɪ/ and /ɜː/-/ɛ/. As can be observed, both groups produced the tense vowels with greater duration than lax vowels, but the duration difference was greater for NES. The results of a GLMM exploring vowel duration with Group and Vowel as fixed effects and a random slope for participants revealed a significant effect of Vowel (*F* (3, 1,802) = 469.25, $p < 0.001$) and a significant Group by Vowel interaction (*F* (3, 1,802) = 58.41, $p < 0.001$), but no significant effect of Group (*F* (1, 1,802) = 3.176, $p = 0.075$). Bonferroni-adjusted pairwise comparisons indicated that the L2 speakers differed from the native speakers in the duration of the tense vowels /ɜː/ and /iː/ ($p < 0.001$), but not regarding the lax vowels /ɪ/ ($p = 0.304$) and /ɛ/ ($p = 0.249$), which explains the significant interaction. Both groups produced highly significant duration differences between all vowels ($p < 0.001$), except for the difference between /ɛ/-/ɪ/ for NES, which was significant at the $p < 0.05$, and the difference between /ɛ/-/iː/, which was not significant for the L2 speakers ($p = 0.747$).

It has been argued that perceptual measures such as native speaker judgements are a more appropriate method for assessing L2 production and accentedness than acoustic analyses (e.g., Munro, 2008; Derwing and Munro, 2015). A preliminary

TABLE 4 | Mean vowel duration (in ms) and duration ratio for L2 speakers (L2S) and English native speakers (NES).

	L2S (*n* = 38)	NES (*n* = 13)
/iː/	154 (38)	190 (29)
/ɪ/	140 (29)	129 (24)
/ɜː/	209 (38)	257 (35)
/ɛ/	153 (28)	141 (24)
/iː/-/ɪ/ ratio	1.11 (0.22)	1.50 (0.24)
/ɜː/-/ɛ/ ratio	1.38 (0.18)	1.88 (0.33)

perceptual analysis of the production data was carried out. Four judges (native speakers of SSBE aged between 25 and 36) rated a selection of all the words produced by the L2 speakers and the native English controls (recall that the current data form part of a larger study involving a greater number of vowels). On a given trial raters first identified the vowel in the stimulus in terms of one of ten English words presented on a computer screen (including meet, sit, set and shirt) and then provided a goodness rating on a 7-point scale where 7 meant native-like. The results show that the L2 speakers' production of English /ɛ/ was very accurate, with 100% identification and an average goodness rating (GR) of 5.8 out of 7. Vowel /ɜː/ received very high identification scores too, 89%, and a relatively high GR of 5. Vowel /iː/ was correctly identified 84% of the time with a GR of 4.9, and finally vowel /ɪ/ was the least accurately produced, reaching only 50% correct identification, with a GR of 4.3, and being heard as /iː/ 43% of the time with a GR of 4.3. Although these results are preliminary as so far only two judgements per stimulus have been obtained, they appear to confirm the acoustic analysis and point to the strongly assimilated vowel /ɛ/ as the most accurately produced vowel, followed by /ɜː/ and /iː/, while /ɪ/ was the least accurately produced.

Discussion of L2 Production Results

The production results show better production of /ɛ/ and /ɜː/ than of /iː/ and /ɪ/. These results are consistent with previous studies involving Spanish learners of English. Flege et al. (1997) reported that two groups of Spanish speakers differing in the amount of time spent in the United States (0.5 vs. 9 years) produced English /ɛ/ very accurately (correctly identified between 91 and 99% of the time by native English judges), while /iː/ and /ɪ/ obtained similarly lower identification scores (51–61 and 57–69%, respectively). Cebrian (2007) also found that English /ɛ/ was the most accurately produced vowel by a group of 30 Spanish/Catalan speakers of English (86% correct identification and a 5.3/7 goodness rating by native English speakers), while /iː/ and /ɪ/ were less accurately produced (/iː/ 73% and GR 4.5/7; /ɪ/: 71%, GR: 3.9). The native speaker judgements were consistent with the acoustic analysis of L2 data, which revealed a large overlap in F1 and F2 space between /iː/ and /ɪ/, while /ɛ/ was more distinctly produced. Further, in a study involving Catalan learners of English, Rallo Fabra and Romero (2012) reported that the learners produced /ɪ/ more accurately than /iː/ (71% vs. 49% correct identification by native speakers). Regarding /ɜː/, the current results are in agreement with Carlet and Cebrian's (2019) study involving Spanish/Catalan learners of English, whose /ɜː/ was judged to be comparatively accurate by native English judges. Raters in that study also judged /iː/ to be better produced than /ɪ/, in line with our preliminary native speaker judgements. Similarly, Carlet and Kivistö-de Souza (2018) found that the productions of /iː/ by Spanish/Catalan learners of English were judged to be more accurate (5.1 on 9-point Likert scale) than those of /ɪ/ (4.8). Overall, then, previous studies tend to find an accurate production of English /ɛ/, and a spectral overlap in the production of /iː/ and /ɪ/, which are found to be less accurately produced than /ɛ/.

Regarding vowel duration, the results of the current study show a smaller reliance on temporal cues for the /iː/-/ɪ/ contrast than previous studies involving Catalan/Spanish L2 English speakers (Cebrian, 2007; Aliaga-Garcia and Mora, 2009; Rallo Fabra and Romero, 2012), as the tense vowel was only 11% longer than the lax vowel (compared to 50% for NES). Still, there was variability among the participants. One of the SSBE native speakers did not produce a temporal difference between /iː/-/ɪ/. Regarding the L2 speakers, eight had duration ratio values for /iː/-/ɪ/ within the range of the SSBE speakers' values, while 15 had ratio of 1 or lower. In order to see if there was an inverse relationship between producing a temporal difference and a spectral difference, the relationship between the /iː/-/ɪ/ duration ratio and the spectral Euclidean distance between their production of /iː/ and /ɪ/ was examined. No significant correlation was found, indicating that those participants who relied more on spectral differences were not necessarily those who relied less on temporal differences.

In summary, the production of /ɛ/ follows predictions for a highly assimilated vowel that may pattern as a near identical vowel. Results for /ɜː/ are unexpected for a comparatively new vowel that was poorly identified. Results for /iː/ and /ɪ/ show that, while /iː/ tends to be better produced, particularly as judged from a native perceptual perspective, it is not as accurately produced as /ɛ/ despite comparable assimilation scores for English /ɛ/ and /iː/ as Spanish /e/ and /i/, respectively. In fact, /iː/ patterns more with /ɪ/ than with /ɛ/ in production accuracy. The analysis also revealed that learners made use of temporal cues, but to a lesser extent than what has been reported in previous studies.

Cross-Task Comparison. Correlational Analysis

One of the goals of this paper is to examine if individual variation in the perceived similarity between L1 and L2 sounds is related to perception and production accuracy, as would be predicted by the SLM-r (Flege and Bohn, 2021). In order to answer this question, a number of Spearman correlations were conducted relating cross-linguistic perceptual similarity measures and L2 perception and production measures. Specifically, the perceptual assimilation measures included in the analysis were percent assimilation, goodness ratings, and the composite fit index score for each vowel. Identification accuracy for each L2 vowel and discrimination accuracy for each vowel contrast were the L2 perception measures. Finally, two types of production measures were explored. First, the extent to which speakers distinguish the two vowels in each contrast, that is, the Euclidean distances between /iː/ and /ɪ/, and between /ɜː/ and /ɛ/. In addition, to explore the degree to which L2 speakers approximated native production, the Euclidean distance between each L2 individual's production of each vowel and the native English speakers' average production was also examined.

Only a subset of the possible correlations reached significance (see **Supplementary Material** in the Appendix for plots of the significant correlations). Accuracy in the identification of vowel /iː/ was moderately correlated with the percent assimilation of English /ɪ/ to Spanish /e/ (r_s = 0.331, p = 0.036, n = 43) and negatively correlated with the fit index of English /ɪ/ to Spanish /i/ (r_s = −0.333, p = 0.029, n = 43). Accurate /iː/-/ɪ/ discrimination was moderately correlated with the fit index of English /iː/ to Spanish /i/ (r_s = 0.364, p = 0.016, n = 43). Similarly, /ɛ/-/ɜː/ discrimination was related to a greater assimilation percentage of English /ɛ/ to Spanish /e/ (r_s = 0.335, p = 0.028, n = 43). These results show that subjects who identified English /iː/ more successfully tended to assimilate English /ɪ/ to Spanish /e/ rather than to Spanish /i/, and that subjects who were more successful at discriminating English /iː/-/ɪ/ and English /ɛ/-/ɜː/ tended to assimilate English /iː/ to Spanish /i/ and English /ɛ/ to Spanish /e/, respectively, more consistently. However, there was no correlation involving the perceived similarity results obtained for /ɪ/ and /ɜː/ and the identification scores obtained for these vowels. No correlations were found between the cross-linguistic similarity measures and the production measures. On the other hand, moderate correlations were found between the results for different vowels in the same task. Identification of English /ɛ/ was correlated with the identification of /ɜː/ (r_s =0 .323, p = 0.035, n = 43) and of /ɪ/ (r_s = 0.329, p = 0.031, n = 43). Within production, a greater Euclidean distance between /iː/ and /ɪ/ was negatively correlated with the degree of difference between an L2 speaker's production and the native English mean in the case of /iː/ production (r_s = −0.370, p = 0.022, n = 38) and /ɪ/ production (r_s = −0.639, p = 0.000, n = 38). This means that the more differently participants produced /ɪ/ and /iː/, the more their productions resembled the native speakers' production. Those participants who produced /ɪ/ more accurately (that is, with values closer to those of native speakers') also tended to produce /iː/ more accurately, though the correlation in this case was marginal (r_s = 0.320, p = 0.05, n = 38).

In conclusion, perceived similarity does not seem to determine accuracy of L2 vowel perception or production at an individual level. The difficulty with the /iː/-/ɪ/ contrast, in particular, does not seem to follow from the pattern of assimilation of each of these two vowels to L1 categories. Some possible explanations for this fact are discussed next.

GENERAL DISCUSSION

Previous studies have found that Catalan/Spanish learners of English, in addition to learners of other language backgrounds, have difficulty distinguishing the English /iː/-/ɪ/ contrast both in perception and in production (e.g., Flege, 1991; Flege et al., 1997; Escudero and Boersma, 2004; Cebrian, 2006; Cebrian, 2007; Kondaurova and Francis, 2008; Morrison, 2008; Morrison, 2009). L2 speech theories relate accurate perception and production of target language phones to the ability to discern differences between L1 and L2 sounds (Best, 1995; Flege, 1995; Escudero, 2005; Best and Tyler, 2007; Flege and Bohn, 2021). This study set out to examine if this difficulty is related to the degree of perceived similarity between L1 and L2 vowels, both in terms of group results and individual judgements of cross-linguistic similarity. To that effect, a group of 43 Catalan/Spanish bilinguals, who were undergraduate students in English Studies at a Spanish university, performed a perceptual assimilation task evaluating the perceived similarity between English /iː, ɪ, ɛ, ɜː/ and the perceptually closest Spanish vowels. English /ɛ, ɜː/ were included for comparison purposes. Previous studies had found that English /iː/ and /ɛ/ have a clear counterpart in the L1, Spanish /i/ and /e/, respectively, while English /ɪ/ and /ɜː/ are less consistently mapped onto an L1 vowel (see **Table 1** above). The goal of the current study was thus to contrast the perceived similarity judgements obtained by a group of Spanish speakers with these same speakers' ability to identify the English vowels, discriminate each vowel pair (/iː, ɪ/, /ɛ, ɜː/) and produce each vowel accurately. In addition, this study investigated if the individual variation often observed in perceived similarity was related to the variability found in L2 performance.

The results of the perceptual assimilation task confirmed previous findings (e.g. Cebrian, 2019) and showed that the English vowels differed in the degree to which they assimilated perceptually to Spanish categories. English /iː/ and /ɛ/ were strongly assimilated to Spanish /i/ and /e/ (93 and 99%, respectively). By contrast, English /ɪ/ and /ɜː/ patterned as more dissimilar to an L1 vowel, with /ɜː/ barely reaching a categorization threshold of 70% (72%) and /ɪ/ being slightly below that threshold (67%). Both /ɪ/ and /ɜː/ obtained low goodness of fit scores, and consequently comparatively low fit indices (4/7 and 3.1/7, respectively; see **Table 2** above for all results). Based on their strong assimilation to Spanish /i/ and /e/, it was hypothesized that English /iː/ and /ɛ/ would be perceived and produced by the L2 learners with similarly high levels of accuracy. Regarding the poorly assimilated vowels, perception and production accuracy may depend on the extent to which the L2 speakers have detected differences between the L1 and the L2 phones to establish separate categories for these vowels (Flege, 1995). This in turn may depend on the amount of authentic L2 input received as well as the amount of L2 use (Flege and Bohn, 2021). Given that L2 learning took place in an instructional setting in the L1 country, with overall limited amount of L2 exposure and use, it is likely that learners were at a stage when the most dissimilar vowels are still challenging. In terms of discrimination ability, the PAM-L2 predicted that for L2 contrasts classified as uncategorized-categorized with partial overlap (/iː/-/ɪ/) and category-goodness assimilation (/ɛ/-/ɜː/), discrimination accuracy depends on the degree of cross-language assimilation overlap (Faris et al., 2018; Tyler, 2021). Thus, the expectation was that /iː/-/ɪ/ would be better discriminated than /ɛ/-/ɜː/, given the greater overlap displayed by the latter (72% as Spanish /e/) than the former (32% as Spanish /i/).

The results of the vowel identification test showed that English /ɛ/ was the most accurately identified vowel, while /ɜː/ was the worst (90 and 47%, respectively). This result may follow from the assimilation PAT results. However, the results for /iː/ and /ɪ/ did not follow the expectations, and, in fact, the latter was better identified than the former (/ɪ/: 72%, /iː/: 58%). This outcome is in line with previous research on the effect of perceptual training that found that Catalan/Spanish learners of English identified /ɪ/ more successfully than /iː/ and /ɜː/ (Cebrian and Carlet, 2014; Carlet and Cebrian, 2019; Fouz-González and Mompean, 2020) and that identification of /iː/ and /ɜː/ improved the most after participants underwent perceptual training. The discrimination results also failed to support the predictions, as /ɛ/-/ɜː/ was more accurately discriminated than /iː/-/ɪ/ (87% vs. 66%, respectively). The discrimination results were replicated in production, as /ɛ/ and /ɜː/ were also more accurately produced than /iː/ and /ɪ/, which tended to be produced as /iː/. In the case of /ɛ/, results were consistent across tasks and in agreement with previous studies involving Spanish L2 English speakers (Flege et al., 1997; Cebrian, 2007). Further studies could investigate if this Spanish vowel would be undetectable by native English listeners when produced in a native English context (Flege, 1992). In fact, our preliminary data from native English listeners yielded very high correct identification scores for the learners' /ɛ/ production (100%, with an average goodness rating of 5.8/7). The results for /ɜː/ showed a surprisingly high level of accuracy, in contrast with its poor identification results. Few studies have examined the production of this vowel by Spanish/Catalan learners, but Carlet and Cebrian (2019) obtained a similar result with a similar population (second-year undergraduate students in an English Studies degree, as opposed to first-year undergraduates in the current study). The relative success in the production of /ɜː/ may indicate that learners are capable of producing a vowel that is different from other L2 (and L1) vowels, even if they cannot successfully identify it. This difference may be methodologically based, as identification involves the ability to distinguish the target sound from other potentially conflicting sounds (e.g., other response options present in the identification task), while production involved the articulation of the sound in a high frequency word, which may make production more successful. The idea that accurate perception precedes accurate production in L2 (e.g., models like Flege (1995) SLM or Best (1995) PAM) is not always supported by the findings (e.g., Llisterri, 1995), and more recent proposals suggest that L2 perception and production may develop without the requirement that one modality precedes the other (Flege and Bohn, 2021, SLM-r).

The more striking results involve the vowel contrast that was the focus of this study, the English /iː/-/ɪ/ contrast. In line with some previous studies, the results showed a high degree of perceptual confusion between /iː/ and /ɪ/, which achieved only 66% correct discrimination, and were often misidentified as one another in the identification task. In addition, the two vowels were produced with a large amount of spectral overlap. In terms of duration, L2 learners' tense vowels were longer than the lax vowels, but this difference in duration was smaller than the duration difference produced by NES and also by other learners of English of a similar background reported in the literature (Cebrian, 2007; Rallo Fabra and Romero, 2012). The possible role of duration in the categorization of the English /iː/-/ɪ/ contrast is further discussed below.

One of the goals of this study was to assess if variability in L2 production and perception was related to individual differences in cross-linguistic perceived similarity. The acquisition of the English /iː/-/ɪ/ contrast by Spanish/Catalan learners of English provides a good ground to explore this issue as previous studies show inconsistent patterns of assimilation of English /ɪ/ to L1 vowels. Recall that English /ɪ/ is generally assimilated to Spanish /i/ and /e/ with varying degrees across studies. Assimilation of English /ɪ/ to Spanish /e/ ranges from 50% or less to 90% or more (see **Table 1** above). The current study found that English /ɪ/ was identified with Spanish /e/ two thirds of the time, and with /i/ one third of the time. An inspection of individual data showed that more than half the participants selected Spanish /e/ as the closest L1 vowel to English /ɪ/, while about a third selected Spanish /i/ and the remaining were not consistent and selected both Spanish /i/ and /e/. Therefore, it was possible that these differences in perceptual assimilation influenced the way learners perceived and produced the English vowels. For instance, participants who predominantly perceived English /ɪ/ as Spanish /e/ may be more successful at discriminating between /ɪ/ and /iː/ than those who assimilate both English /iː/ and /ɪ/ as Spanish /i/. However, the correlational analysis involving the perceptual similarity measures, the identification and discrimination results and the production results yielded very little evidence of a relationship between individuals' cross-linguistic perceived similarity and their L2 performance. Those learners who assimilated English /ɪ/ to Spanish /e/ and who judged English /iː/ to be closest to Spanish /i/, tended to be better at identifying English /iː/ and discriminating between /iː/ and /ɪ/. Still, no correlation involving the assimilation patterns obtained for /ɪ/ and /ɜː/ and their identification or production accuracy were found. Therefore, the current study presents only a weak relationship between cross-linguistic perceived similarity and L2 performance at an individual level and hence it does not provide evidence to the claim that learners' accuracy in perception and production is related to individual variation in the perceived similarity between L1 and L2 sounds (Flege and Bohn, 2021). Future research is necessary to examine if a relationship can be found with other measures of similarity and of production and perception and with learners of different levels of experience.

In summary, the perception and production of the English /iː/-/ɪ/ contrast by the L2 speakers in the current study does not follow from the perceived similarity relationships between the target and the native vowels, neither at a group level nor at an individual level. This outcome is consistent with the results reported in previous studies involving Spanish and Catalan learners of English showing comparatively poor perception and production of both /iː/-/ɪ/, as discussed in *Results and Interim Discussions*. In fact, similar results are reported for learners of other L1 backgrounds like Italian and Japanese (e.g., Flege et al., 1998; Grenon et al., 2019).

What is it then about the English /iː/-/ɪ/ contrast that makes it difficult for L2 speakers to acquire? And why is it that a target vowel like English /iː/ that is perceptually very close to an L1 vowel is often found to be poorly perceived and produced? Flege (2018) explains that cases of inaccurate L2 performance that have been attributed to variables such as a late starting age of learning or number of years of L2 use can be accounted for in terms of the quality or quantity of the input, that is, whether learners are exposed to authentic native-like input or to variable and often accented input (see also Flege and Bohn, 2021). For example, Casillas (2015) reports that early Spanish-English bilinguals, who had been exposed to English since age 4–6 and were dominant in English as adults still differed from native English speakers in their use of spectral and temporal cues in their perception (but not the production) of English /iː/-/ɪ/. A possible explanation offered is exposure to Spanish-accented English. Morrison (2012), who also found that Spanish speakers identified English /iː/ and /ɪ/ with Spanish /i/ and /e/, respectively, argued that a perceptual explanation for Spanish speakers' difficulty with the /iː/-/ɪ/ contrast is not satisfactory and points to the effects of pronunciation instruction and orthography as possible causes. Indeed, orthography has been found to influence the pronunciation of L2 words (e.g., Morrison, 2009; Escudero and Wanrooij, 2010). This influence may be particularly notable in foreign language learning contexts where L2 words are often first encountered in writing. For example, the fact that the English lax vowel /ɪ/ is typically spelt with the letter <i>, which represents the sound /i/ in many languages like Spanish, Catalan and Italian, may result in the mispronunciation of this vowel (Morrison, 2008). English /ɛ/, on the other hand, is often spelt with the same letter as in Spanish, <e>. The effect of orthography is likely enhanced by the presence of many cognate words, particularly for speakers of Romance languages (e.g., words containing <i> such as cinema, city, or minimum). Further, the tense vowel /iː/ is often represented by a two-letter grapheme, e.g., <ee> in feet or <ea> in beat, which may suggest a longer duration. In addition, the effect of explicit instruction that describes the contrast as merely a duration contrast (e.g., long /i/ vs. short /i/) may result in a misrepresentation of the tense-lax contrast as a purely duration contrast (Flege et al., 1997; Wang and Munro, 1999).

As discussed in *Introduction*, L2 English speakers' have been found to rely on duration to implement the English /iː/-/ɪ/ (e.g., Escudero and Boersma, 2004; Cebrian, 2006; Kondaurova and Francis, 2008). A combination of factors may account for this fact. In addition to the issues described above, there are linguistic factors such as sensitivity to phonetic temporal distinctions in the L1, e.g., as a result of stress or final obstruent voicing (Kondaurova and Francis, 2008), desensitization to spectral contrasts not used in the L1 (Bohn, 1995), and statistical learning of the characteristics of each vowel with the consequent detection that /iː/ tends to be longer than /ɪ/ (Escudero and Boersma, 2004; Morrison, 2008). For example, Morrison suggests that learners go through different stages in the process of learning the /iː/-/ɪ/ contrast which go from the inability to discern the two vowels, to establishing a duration contrast, to gradually detecting and incorporating spectral differences. Following this interpretation, we can speculate that when learners start to detect spectral differences between English /ɪ/ and Spanish or Catalan /i/, their L2 category starts to drift toward a more /ɪ/-like vowel. However, since at this stage both L2 vowels, /iː/-/ɪ/, share a single spectral category, possibly distinguished temporally, the drift toward /ɪ/ affects the perception and production of not only /ɪ/, but also /iː/. This would result in a "deterioration" of /iː/, a vowel that by itself should not be problematic given the high assimilation rates to an L1 vowel. In a similar vein, Major (1987) found that Brazilian Portuguese speakers produced the similar English vowel /ɛ/ with increasing less accuracy as their production of the new vowel /æ/ improved. In any event, the comparatively poor performance with English /iː/ shows that L2 vowels are not learned individually but as part of a system of contrasting phones. This factor, together with the effect of non-linguistic factors like the roles of orthography and explicit instruction, may account for why cross-linguistic similarity may not always make the right predictions, and why a target vowel that should in principle be "easy" to learn on the basis of its strong perceptual assimilation to an L1 vowel, like English /iː/ for Spanish speakers, is not accurately perceived and produced in the L2.

IMPLICATIONS, LIMITATIONS AND CONCLUSION

The observations made in the previous section have some pedagogical implications for the teaching and learning of this vowel contrast. First of all, the contrast between English /iː/ and /ɪ/ should be set as a priority in L2 speech teaching given the contrast's high functional load, and, consequently, the fact that mispronunciation may contribute to loss of intelligibility and reduction of comprehensibility [Brown, 1988; Levis, 2018; Munro, 2021 (this volume); O'Brien, 2021 (this volume)]. Secondly, instructors should avoid characterizing the English vowels /iː/ and /ɪ/ as "long" and "short" respectively, as this may lead to an inaccurate simplification of the difference between them and may make learners ignore vowel quality differences. Although the duration difference is an important feature to illustrate, the focus should be on the qualitative difference. In that sense, the results of cross-linguistic perceived similarity tasks could be useful. For instance, given that English /ɪ/ is perceptually closer to Spanish /e/ than to Spanish /i/, learners' attention could be drawn to this L1 vowel quality difference as a way of helping learners detect differences between English /iː/ and /ɪ/. On a similar note, cognate words could be used to illustrate differences in L1 and L2 pronunciation (e.g., Spanish "lista, video" and English "list, video"). Further, instruction should avoid exercises involving the written form of words, at least at initial stages, and concentrate on auditory input as much as possible, in order to avoid the confusion that spelling might cause in the acquisition of this contrast. The use of phonetic symbols is a possible option; Fouz-González and Mompean (2020) report that both the use of keywords and phonetic symbols in high variability phonetic training enhances identification accuracy of L2 vowels. See O'Brien (2021) (this volume) for an overview of approaches to and suggestions for pronunciation teaching. Finally, note that, as Munro (2021) (this volume) demonstrates, while there is some systematicity in the L2 pronunciation difficulties observed for a given L1 group, there is considerable variation in the pronunciation difficulties experienced by different individuals. This fact underscores the importance of individual pronunciation diagnosis and pronunciation practice targeting individual difficulties.

The current study has some limitations that should be acknowledged. The first limitation is the cross-sectional nature of the current study. A single group of learners with a similar level of proficiency was tested at a single point in time. Additional groups differing in amount of experience or a longitudinal approach would be necessary to fully evaluate the relationship between perceived similarity and L2 performance. In fact, L2 experience has been found to influence the way learners categorize the /iː/-/ɪ/ contrast (Aliaga-García and Mora, 2009; Ylinen et al., 2009; Grenon et al., 2019). For instance, Grenon et al. found that a subset of the inexperienced Japanese learners of English trained to attend to spectral differences between English /iː/ and /ɪ/ managed to modify their original reliance on duration and to attend to spectral cues, although their perception of the English contrast was not completely native-like. Another shortcoming involves the lack of L1 production data to compare to the L2 data so as to have a measure of how differently the L1 and L2 vowels are produced, particularly when evaluating cases of near-identical L1-L2 vowels. An additional limitation is related to the bilingual nature of the participants in this study, as, in addition to speaking Spanish, many participants spoke Catalan, with varying degrees of dominance (see *Participants*). Given the difference in vowel inventory between the two languages (five Spanish vowels vs. seven Catalan vowels), it is difficult to determine, for instance, if the successful perception and production of English /ɛ/ is related to the perceived similarity between this English vowel and Spanish /e/ or Catalan /ɛ/[4]. Even if differences in Catalan/Spanish dominance were not found to affect similarity judgements (see *Perceptual Assimilation Tasks Results*), and the results of the PAT in the current study closely replicated the results obtained by Cebrian (2019) with monolingual Spanish speakers, there is a potential effect of speaking more than one language on L2 perception and production. Further, production data was mainly measured acoustically instead of by means of native speaker judgements, a measure that has been argued to be more appropriate for measuring L2 production and accentedness (Munro, 2008). Finally, the limited number of vowels examined, two vowel contrasts, prevents the comparison of the current results with those for other vowel combinations, for example in the discrimination task. These issues are left for further research.

To conclude, the perception and production of L2 phones do not always follow from cross-linguistic similarity relations, contrary to the predictions of most current L2 speech models. In addition, no clear explanation has been found based on individual variation, as variability in L2 to L1 mapping does not seem to be related to the degree of L2 perception or production accuracy, at least with the measures and the level of proficiency evaluated in this study. Still, some sounds may be readily perceived and produced successfully in terms of the L1 category (e.g., English /ɛ/ by Spanish speakers). Further, similarity predictions may differ for production and perception accuracy, as was the case of /ɜː/, a vowel that was very poorly perceived but relatively accurately produced, which may be related to different demands for perception and production tasks. The results for the /iː/-/ɪ/ contrast suggest that different factors are at play in the acquisition of this L2 contrast, including non-native overreliance on acoustic cues (triggered by statistical learning or by metalinguistic information present in pronunciation instruction), the effect of orthography, and the need to maintain an L2 contrast with a high functional load. These factors may override the effect of cross-linguistic similarity, as the need to categorize the most dissimilar sound may bring with it the deterioration of the more similar sound, at least at some stages. These issues may be behind the reason why a presumably “easy” sound like English /iː/ may be more challenging than expected.

DATA AVAILABILITY STATEMENT

Data from this study are available by request if required to form part of a metaanalysis or similar form of research. Requests to access the datasets should be directed to juli.cebrian@uab.cat.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JC contributed to conception and design of the study. JC, CG and NG collected the data, analyzed the results and performed the statistical analysis. JC wrote the first draft of the manuscript. CG and NG wrote sections of the manuscript.

FUNDING

This work was supported by research grants FFI2017-88016-P the Spanish Ministry of Economy and Competitiveness, and by the research group Grant 2017SGR34 from the Catalan Government.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.660917/full#supplementary-material

[4] See Cebrian (2021) for a study on the perceived similarity between a near-complete set of Catalan and SSBE vowels.

REFERENCES

Adank, P., Smits, R., and van Hout, R. (2004). A Comparison of Vowel Normalization Procedures for Language Variation Research. *J. Acoust. Soc. Am.* 116, 3099–3107. doi:10.1121/1.1795335

Aliaga-García, C. (2011). "Measuring Perceptual Cue Weighting after Training: A Comparison of Auditory vs Articulatory Training Methods," in *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*. Editors M. Wrembel, M. Kul, and K. Dziubalska-Kołaczyk (Frankfurt am Main, Germany: Peter Lang), 15–28.

Aliaga-García, C., and Mora, J. C. (2009). "Assessing the Effects of Phonetic Training on L2 Sound Perception and Production," in *Recent Research in Second Language Phonetics/Phonology: Perception and Production*. Editors M. A. Watkins, A. S. Rauber, and B. O. Baptista (Newcastle upon Tyne, United Kingdom: Cambridge Scholars Publishing), 2–31.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (2004). Perceived Phonetic Dissimilarity and L2 Speech Learning: The Case of Japanese /r/ and English /l/ and /r/. *J. Phon.* 32 (2), 233–250. doi:10.1016/S0095-4470(03)00036-6

Baigorri, M., Campanelli, L., and Levy, E. S. (2019). Perception of American-English Vowels by Early and Late Spanish-English Bilinguals. *Lang. Speech.* 62 (4), 681–700. doi:10.1177/0023830918806933

Best, C. T. (1995). "A Direct Realist View of Cross-Language Speech Perception," in *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Editor W. Strange (Timonium, MD: York Press), 171–204.

Best, C. T., and Strange, W. (1992). Effects of Phonological and Phonetic Factors on Cross-Language Perception of Approximants. *J. Phon.* 20 (3), 305–330. doi:10.1016/S0095-4470(19)30637-0

Best, C. T., and Tyler, M. D. (2007). "Nonnative and Second-Language Speech Perception," in *Honor of James Emil Flege*. Editors O. S. Bohn and M. J. Munro (Amsterdam, Netherlands: John Benjamins), 13–34. doi:10.1075/lllt.17.07bes

Birdsong, D., Gertken, L. B., 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. (2018). Praat: Doing Phonetics by Computer [Computer Program]. version 6.0.43. Available at: http://www.praat.org/ (Accessed September 8, 2018).

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*. Editor W. Strange (Timonium, MD: York Press), 279–304.

Bohn, O.-S., and Flege, J. E. (1992). The Production of New and Similar Vowels by Adult German Learners of English. *Stud. Second Lang. Acquis.* 14 (02), 131–158. doi:10.1017/S0272263100010792

Bohn, O.-S., and Steinlen, A. K. (2003). "Consonantal Context Affects Cross-Language Perception of Vowels," in Proceedings of the XV International Congress of Phonetic Sciences. Editors D. Recasens, M. J. Solé, and J. Romero (Barcelona, Spain: Causal Productions), 2289–2292.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. i. (1997). Training Japanese Listeners to Identify English /r/ and /l/: IV. Some Effects of Perceptual Learning on Speech Production. *J. Acoust. Soc. Am.* 101, 2299–2310. doi:10.1121/1.418276

Brown, A. (1988). Functional Load and the Teaching of Pronunciation. *TESOL Q.* 22 (4), 593–606. doi:10.2307/3587258

Bundgaard-Nielsen, R. L., Best, C. T., and Tyler, M. D. (2011). Vocabulary Size Matters: The Assimilation of Second-Language Australian English Vowels to First-Language Japanese Vowel Categories. *Appl. Psycholinguist.* 32, 51–67. doi:10.1017/S0142716410000287

Busà, M. G. (1992). "On the Production of English Vowels by Italian Speakers with Different Degrees of Accent," in New Sounds 92, Proceedings of the 1992 Amsterdam Symposium on the Acquisition of Second-Language Speech. Editors J. Leather and J. Allan (Amsterdam, Netherlands: University of Amsterdam), 47–63.

Carlet, A., and Cebrian, J. (2019). "Assessing the Effect of Perceptual Training on L2 Vowel Identification, Generalization and Long-Term Effects," in *A Sound Approach to Language Matters. In Honor of Ocke-Schwen Bohn*. Editors A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, and M. H. Sørensen (Aarhus, Denmark: Deparment of English, School of Communication and Culture, Aarhus University), 91–119.

Carlet, A., and Kivistö-de Souza, H. (2018). Improving L2 Pronunciation Inside and Outside the Classroom. *Ilha do Desterro.* 71 (3), 99–124. doi:10.5007/2175-8026.2018v71n3p99

Casillas, J. (2015). Production and Perception of the /i/-/I/ Vowel Contrast: The Case of L2-Dominant Early Learners of English. *Phonetica* 72, 182–205. doi:10.1159/000431101

Cebrian, J. (2006). Experience and the Use of Non-Native Duration in L2 Vowel Categorization. *J. Phon.* 34, 372–387. doi:10.1016/j.wocn.2005.08.003

Cebrian, J. (2007). "Old Sounds in New Contrasts: L2 Production of the English Tense-Tax Vowel Distinction," in Proceedings of the 16th International Congress of Phonetic Sciences. Editors J. Trouvain and W. J. Barry (Saarbrucken, Germany: University of Saarland), 1637–1640.

Cebrian, J. (2019). Perceptual Assimilation of British English Vowels to Spanish Monophthongs and Diphthongs. *J. Acoust. Soc. Am.* 145, EL52–EL58. doi:10.1121/1.5087645

Cebrian, J. (2021). Perception of English and Catalan Vowels by English and Catalan Listeners: A Study of Reciprocal Cross-Linguistic Similarity. *J. Acoust. Soc. Am.* 149 (4), 2671–2685. doi:10.1121/10.0004257

Cebrian, J., and Carlet, A. (2014). Second-Language Learners' Identification of Target-Language Phonemes: A Short-Term Phonetic Training Study. *Can. Mod. Lang. Rev.* 70 (4), 474–499. doi:10.3138/cmlr.2318

Cerviño, E., and Mora, J. C. (2009). Duration as a Phonetic Cue in the Categorization of /iː/-/ɪ/ and /s/-/z/ by Spanish/Catalan Learners of English. *J. Acoust. Soc. Am.* 125, 2764. doi:10.1121/1.4784683

Derwing, T. M., and Munro, M. J. (2015). A Prospectus for Pronunciation Research in the 21st Century. A Point of View. *J. Second Lang. Pronunciation* 1 (1), 11–42. doi:10.1075/jslp.1.1.01mun

Escudero, P. (2005). Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization. PhD Thesis. Amsterdam (Netherlands): Netherlands Graduate School of Linguistics.

Escudero, P., and Boersma, P. (2004). Bridging the Gap Between L2 Speech Perception Research and Phonological Theory. *Stud. Second Lang. Acquis.* 26 (4), 551–585. doi:10.1017/S0272263104040021

Escudero, P., and Chládková, K. (2010). Spanish Listeners' Perception of American and Southern British English Vowels. *J. Acoust. Soc. Am.* 128, EL254–EL260. doi:10.1121/1.3488794

Escudero, P., and Wanrooij, K. (2010). The Effect of L1 Orthography on Non-Native Vowel Perception. *Lang. Speech.* 53 (3), 343–365. doi:10.1177/0023830910371447

Faris, M. M., Best, C. T., and Tyler, M. D. (2018). Discrimination of Uncategorised Non-native Vowel Contrasts is Modulated by Perceived Overlap with Native Phonological Categories. *J. Phon.* 70, 1–19. doi:10.1016/j.wocn.2018.05.003

Flege, J. E. (1991). The Interlingual Identification of Spanish and English Vowels: Orthographic Evidence. *Q. J. Exp. Psychol. A.* 43 (3), 701–731. doi:10.1080/14640749108400993

Flege, J. E. (1992). "Speech Learning in a Second Language," in *Phonological Development: Models, Research, Implications*. Editors C. A. Ferguson, L. Menn, and C. Stoel-Gammon (Timonium, MD: York Press), 565–604.

Flege, J. E. (1995). "Second Language Speech Learning: Theory, Findings and Problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Editor W. Strange (Baltimore, MD: York Press), 233–277.

Flege, J. E. (2003). "Assessing Constraints on Second-Language Segmental Production and Perception," in *Phonetics and Phonology in Language Comprehension and Production*. Editors A. Meyer and N. Schiller (Berlin, Germany: Mouton de Gruyter), 319–355.

Flege, J. E. (2018). It's Input That Matters Most, Not Age. *Biling. Lang. Cogn.* 21 (5), 919–920. doi:10.1017/S136672891800010X

Flege, J. E., Frieda, E. M., Walley, A. C., and Fandazza, L. A. (1998). Lexical Factors and Segmental Accuracy in Second Language Speech Production. *Stud. Second Lang. Acquis.* 20 (2), 155–187. doi:10.1017/S0272263198002034

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 (4), 437–470. doi:10.1006/jpho.1997.0052

Flege, J. E., and Bohn, O.-S. (1989). The Perception of English Vowels by Native Spanish Speakers. *J. Acoust. Soc. Am.* 85 (1), S85. doi:10.1121/1.2027177

Flege, J. E., and Bohn, O.-S. (2021). "The Revised Speech Learning Model (SLM-R)," in *Second Language Speech Learning: Theoretical and Empirical Progress*. Editor R. Wayland (Cambridge, United Kingdom: Cambridge University Press), 3–83.

Flege, J. E., Schirru, C., and MacKay, I. R. A. (2003). Interaction Between the Native and Second Language Phonetic Subsystems. *Speech Commun.* 40 (4), 467–491. doi:10.1016/S0167-6393(02)00128-0

Fouz-González, J., and Mompean, J. A. (2020). Exploring the Potential of Phonetic Symbols and Keywords as Labels for Perceptual Training. *Stud. Second Lang. Acquis.* 1–32. doi:10.1017/S0272263120000455

Fullana-Rivera, N., and MacKay, I. R. (2003). "Production of English Sounds by EFL Learners: The Case of /i/ and /ɪ/," in Proceedings of the 15th International Congress of Phonetic Sciences, ICPhS. Editors M. J. Solé, D. Recasens, and J. Romero (Barcelona, Spain: Universitat Autònoma de Barcelona), 1525–1528.

Gilichinskaya, Y. D., and Strange, W. (2010). Perceptual Assimilation of American English Vowels by Inexperienced Russian Listeners. *J. Acoust. Soc. Am.* 128, EL80–EL85. doi:10.1121/1.3462988

Grenon, I., Kubota, M., and Sheppard, C. (2019). The Creation of a New Vowel Category by Adult Learners After Adaptive Phonetic Training. *J. Phon.* 72, 17–34. doi:10.1016/j.wocn.2018.10.005

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (2000). An Investigation of Current Models of Second Language Speech Perception: The Case of Japanese Adults' Perception of English Consonants. *J. Acoust. Soc. Am.* 107 (5), 2711–2724. doi:10.1121/1.428657

Hacquard, V., Walter, M. A., and Marantz, A. (2007). The Effects of Inventory on Vowel Perception in French and Spanish: An MEG Study. *Brain Lang.* 100 (3), 295–300. doi:10.1016/j.bandl.2006.04.009

Higgins, J. (2018). Minimal Pairs for English RP: Lists by John Higgins. Available at: http://minimal.marlodge.net/minimal.html (Accessed January 29, 2021).

Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (2000). Some Effects of Duration on Vowel Recognition. *J. Acoust. Soc. Am.* 108 (6), 3013–3022. doi:10.1121/1.1323463

Højen, A., and Flege, J. E. (2006). Early Learners' Discrimination of Second-Language Vowels. *J. Acoust. Soc. Am.* 119 (5), 3072–3084. doi:10.1121/1.2184289

Holt, L. L., and Lotto, A. J. (2006). Cue Weighting in Auditory Categorization: Implications for First and Second Language Acquisition. *J. Acoust. Soc. Am.* 119 (5), 3059–3071. doi:10.1121/1.2188377

IBM Corp (2017). *IBM SPSS Statistics for Windows*. Version 25.0. Armonk, NY: IBM Corp.

Jun, S.-A., and Cowie, I. (1994). Interference for 'New' and 'Similar' Vowels in Korean Speakers of English. *Ohio State Univ. Working Papers.* 43, 117–130.

Kivistö-de Souza, H., Carlet, A., Julkowska, I., and Rato, A. (2017). Vowel Inventory Size Matters: Assessing Cue-Weighting in L2 Vowel Perception. *Ilha do Desterro.* 70 (3), 33–46. doi:10.5007/2175-8026.2017v70n3p33

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 (6), 3959–3971. doi:10.1121/1.2999341

Kuhl, P. K., and Iverson, P. (1995). "Linguistic Experience and the Perceptual Magnet Effect," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Editor W. Strange (Timonium, MD: York Press), 121–154.

Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, MI: University of Michigan Press.

Levis, J. M. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. Cambridge, United Kingdom: Cambridge University Press.

Levy, E. S. (2009b). Language Experience and Consonantal Context Effects on Perceptual Assimilation of French Vowels by American-English Learners of French. *J. Acoust. Soc. America.* 125, 1138–1152. doi:10.1121/1.3050256

Levy, E. S. (2009a). On the Assimilation-Discrimination Relationship in American English Adults' French Vowel Learning. *J. Acoust. Soc. Am.* 126 (5), 2670–2682. doi:10.1121/1.3224715

Llisterri, J. (1995). "Relationships between Speech Production and Speech Perception in a Second Language," in Proceedings of the 13th International Congress of Phonetic Sciences. Editors K. Elenius and P. Branderud (Stockholm, Sweden: KTH and Stockholm University), 92–99.

Lobanov, B. M. (1971). Classification of Russian Vowels Spoken by Different Speakers. *J. Acoust. Soc. Am.* 49, 606–608. doi:10.1121/1.1912396

Major, R. C. (1987). English Voiceless Stop Production by Speakers of Brazilian Portuguese. *J. Phon.* 15 (2), 197–202. doi:10.1016/S0095-4470(19)30560-1

Mora, J. C., and Fullana, N. (2007). "Production and Perception of English/iː/-/ɪ/ and /æ/-/ʌ/ in a Formal Setting: Investigating the Effects of Experience and Starting Age," in Proceedings of the 16th International Congress of Phonetic Sciences 3. Editors J. Trouvain and W. Barry (Saarbrücken, Germany: University of Saarland), 1613–1616.

Morrison, G. S. (2008). L1-Spanish Speakers' Acquisition of the English /i /-/ɪ/ Contrast: Duration-Based Perception is Not the Initial Developmental Stage. *Lang. Speech.* 51 (4), 285–315. doi:10.1177/0023830908099067

Morrison, G. S. (2009). L1-Spanish Speakers' Acquisition of the English /i/-/ɪ/ Contrast II: Perception of Vowel Inherent Spectral Change1. *Lang. Speech.* 52 (4), 437–462. doi:10.1177/0023830909336583

Morrison, G. S. (2012). Perception of Natural Vowels by Monolingual Canadian-English, Mexican-Spanish, and Peninsular-Spanish Listeners. *Can. Acoust.* 40 (4), 29–40. doi:10.1121/1.3587895

Munro, M. J. (2008). "7. Foreign Accent and Speech Intelligibility," in *Phonology and Second Language Acquisition*. Editors Hansen Edwards, J. G., and Zampini, M. L. (Burnaby, Canada: Simon Fraser University), 193–218.

Munro, M. J. (Forthcoming 2021). On the Difficulty of Defining "Difficult" in Second-Language Vowel Acquisition. *Front. Commun.* doi:10.3389/fcomm.2021.639398

Munro, M. J., Flege, J. E., and MacKay, I. R. A. (1996). The Effects of Age of Second Language Learning on the Production of English Vowels. *Appl. Psycholinguist.* 17, 313–334. doi:10.1017/s0142716400007967

O'Brien, M. G. (2021). Ease and Difficulty in L2 Pronunciation Teaching: A Mini-Review. *Front. Commun.* 5, 1–7. doi:10.3389/fcomm.2020.626985

Piske, T., MacKay, I. R. A., and Flege, J. E. (2001). Factors Affecting Degree of Foreign Accent in an L2: A Review. *J. Phon.* 29 (2), 191–215. doi:10.1006/jpho.2001.0134

Rallo Fabra, L., and Romero, J. (2012). Native Catalan Learners' Perception and Production of English Vowels. *J. Phon.* 40, 491–508. doi:10.1016/j.wocn.2012.01.001

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

Rochet, B. (1995). "Perception and Production of Second-Language Speech Sounds by Adults," in *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Editor W. Strange (Timonium, MD: York Press), 379–410.

Strange, W. (1995). *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Timonium, MD: York Press.

Strange, W., Hisagi, M., Akahane-Yamada, R., and Kubo, R. (2011). Cross-Language Perceptual Similarity Predicts Categorial Discrimination of American Vowels by Naïve Japanese Listeners. *J. Acoust. Soc. Am.* 130, EL226–EL231. doi:10.1121/1.3630221

Strange, W., Levy, E. S., and Law, F. F. (2009). Cross-Language Categorization of French and German Vowels by Naïve American Listeners. *J. Acoust. Soc. Am.* 126 (3), 1461–1476. doi:10.1121/1.3179666

Thomas, E., and Kendall, T. (2007). NORM: The Vowel Normalization and Plotting Suite. Available at: http://ncslaap.lib.ncsu.edu/tools/norm/ (Accessed January 29, 2021).

Trubetzkoy, N. (1939). *Principles of Phonology*. Berkeley, CA: University of California Press.

Tyler, M. D. (2021). "Phonetic and Phonological Influences on the Discrimination of Non-Native Phones," in *Second Language Speech Learning: Theoretical and Empirical Progress*. Editor R. Wayland (Cambridge, United Kingdom: Cambridge University Press), 157–174. doi:10.1017/9781108886901.005

Tyler, M. D., Best, C. T., Faber, A., and Levitt, A. G. (2014). Perceptual Assimilation and Discrimination of Non-Native Vowel Contrasts. *Phonetica* 71, 4–21. doi:10.1159/000356237

Wang, X., and Munro, M. J. (1999). "The Perception of English Tense-Lax Vowel Pairs by Native Mandarin Speakers: The Effect of Training on Attention to Temporal and Spectral Cues," in Proceedings of the 14th International Congress of Phonetic Sciences. Editors J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Baily (San Francisco, CA: University of California, Berkeley), 125–129.

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., et al. (2009). Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-Language Users of English. *J. Cogn. Neurosci.* 22 (6), 1319–1332. doi:10.1162/jocn.2009.21272

Conflict of Interest: 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.

frontiers in Communication

ORIGINAL RESEARCH
published: 05 August 2021
doi: 10.3389/fcomm.2021.639398

On the Difficulty of Defining "Difficult" in Second-Language Vowel Acquisition

*Murray J. Munro**

Department of Linguistics, Simon Fraser University, Burnaby, BC, Canada

Hierarchies of difficulty in second-language (L2) phonology have long played a role in the postulation and evaluation of learning models. In L2 pronunciation teaching, hierarchies are assumed to be helpful in the development of instructional strategies based on anticipated areas of difficulty. This investigation addressed the practicality of defining a pedagogically useful hierarchy of difficulty for English tense and lax close vowels (/i ɪ u ʊ/) produced by Cantonese speakers. Unlike their English counterparts, Cantonese close tense-lax pairs are allophonic variants with [i u] occurring before alveolars and [ɪ ʊ] before velars. Each tense-lax pair represents a "phonemic split" in which members of a single L1 category are realized contrastively in L2. Despite evidence that English tense-lax distinctions are challenging for Cantonese speakers, no previous empirical work has closely considered the problem from the standpoint of vowel intelligibility across multiple phonetic contexts and in different words sharing the same rhyme. In a picture-based word-elicitation task, 18 Cantonese-speaking participants produced 31 high-frequency CV and CVC words. Vowels were evaluated for intelligibility by phonetically-trained judges. A series of mixed-effects binary logistic models were fitted to the scores, with vowel quality, phonetic context (rhyme) and word as factors, and length of Canadian residence and daily use of English as co-variates. As expected, the general hierarchy of difficulty for vowels that emerged (/i/ > /u/ > /ʊ/ > /ɪ/) was complicated by large differences across phonetic contexts. Results were not readily explicable in terms of transfer; moreover, different words with the same rhyme were not produced with equal intelligibility. The most serious modeling complication was the sizeable inter-speaker variability in difficulties, which could not be accounted for by model co-variates. Although some difficulties were roughly systematic at the group level, it is argued that establishing a pedagogically useful hierarchy on such data would prove intractable. Rather, L2 learners might be better served by assessment and instructional targeting of their individual problem areas than by a focus on errors predicted from hierarchies of difficulty.

Keywords: second language, intelligibility, vowels, Cantonese, phonology

OPEN ACCESS

Edited by:
Andrew Sewell,
Lingnan University, China

Reviewed by:
Ocke-Schwen Bohn,
Aarhus University, Denmark
Charles Nagle,
Iowa State University, United States

Correspondence:
Murray J. Munro
mjmunro@sfu.ca

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *08 December 2020*
Accepted: *08 March 2021*
Published: *05 August 2021*

Citation:
Munro MJ (2021) On the Difficulty of Defining "Difficult" in Second-Language Vowel Acquisition. Front. Commun. 6:639398. doi: 10.3389/fcomm.2021.639398

INTRODUCTION

A frequently cited goal of applied linguistics research is to unearth findings that advance classroom teaching practices. Although many research outcomes in the field are indeed pedagogically valuable, a connection between research and practice is not a *necessary* component of studies of second language (L2) learning. In fact, much L2 research is designed to resolve theoretical questions without direct relevance to instruction (White, 2018). Studies conducted without explicit pedagogical motivation may, of course, yield incidental implications for teaching, but that outcome is more likely the exception than the rule.

Specifically within the realm of pronunciation, Brinton (2018, p. 283) observes that "the nature of second language pronunciation research often precludes its application to the classroom." In addition to the theoretical emphasis mentioned above, at least two further reasons may be offered for the lack of relevance she describes. One of these is a long-standing pre-occupation with native-like accuracy in assessments of L2 production, in accordance with the "nativeness principle" (Levis, 2005, 2020). In their historical overview, Munro and Derwing (2011) pointed to a dearth of empirical pronunciation studies motivated by the opposing "intelligibility principle," despite repeated calls for an instructional focus on intelligible L2 speech going back at least as far as Sweet (1900). In a narrative summary of 75 mainly twenty-first century studies of instructional effectiveness, Thomson and Derwing (2015) reported that 63% were aligned with the nativeness principle. That orientation emphasizes *accentedness*, the degree to which speech is judged to differ from some particular variety (Munro and Derwing, 1995, 2020). For pedagogy, such a focus is deeply problematic in that empirical data conclusively demonstrate that native-sounding pronunciation is neither necessary for effective communication, nor likely to be achievable in the majority of L2 learners (for an overview, see Derwing and Munro, 2015; for a replication study, see Nagle and Huensch, 2020). Rather, even heavily-accented L2 speech can be highly *intelligible* (understood as the speaker intends) and highly *comprehensible* (easy to process). Given the incongruence between accentedness and the other dimensions, findings from nativeness-driven research are apt to create misapprehensions among teachers and pronunciation researchers. For instance, an intervention that leads to no change in learners' accentedness might be wrongly interpreted as evidence of the ineffectiveness of instruction, even though empirical findings show that L2 speech can become more comprehensible through instruction despite no change in accentedness (e.g., Derwing et al., 1998; Derwing and Rossiter, 2003).

A partial explanation for the quasi-independence of accentedness from the other two dimensions comes from a functional load (FL) account of error gravity (Catford, 1987; Levis and Cortes, 2008; Sewell, 2017), which assumes that certain linguistic contrasts play a greater role in distinguishing meanings than do other contrasts. Although FL itself is not a focus of the present study, the vowel sounds at issue (English close vowels) provide a useful example of the concept. The English /i/–/ɪ/ contrast is considered to have a high FL for several reasons, including that it distinguishes a large number of minimal pairs, such as *seat- sit*, *feel-fill*, and *leave-live*; that many of these pairs are easily confusable because of identical syntactic functions (*leave* and *live* can both function as verbs); and that the words at issue are of relatively high frequency. In contrast, the /u/–/ʊ/ distinction has a low FL because of its relative rarity, especially in commonly-used, confusable words. Examples include *Luke-look* (unlikely to be confused), *wooed-wood* (unlikely to be confused; first item is rare), and *tuque-took* (unlikely to be confused; first item, a type of hat, is largely unknown outside Canada). Findings from Munro and Derwing (2006) suggest that a failure to produce high FL distinctions can reduce comprehensibility more than a failure to produce low FL distinctions, yet the two types appear to increase accentedness to a similar degree. It follows that if intelligibility is an instructional goal, prioritizing high FL difficulties is likely to be more effective than an approach that treats all difficulties as equally problematic.

Returning to Brinton's (2018) point mentioned above, an additional factor in the lack of relevance of pronunciation research has been a preoccupation with differences in group-level performance in L2 speech production coupled with insufficient attention to inter-learner variability. In spite of twenty-First-century, rethinking of the uses of statistics in the social sciences (Larson-Hall, 2015), quantitative applied linguistics research relies heavily on null hypothesis statistical testing (NHST), which entails reporting of means, dispersion and effect sizes to evaluate between-group differences. In spite of the theoretical value of capturing such tendencies in pronunciation data, that emphasis appears to be at odds both with common anecdotal reports and with longitudinal data (Munro et al., 2015) indicating large inter-learner differences in pronunciation learning trajectories. Recent work by Wade et al. (2020) illustrates how an excessive focus on group means "masks the complexity of phonetic behaviors." Their native English VOT data showed that individual departures from group performance were often stable, indicating that between-speaker variability was not simply "noise," but reflected systematic individual differences from group norms that were not explicable on linguistic or sociolinguistic grounds. One might expect a parallel phenomenon in L2 production.

One issue in L2 pronunciation research that is commonly cited as pedagogically relevant is the determination of segmental "hierarchies of difficulty." A long-standing assumption is that theory-driven research should facilitate advance prediction of the L2 segmental difficulties faced by learners from particular L1s, and that teachers would benefit from such knowledge. In his skeptical coverage, Walz (1980), for instance, commented on the view among applied linguists that such research would lay the groundwork for good teaching materials and techniques. And years earlier, Moulton (1962) had argued that theoretical explanations of the sources of L2 pronunciation errors would lead to improved instruction. Nearly half a century later, research on L2 segmental difficulties experienced by particular L1 groups is ubiquitous and has expanded beyond L2 English. Even when only vowels are considered, the diverse range of L2s covered includes Danish (Bohn and Garibaldi, 2017), Dutch (Burgos et al., 2014), German (O'Brien and Smith, 2010; Darcy and Krüger, 2012; Nimz and Khattab, 2020), Polish (Sypiańska, 2016) and Japanese (Okuno and Hardison, 2016). Such studies attest to the health of the field of L2 pronunciation research and promise advancements in our understanding of the process of L2 speech acquisition. At the same time, judging by the continued popularity of texts presenting lists of difficult L2 phones organized by L1 (e.g., for English, Swan and Smith, 2001; Nilsen and Nilsen, 2010), there is still considerable enthusiasm among teachers and learners for predicting learner difficulties in advance (Munro, 2018). Whether that enthusiasm is warranted, however, is open to question.

Numerous accounts of the sources of difficulty in L2 phonological acquisition have been proposed. Among these, the best known concept is *transfer*, widely recognized as a fundamental influence on L2 speech production (Archibald, 2017). Attempts to characterize L2 phonology strictly in terms of transfer took the form of the strong version of the Contrastive Analysis Hypothesis (CAH). However, that approach failed at the levels of both segments (Brière, 1966; Wardhaugh, 1970) and prosody (Liu, 2021). In his detailed survey of twentieth - century L2 phonology, Eckman (2004) discusses how research has been extended to encompass other predictive factors as a way of improving upon transfer-based analyses. At the linguistic level these include similarity, markedness, and phonological constraints. It is also recognized that individual differences in learning are worthy of close attention (Edwards, 2017), and considerable work goes beyond linguistic considerations to include an array of psycho-social variables such as age of second language learning (AOL) and language experience (e.g., the Speech Learning Model; Flege, 1995), group affiliation (Gatbonton et al., 2005), cognitive processing ability (Darcy et al., 2015) and motivation (Saito et al., 2017; Nagle, 2018).

The present study concerns aspects of English close vowel acquisition that pose difficulty for Cantonese-speaking immigrants in Canada. In particular, I ask whether it is feasible to identify a pedagogically useful hierarchy of difficulty, given two contemporary concerns described above: the intelligibility principle and individual learner differences. For the purposes of this investigation it is irrelevant whether any particular theoretical model proves statistically more accurate than another. Rather, no specific theoretical framework will be assumed. Instead I use a *post-hoc* approach similar to that of Walz (1980) without seeking detailed explanations for patterns of performance. In this case, L2 productions of vowels in selected phonetic contexts are elicited from a cohort of Cantonese-speaking immigrants. These are evaluated to determine the relative difficulty of the vowels in the contexts at issue. Then I consider whether the *post-hoc* patterns of difficulty that emerge can be described in such a way that a language teacher could hypothetically apply the results directly to instruction given to Cantonese-speaking learners of English. In simple terms, my concern is with the "what," rather than the "why" of L2 vowel production.

Findings from some previous work on L2 vowel intelligibility call into question whether the proposed hierarchy is achievable. For instance, Munro et al. (1996) examined English vowel production by 240 Italian speakers. Their focus was the impact on production of the speakers' AOL (ranging from early childhood to young adulthood) with respect to both vowel nativeness and vowel intelligibility. The results for nativeness were theoretically interesting in that a strong, linear AOL effect emerged, with little indication of native-like productions among late teen and young adult learners. However, the AOL effect was much weaker for intelligibility. Canadian English /oʊ/ and /ɒ/, for instance, were produced with high intelligibility by nearly all AOL groups, despite their being judged as heavily "accented" in many cases. Moreover, individual speakers' performance on vowels varied widely, leading the authors to conclude that strong generalizations about Italian speakers' areas of difficulty could not be made without ignoring important individual differences.

Additional concerns arise from a longitudinal investigation of English vowel intelligibility in Mandarin and Slavic Language speakers (Munro and Derwing, 2008). Even without focused pronunciation instruction, vowel intelligibility increased for both groups during the first 6 to 8 months of Canadian residence. However, improvement was not uniform either across vowels or within groups. Instead, learning trajectories varied from one learner to another, and no full explanation for the differences could be offered.

Another potential complexity in developing a difficulty hierarchy concerns the generalizability of vowel learning across different phonetic contexts and in similar phonetic contexts across different words. While the simplest possible hierarchy would be a list of vowels in order of difficulty, evidence suggests that this is not achievable. For instance, in Thomson's (2011) study of the benefits of perceptual training on vowel production, the effects of the intervention, assessed in terms of improved intelligibility, generalized to some, but not all phonetic contexts. Interestingly, the context at issue was the consonant *preceding* the trained vowel. That outcome supports the view that L2 vowel intelligibility is at least to some degree context-dependent, and that a difficulty hierarchy would have to specify different levels of difficulty according to context. A related issue is whether generalization can occur across the similar phonetic contexts of rhyming words, such as *seat*, *heat*, and *feet*. That question, which has yet to be closely examined in an intelligibility study, will be given close attention here.

The present study uses a descriptive approach to address the problem areas identified above: it assesses vowel intelligibility across a cohort of L2 English speakers from a shared L1 background (Cantonese), taking into account different phonetic contexts and different words that are phonetically similar. This close examination of individual learners' productions will provide insight into the feasibility of characterizing learners' performance in terms of hierarchy of vowel difficulty. Although intelligibility will be considered at the group level, considerable attention will be focused on how well group-based generalizations may obscure important individual variability in individual speakers. In particular, highly idiosyncratic performance would pose a serious problem for establishment of a useful hierarchy.

METHODS

Focus of the Investigation

The vowels / i ɪ u ʊ / were selected for study because of their different distributions in English and Cantonese. While the four are phonemically contrastive on the basis of quality in Canadian English, Cantonese makes only a two-way phonemic distinction in which tense and lax vowels are allophonic variants. Chan and Li (2000) state that front /i/ is realized as [iː] before labials and alveolars and as [ɪ] before velars, and that back /u/ is produced as [uː] before alveolars and as [ʊ] before velars. Roughly, then, the English rhymes /it/, /ut/, /ɪk/, /ʊk/ match sequences that occur in Cantonese, while /ɪt/, /ʊt/, /ik/, /uk/ do not. From the standpoint of classical Contrastive Analysis, the problem faced by learners is referred to as a "split category" or "allophonic split" (Brière, 1966; Eckman et al., 2003). In addition, Cantonese syllable codas are restricted to nasals and voiceless plosives. Therefore, no English vowel + /d/ rhymes have matches in Cantonese. Finally, Zee (1991) observes that both /ɪ/ and /ʊ/ are relatively lowered, suggesting that they are close but not exact matches for their Canadian English counterparts. All three of these differences will be considered in this investigation.

Research on L2 English vowel production by Cantonese speakers points to difficulties with close vowels. Meng et al. (2007), for example, note that common errors include pronunciation of /ɪ/ as /i/ in words like *sit* and of /ʊ/ as /u/ in *full*. However, intervention studies indicate that adult L2 speakers from Cantonese backgrounds are capable of improved intelligibility after training. Wong (2015) reported more accurate production of close tense and lax vowels after high variability perceptual training (HVPT), though improvement was greater in the front pair than in the back one. Wong (2013) also found intervention to be beneficial, but neither study examined contextual effects closely, and neither considered the consistency of vowel accuracy across different words with identical rhymes (e.g., *seat*, *heat*).

Speakers

English is well-established in Hong Kong, and the characteristics of the Hong Kong variety of English have been discussed by Hung (2000). Nonetheless, individual speakers' English proficiency varies widely. Some immigrants to Canada from Hong Kong have little experience using English for social or work-related communication and regard themselves as having insufficient English skills to function in Canadian society. They therefore view themselves as English learners and may choose to enroll in English classes after arrival. Speakers from that demographic were targeted for this investigation.

Speech data were collected from 18 native speakers of Cantonese (10 female, 8 male) recruited via post-secondary student e-mail lists and word-of-mouth on campuses in the lower mainland of British Columbia (i.e., Metro Vancouver). Informational materials called for participants from Hong Kong who self-identified as second-language speakers of English, who were between 19 and 30 years of age with normal hearing, and who had arrived in Canada at age 15 or later. A CAD$25 payment was given as an incentive. One participant who initially enrolled was dropped from the study (and replaced) for not meeting the age criterion.

Prior to recording, each speaker completed a language background questionnaire (LBQ). The speakers had been born and raised in Hong Kong and had moved to Canada at a mean age of arrival (AOA) of 18 years [range: 15–25 years]. Mean age at the time of the study was 23 years [range: 19–28], and length of residence (LOR) in Canada averaged 4.9 years [range: 9 months−6.9 years]. All had grown up in households with native Cantonese-speaking parents who used Cantonese as their household language. Though all speakers had studied English from the beginning of their schooling, as was typical in Hong Kong at the time, none reported regular use of English for social purposes until arrival in Canada. Eleven speakers had attended Canadian ESL classes, with a mean duration of 11.3 months, and two speakers reported completion of a 1–2 mo. course in English pronunciation. At the time of the study all had enough proficiency in English to be enrolled in English-speaking Canadian post-secondary institutions, where they were registered for degree credit. On the LBQ, each speaker completed a grid to estimate personal use of spoken English and Cantonese during each morning, afternoon and evening in a typical week. On average, the speakers reported using English 26% of the time [range: 0–93%]. The cases of 0% English use [$n = 2$] are attributable to recruitment during the summer semester when some speakers were not enrolled in classes and reported exclusive use of their L1 at home and work.

For reliability assessment of the judges' evaluations (see below), recordings of two native English speakers (1 female, 1 male) were also used. These were randomly selected from a database of 18 speakers of Canadian English, also post-secondary students. All 20 speakers in the study passed a pure-tone hearing screen (250–4,000 Hz at 20 $\mathrm{dB_{HL}}$).

Test Items

The test items were English CV(C) words covering the 14 actually-occurring English rhymes consisting of the vowels /i/, /ɪ/, /u/, and /ʊ/ in open syllables and in checked syllables before /t/, /k/, and /d/. These combinations were selected because they appear to be an optimal set for this study, given that Cantonese allows rough equivalents to some of them, but not to others. For instance, it is possible to examine whether the L2 speakers perform better on the "matching" /it/ rhyme than on "non-matching" /ɪt/. Target words with /d/ as a coda were also included. These are interesting in that final /d/ is absent in Cantonese, but rhymes containing (homorganic) final /t/ are possible. It is unknown whether the speakers might be able to generalize knowledge of Vt rhymes to rhymes matching in place of articulation but not in voicing. Although other syllable codas could have been included, the number of tokens needed to cover the above combinations is already large. Expanding the list would

TABLE 1 | Target words from the picture-naming task.

Vowel	Coda	Words
/i/	#	key see tea
	/t/	feet, heat, seat
	/k/	cheek, speak
	/d/	feed, read
/ɪ/	/t/	hit, sit
	/k/	chick, kick, sick
	/d/	kid, lid
/u/	#	Sue, two
	/t/	boot, suit
	/k/	Luke, tuque
	/d/	food
/ʊ/	/t/	foot, put
	/k/	book, cook, look
	/d/	good, wood

add more work to an already onerous task for the judges (see section Judges' Evaluations).

To successfully elicit productions it was necessary to find familiar words that could be easily depicted in drawings. For that reason, it was not possible to generate a set of suitable words such that equal numbers of each rhyme were represented; nor was it possible to fully match words for initial consonants or to incorporate a complete set of minimal pairs. The final word list is given in **Table 1**.

Speaking Tasks

During individual recording sessions, the speakers sat in an audiometric booth wearing a Shure Beta 54 head-mounted microphone connected to a Symetrix 302 microphone preamplifier and an HHB Professional digital recorder (CDR-830) set to a sampling rate of 44.1 kHz with 16-bit resolution.

A picture naming task was used to elicit the target words, so that the productions could be assumed to be based on stored representations developed through experience with English. It was expected that avoiding written forms would minimize orthographic influences. Participants named the target words from a randomly-ordered set of drawings, each accompanied by the first letter of the word as a cue. For instance, a drawing of a chair, along with "S" was used to elicit *seat*, and an arrow pointing to a foot with "F" indicated *foot*. The drawings were mounted on individual 8.5″ by 11″ cards that were randomized by shuffling. Each card included a stimulus number (1 to 31). During an unrecorded practice session, a research assistant went through the entire pack, and speakers guessed each target item. They then practiced producing the stimulus number and the target item in the following sentence frame: "Number __. The next word is __." The stimulus number was used for later sorting of the randomized recorded items. In case of a wrong guess, the speaker was required to guess again until the correct word was produced. *Sue* and *Luke* were depicted by a female and male face, respectively, and the assistant explained that proper names beginning with "S" and "L" were required. During the practice elicitations it was discovered, as expected, that none of the Cantonese speakers were familiar with the word *tuque* (/tuk/, also spelled *toque*), even though this word is widely used in Canada to refer to a winter hat and was produced without hesitation by all native English speakers in the database. The research assistant therefore modeled the word and allowed each speaker to practice it a few times prior to recording.

On completion of the practice session, the pack of cards was shuffled. Each participant then recorded the full stimulus set three separate times, with a shuffling of the cards after each run-through. In case of a hesitation, a repetition was elicited (fewer than 1% of cases). During each run-through, extra items (not used in any analyses) were added to the beginning and end of the pack to minimize effects of list intonation in production. The picture elicitation task was followed by additional speaking and listening tasks to be reported elsewhere.

Judges' Evaluations

The recorded words from all 20 speakers were extracted digitally and saved as peak-normalized, single-word audio files. Prior to evaluation by the judges, the author informally pre-screened the recordings and judged that some vowel productions resembled English vowels other than the actual target items. A majority of productions were heard to fall into one of the seven English categories / i ɪ eɪ ɛ u ʊ oʊ /. However, some tokens were ambiguous, falling between two of the categories, and a small number (<1%) were indeterminate.

The listener-judges were four linguistically-trained assistants, all familiar with IPA, and all having passed the same pure-tone hearing screen as the speakers. During individual listening sessions (nine per judge) held on different days, each judge heard the recorded words one at a time and identified each vowel in a forced-choice task (as in Munro and Derwing, 2008). The judges sat in a sound-treated room, each hearing a different random presentation via custom-designed playback software through AKG K141 professional studio headphones. They responded by clicking computer buttons marked with the phonetic symbols representing the seven vowel categories described above. An extra button marked "Other" was available for vowels not heard as one of the possible choices. Judges were advised that a vowel might sound foreign-accented, yet still clearly belong to one of the possible categories. In case of ambiguity, they were told to choose the vowel symbol that best represented the production. The task was self-paced: once a response was entered, the computer automatically played the next item. A practice session of about 5 min was given, during which the listening volume was adjusted to a comfortable level. Over the nine sessions, each judge provided a total of 1,860 responses.

RESULTS

Reliability of Judges

For assessment of inter- and intra-judge reliability and for the intelligibility analyses that follow, the judges' responses were coded as binary values to indicate on-target (i.e., "correct

TABLE 2 | Type III tests of fixed effects (vowel).

Effect	*df* (Num)	*df* (Den)	*F*	*Pr* > *F*
Vowel	3	1,866	188.56	<0.0001
LOR	1	14.62	4.33	0.0554
% USE	1	14.84	2.79	0.1159

category") or off-target ("incorrect category") identifications. Inter-judge agreement was operationalized as the number of times the judges agreed with each other on whether or not a token was on-target. At least 3 out of 4 judges agreed on 91% of the Cantonese speakers' tokens, with 4-way agreement on 70%. These results are very similar to those reported by Munro and Derwing (2008). Fleiss' Kappa was computed at κ = 0.56, $p < 0.001$, CI [0.542, 579], indicating moderate agreement according the Landis and Koch (1977) benchmarks. It should be noted that high reliability was not expected here because of the ambiguity in many of the productions that was noted above. That is, many productions seemed to straddle more than one vowel category such that a judge would have difficulty making a straightforward classification.

Intra-judge consistency was assessed by requiring the judges to re-evaluate 84 tokens during a separate listening session held on a different day after the original evaluations were completed. Consistency was computed by determining the percentage of times each judge evaluated items the same way (on-target/off-target) both times. Consistency ranged from 89 to 93% across the four judges. Given the ambiguous nature of some of the productions, this was deemed very acceptable.

Intelligibility

For the Cantonese speakers' vowels, the mean correct ID rate was 73%. For the native English tokens, which were included strictly as a means of verifying correct use of the vowel symbols, mean correct ID was 99%, suggesting a high level of accuracy in symbol use by the judges. In the analyses that follow, the results for only the Cantonese productions will be discussed.

On-target identifications on the open-syllable vowels /i/ and /u/ were 100%; therefore, scores on these items were excluded from statistical computations. The data from the checked-syllable productions were submitted to generalized mixed-effects modeling with Glimmix in SAS 9.4 (2018, SAS Institute, Cary NC), using the Kenward-Roger approximation for degrees of freedom. Speakers and judges were entered as random effects, with percentage of weekly English use as reported by participants in the LBQ (%USE) and length of residence in months (LOR) as co-variates. Vowel, Rhyme and Word were fixed effects. Because the fixed factors were not independent, it was necessary to compute a separate model for each one. Except where indicated otherwise, Tukey-Kramer *post-hoc* tests were used for follow-up analyses, with overall $p < 0.05$ as the criterion for significance.

Vowel Comparisons

For checked-syllable vowels, the order of intelligibility was /i/ (92%) > /u/ (82%) > /ʊ/ (59%) > /ɪ/ (51%). Mixed-effects modeling (summarized in **Table 2**) yielded a significant effect of Vowel, with a non-significant contribution of %USE and a marginally non-significant contribution of LOR.

Post hoc tests revealed significant differences for all pairwise combinations, according to the ordering given above. In sum, the two lax vowels were produced significantly less intelligibly than their tense counterparts. For the tense pair, the front vowel was significantly more intelligible than the back vowel, while the reverse was true for the lax pair.

Rhyme Comparisons

Performance on the checked-syllable rhymes is summarized in **Table 3**, according to speaker and rhyme. Speakers are ordered vertically from highest to lowest mean correct identifications, and rhymes are ordered from left to right. For individual speakers, intelligibility rates ranged from 89 to 58%, while the range for individual rhymes was from 97% for /id/ to only 33% for /ʊk/. Also indicated in the table (by "+") are cases in which a particular speaker could be considered to have "acquired" a particular rhyme, according to the 80% correct criterion suggested by Carlisle (1998). Across speakers the total number of intelligible rhymes (second-to-last column) ranged from 10 to 3, out of a possible 12. The number of speakers reaching criterion on particular rhymes (second-to-last row) ranged from 17 for /id/ to just a single speaker for /ʊk/.

Model results for Rhyme, shown in **Table 4**, were parallel to those for the Vowel analysis. They indicated a significant effect of Rhyme, and non-significant contributions of LOR and %USE.

Post hoc comparisons indicated that performance on /id/ was significantly better than on all other rhymes except /ik/ while the mean score on /ʊk/ was significantly worse than on all other rhymes. The number of possible pairwise comparisons for the 12 rhymes is 66; however, only a subset of the outcomes—those most pertinent to the questions addressed in this study—will be considered here. First, **Table 5** compares performance on Vt and Vk rhymes, with the left column (matching VCs) showing rhymes that have approximate counterparts in Cantonese and the right column (non-matching) showing those that do not. In none of the 4 pairwise comparisons did performance on matching VCs statistically exceed scores on non-matching VCs. In fact, the significant differences that emerged in three comparisons all favored the non-matching VC, while for /ɪk/ and /ɪt/, intelligibility was relatively low in both cases (52 and 44%), and the difference failed to reach significance.

Table 6 shows performance on Vt and Vd rhymes for each of the four vowels. In three of the four cases, Vd performance significantly exceeded Vt, with no statistical difference for the fourth, /ut/ vs. /ud/.

Word Comparisons

The results of the mixed-effects analysis for Word are given in **Table 7**. Again, these parallel the results of the previous modeling, with Word yielding a significant effect, and non-significant effects for the two co-variates.

The number of possible pairwise word comparisons (325) is very large, but the comparisons of interest for the purposes of this study are the smaller subset of words with identical

TABLE 3 | Intelligibility of vowels by rhyme and speaker.

SPKR	LOR[a]	%USE[b]	id	ik	uk	ʊd	it	ut	ʊt	ud	ɪd	ɪk	ɪt	ʊk	Acquired rhymes[c]	Mean %-ID
c009[ESL]	6.8	72	+	+	+	+	+	+		+	+	+	+		10	89
c014	5.7	33	+	+	+		+	+	+	+	+			+	9	85
c018	6	16	+	+	+	+	+	+		+					7	84
c015	6.8	16	+	+	+	+	+		+		+		+		8	81
c016[ESL]	3.4	32	+	+		+	+				+	+			6	79
c022[ESL]	5.9	0	+	+	+	+	+				+		+		7	78
c012[ESL]	6.8	93	+	+	+	+						+			5	78
c020[ESL]	6.7	71	+	+	+	+	+		+						6	72
c019	3.1	48	+	+	+		+	+		+					6	72
c013[ESL]	6.7	6		+		+			+						3	71
c021	6	2	+	+	+	+			+						5	69
c005[P]	0.8	33	+	+	+	+	+	+		+					7	68
c004[ESL]	3.8	0	+	+	+		+			+					5	68
c002[ESL]	4.6	17	+	+	+	+			+						5	67
c003	2.7	6	+	+	+	+	+			+					6	65
c008[ESL]	5.6	9	+	+	+	+		+							5	65
c007[ESL]	4.9	4	+			+			+						3	64
c006[ESL,P]	2.8	13	+		+					+					3	58
Total speakers			17	16	15	14	11	6	7	8	5	3	3	1		
Mean %-ID			97	93	88	86	81	70	69	62	55	52	44	33		

"+" indicates that the vowel in the rhyme was accurately identified at least 80% of the time by the four judges. [a]Length of residence in Canada in years. [b]Weekly use of English. [c]Total rhymes (maximum 12) showing 80% correct ID or higher. [ESL]Participant had received ESL instruction in Canada. [P]Participant had received focused instruction in English pronunciation.

TABLE 4 | Type III tests of fixed effects (rhyme).

Effect	*df* (Num)	*df* (Den)	*F*	*Pr* > *F*
Rhyme	11	1,858	80.18	<0.0001
LOR	1	14.66	4.36	0.0547
% USE	1	14.84	2.78	0.1167

TABLE 5 | Intelligibility of vowels in syllables according to approximate match or non-match of rhymes in English and Cantonese.

VC [matching]	Results*	VC [non-matching]
it (81)	<	ik (93)
ɪk (52)	=	ɪt (44)
ut (70)	<	uk (88)
ʊk (33)	<	ʊt (69)

** < indicates that performance on the matching VC was significantly worse; = indicates no statistical difference between matching and non-matching.*

TABLE 6 | Intelligibility of vowels in Vt and Vd rhymes.

Vt	Results*	Vd
it (81)	<	id (97)
ɪt (44)	<	ɪd (55)
ut (70)	=	ud (62)
ʊt (69)	<	ʊd (86)

** < indicates that performance on the Vt target was significantly worse; = indicates no statistical difference between Vt and Vd targets.*

TABLE 7 | Type III tests of fixed effects (word).

Effect	*df* (Num)	*df* (Den)	*F*	*Pr* > *F*
Word	25	1,844	37.08	<0.0001
LOR	1	14.67	4.37	0.0543
% USE	1	14.67	2.77	0.1171

rhymes. Accordingly, a total of 17 Bonferroni-adjusted *t*-tests (with overall $p < 0.05$) were performed, the results of which are summarized in **Table 8**. Here, the central issue is whether vowels in words with identical rhymes showed approximately equal levels of accuracy. As can be seen, nearly half (eight) of the pairwise comparisons yielded significant differences. The *heat* and *chick* vowels were produced less intelligibly than the vowels in the other two words with identical rhymes, and the *cheek*, *foot*, *Luke* and *lid* vowels were less intelligible than those in the rhyming words *speak*, *put*, *tuque*, and *kid,* respectively.

Individual Speaker Performance

Individual speaker data on words with identical rhymes revealed noteworthy differences in words across speakers. Illustrative examples are provided in **Figures 1–3**. **Figure 1** shows individual performance on *kid* and *lid*. Note that the group means

TABLE 8 | Statistical comparisons of %-correct ID for words with identical rhymes.

Rhyme	Results*
it	[seat = feet] > heat
ɪt	hit = sit
ik	speak > cheek
ɪk	[sick = kick] > chick
ut	boot = suit
ʊt	put > foot
uk	Tuque > luke
ʊk	book = cook = look
id	feed = read
ɪd	kid > lid
ʊd	good = wood

*No difference shown by "="; significant difference (Bonferroni $p < 0.05$) shown by ">".

indicate significantly better performance on *kid*, and in fact, no speaker performed better on *lid*. However, several differences are evident across individual speakers. Three of them showed perfect performance on both words (c005, c009, and c014) and three showed 0% (or nearly 0%) on both (c003, c004, and c005). Discrepancies of 50 to 100 points between the two words occurred for seven speakers, and in those cases, the speaker performed better on *kid* than on *lid*.

No statistical difference was observed in performance on *hit* and *sit*. The individual speaker data shown in **Figure 2** reveal that eight speakers showed either identical or similar (within 10 points) correct ID scores for these words, suggesting comparable difficulty. However, the remaining data present a very different picture. Across all speakers, performance ranged from 0 to 100% for *sit* and from 0 to 83% for *hit*, suggesting slightly greater difficulty for the latter, while individual scores revealed a wide disparity in which word posed a greater challenge. Speaker c016 performed perfectly on *sit*, while scoring 0% on *hit*, with

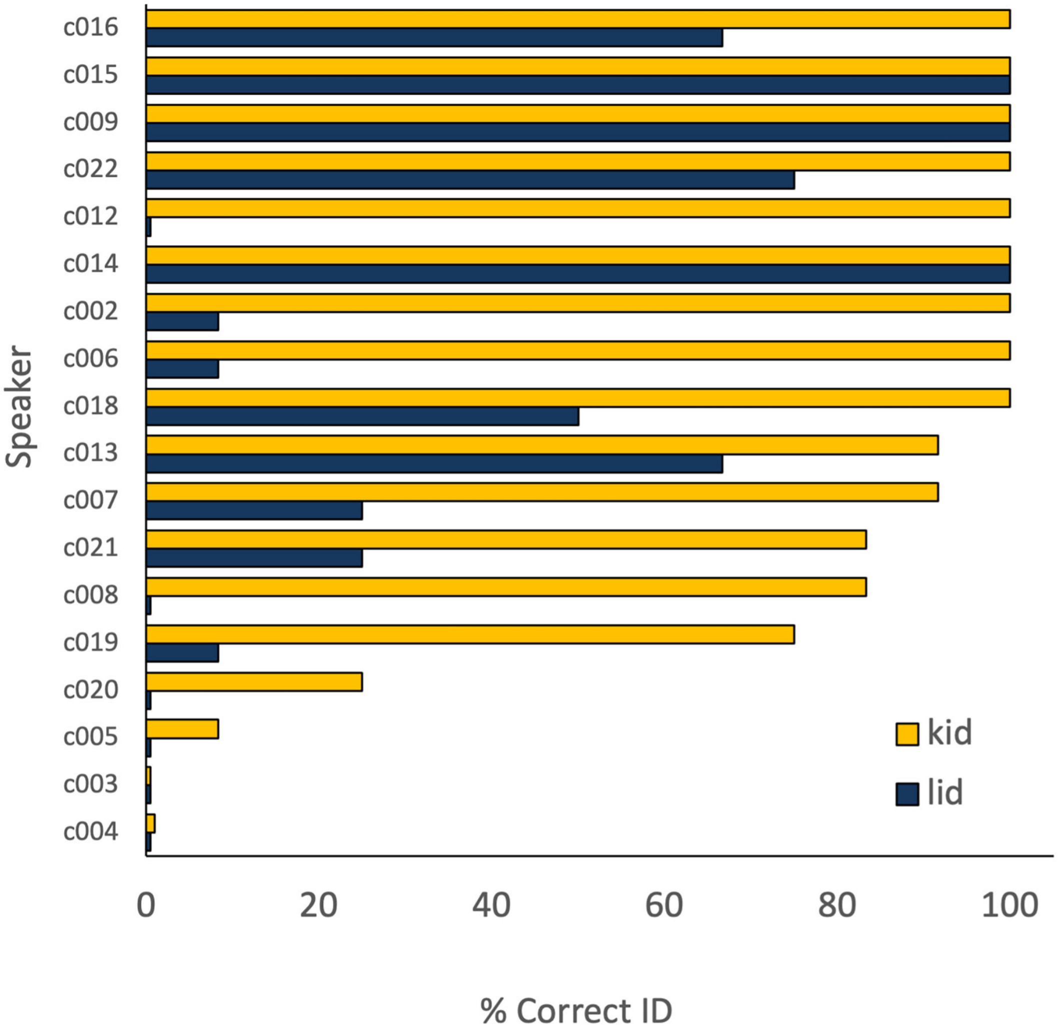


FIGURE 1 | Individual speaker performance (% intelligibility) on *kid* and *lid*.

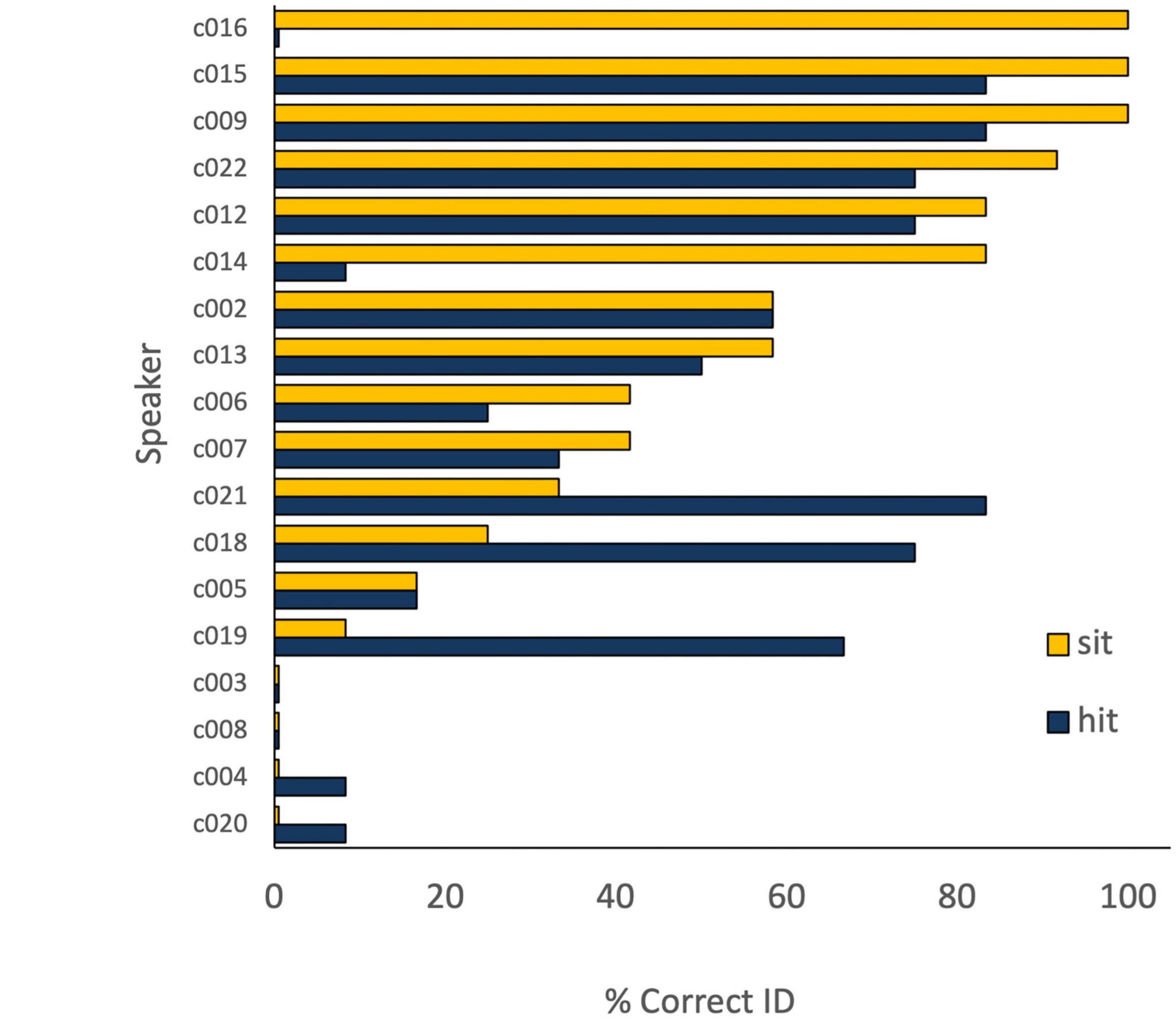


FIGURE 2 | Individual speaker performance (% intelligibility) on *hit* and *sit*.

speaker c014 also showing a sizeable discrepancy in the same direction. The pattern was reversed for speakers c021, c018, and c019, all of whom performed much better on *hit* than on *sit*. It appears, then, that the absence of a *post-hoc* statistical difference between the words was not the result of consistent performance across speakers. Rather, it was due to evaluating the average performance of speakers who showed opposite patterns of difficulty.

Similarly, for *sick* and *kick* no statistical difference was observed in vowel intelligibility. But as illustrated in **Figure 3**, that generalization does not accurately capture individual performance. Speakers c015, c018, and c002 all showed much better intelligibility for *sick*, while c012 and c014 performed much better on *kick*. Once again the lack of a statistical difference belies considerable variability among speakers.

Assuming a criterion of $p < 0.05$, none of the mixed-effects analyses yielded significant effects of either %USE or LOR. However, LOR just missed significance in all cases. For that reason, the data were examined impressionistically for any general tendences. With respect to rhymes, two of the top performers in **Table 3** (12 correct rhymes each) were tied for the longest LOR (6.8 years); however, a speaker with only slightly less time in Canada (6.7 years) was tied for poorest performance with two other speakers. Moreover, the speaker with the shortest LOR performed better than 12 of the other speakers. Daily use of English was also relatively high (72%) for the top performer (c009). However, c015 and c014, both of whom reported relatively low usage (16%, 33%), were the next two best performers.

Vowel Confusions

Confusion patterns for the vowel productions are shown in **Table 9**. Nearly all non-target /i/ productions were classified by the judges as /ɪ/, irrespective of coda consonant. However, while non-target /ɪ/ was nearly always /i/ before /t/ and /d/, it was rarely so before /k/. Instead, target /ɪk/ was judged to contain

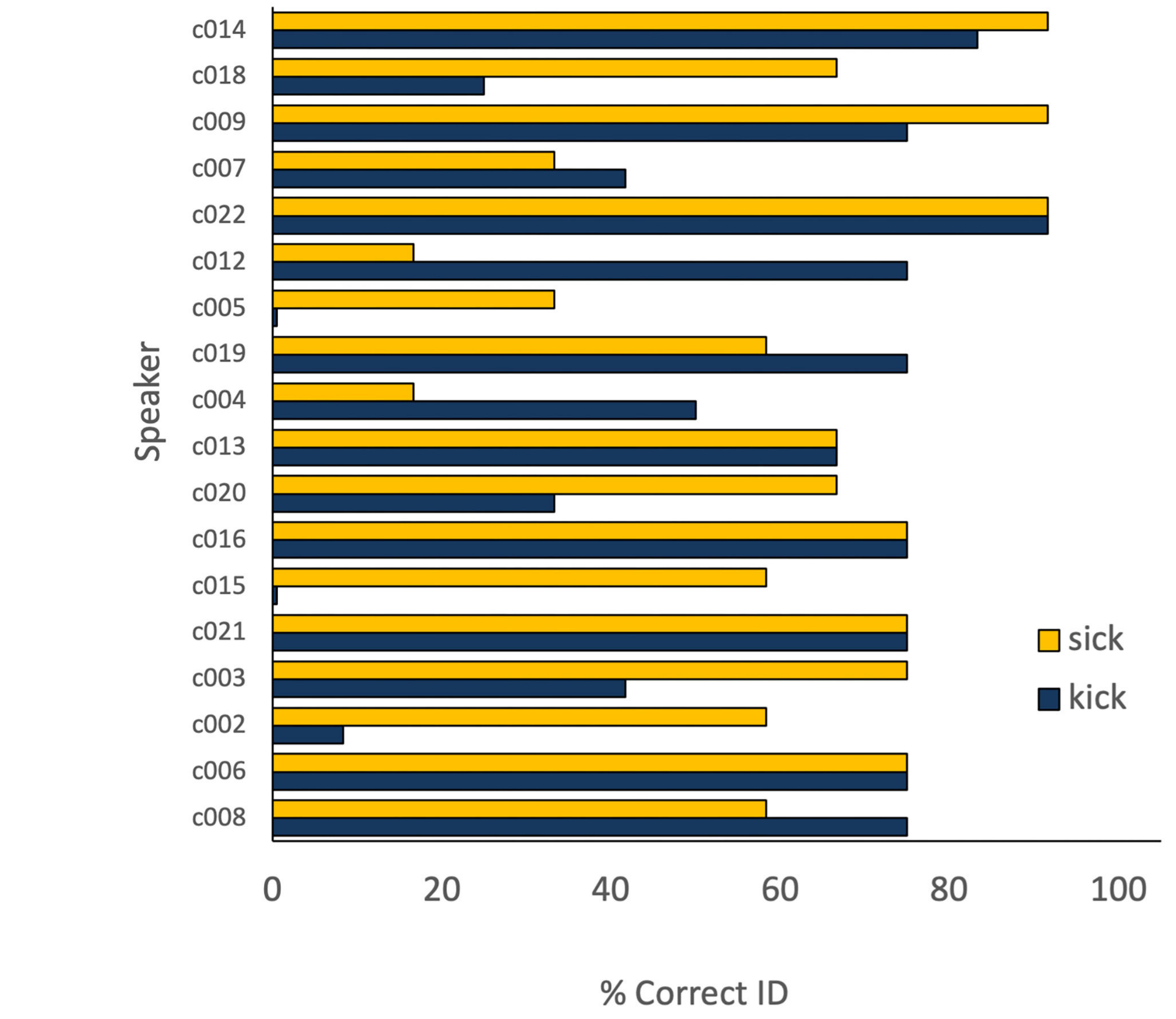


FIGURE 3 | Individual speaker performance (% intelligibility) on *sick* and *kick*.

/e/ in 25% of cases, and some other vowel—most often /ϵ/—in another 14%. A parallel but larger discrepancy occurred for the back targets. Nearly all incorrect /u/ tokens were judged to be /ʊ/, and although non-target /ʊ/ items tended to be /u/ before /t/ and /d/, the majority of /ʊk/ targets (58%) were judged to contain /o/.

DISCUSSION

The main question at issue in this study was whether or not it is feasible to specify a pedagogically useful hierarchy of difficulty for English close vowel production by Cantonese speakers of English. In this case, "difficulty" was operationalized in terms of intelligibility as assessed by a panel of four judges, and English tense-lax contrasts were selected for consideration because they are known to pose a challenge for Cantonese speakers due to different distributions in the two languages. "Pedagogically useful" refers to whether or not an empirically-based hierarchy could be expected to facilitate instruction by prioritizing certain vowels or vowel + final consonant combinations for attention from teachers and learners.

Difficulty by Vowel

From a statistical standpoint, a number of generalizations can be made to characterize the group performance observed in the study. First, open-syllable /i/ and /u/ were produced with nearly perfect intelligibility. Second, in checked syllables the following ordering emerged, with all differences being significant: /i/ > /u/ > /ʊ/ > /ɪ/. Thus, the two lax vowels were more difficult than the tense ones, and /ɪ/ proved most difficult of all, just as in Munro and Derwing's (2008) study of Mandarin and Slavic language speakers.

Difficulty by Rhyme

It was unsurprising that vowels in some rhymes were more difficult than in others, given Cantonese phonotactics. However,

TABLE 9 | Confusions (%) by target rhyme.

Target Rhyme	Vowel identified (by judges)						
	i	ɪ	e	u	ʊ	o	Other
it	83	14	2	0	0	0	1
ɪt	46	50	2	0	0	0	3
ik	93	5	1	0	0	0	1
ɪk	5	56	25	0	0	0	14
ut	0	1	0	73	18	2	6
ʊt	0	0	0	27	72	0	1
uk	0	0	0	89	9	0	1
ʊk	0	0	0	3	38	58	1
id	97	3	0	0	0	0	0
ɪd	40	60	0	0	0	0	0
ud	0	0	0	65	33	0	2
ʊd	0	0	0	11	87	0	2

Shaded boxes show the largest ID category.

the patterns of difficulty did not fit a straightforward transfer-based explanation. In particular, vowels in "matching" rhymes (i.e., those that were roughly equivalent in both languages), did not exhibit higher rates of intelligibility. For instance, /ʊk/ (matching) exhibited much worse intelligibility than did /ʊt/ (non-matching), and both /ɪk/ (matching), and /ɪt/ (non-matching) showed similarly poor intelligibility overall. Nor could transfer be invoked to account for performance on vowels in Vd syllables, which do not occur in Cantonese. In fact, contrary to what a transfer account would suggest, these vowels were produced more intelligibly overall than the ones in matching Vt syllables.

The vowel confusion patterns (**Table 9**) indicated interesting asymmetries in performance. For the front targets, /ɪ/ was often produced as /i/ before /t/, but sometimes as /e/ (i.e., as lower than the target) before /k/. For back vowels, /ʊ/ was judged to be /u/ in most instances before /t/, but very often as /o/ (again lower) before /k/. The pattern of apparent tongue-lowering seen here brings to mind an observation made by Brière (1966) that a full understanding of transfer effects requires attention to relatively subtle phonetic details. In particular, as noted earlier, Zee (1991) reported relatively low tongue positions for the two lax vowels in Cantonese. It seems plausible that transfer of Cantonese articulatory patterns could lead to a perception on the part of the judges of lower-than-target English vowel categories. Nonetheless, this phonetic detail alone is insufficient to explain the different patterns between speakers. The reason why some speakers tended to produce a more /o/-like /ʊ/ while others did not must pertain to some aspect of individual learners' phonological knowledge. An examination of these speakers' L1 productions might prove useful in addressing this issue.

Difficulty by Word

In contrast to the statistical significance of the rhyme effect, the wide-ranging word effect in these data was unexpected. Specifically, vowels in words sharing the same rhyme frequently showed statistically different degrees of difficulty. This was true for each of the four vowels: for instance, /i/ in *heat* was produced less intelligibly than /i/ in *seat* or *feet*, /ɪ/ in *chick* was less intelligible than /ɪ/ in *sick* or *kick*, /u/ was less intelligible in *Luke* than in *tuque*, and /ʊ/ was less intelligible in *foot* than in *put*. It is possible to speculate about multiple reasons for this patterning, including effects of the initial consonant (as in Thomson, 2011), differential frequency of experience with the words, or prior encounters through borrowings of the same or similar words into Cantonese. None of these explanations can be verified as a causal factor in this study. More importantly, however, pursuing such lines of explanation may ultimately prove fruitless because of the complications of individual speaker performance described in the next section.

Difficulty by Speaker

In spite of the group-based statistical tendencies discussed above, the most striking outcome of the study was the high degree of variability in performance across individual speakers. First, the between-speaker differences on different rhymes (**Table 3**) pose a serious problem for development of a detailed difficulty hierarchy. These differences do not correlate significantly with the socio-linguistic variables included in the study, i.e., use of English or length of residence in Canada. Nor can they be explained by an appeal to a "natural order of acquisition" for vowels in rhymes. If the latter applied, an implicational hierarchy should be visible within **Table 3**, such that success on some particular rhyme strongly predicts success on some statistically easier rhyme. However, indications of such patterning are exceedingly weak. Speaker c016, for instance, performed well on two of the very difficult rhymes (/ɪd/ and /ɪk/) but not on several that were statistically easier (e.g., /uk/, /ut/, /ud/). A similar point can be made about speakers c022 and c012. And speaker c014, who was the sole speaker to produce /ʊk/ rhymes at criterion, failed to produce an intelligible /ʊd/. Yet /ʊd/ was one of the easiest rhymes, having been produced intelligibly by 14 of the 18 participants.

The prognosis for establishing a meaningful hierarchy becomes considerably worse when one considers variability across words sharing the same rhyme. On the one hand, it might be proposed that **Figure 1** illustrates a hierarchy of difficult at the level of the word. Not only was the *lid* vowel less intelligible overall than the *kid* vowel, but no speaker performed better on *lid* than on *kid*, and 13 speakers performed above the 80% criterion on *kid*. That outcome is consistent with a learning pattern in which the vowels in both words can eventually be acquired, but *kid* is mastered first. Because no longitudinal data were collected for this study, that account can be neither confirmed nor rejected. However, even if it proved true, other data do not appear explicable in terms of word-level hierarchies. **Figure 2** is an especially compelling example of the Wade et al.'s (2020) observation that focussing on group-level performance can obscure important individual variability in speech data. Despite no statistical difference between the vowels in *sit* and *hit*, some speakers clearly performed much better on *sit*, while others did better on *hit*. The same type of concern can be raised about *sick* and *kick* in **Figure 3**.

Linguistic and Sociolinguistic Predictors of a Hierarchy of Difficulty

The problems cited above are not the result of a lack of systematicity in the data. The mixed effects analyses in fact confirmed that broad generalizations could be made about the speakers' difficulties in terms of linguistic predictors: vowel (e.g., /ɪ/ was the most difficult of the four), rhyme (/ʊk/ was very difficult) and word (*lid* was especially difficult). Nonetheless, an analysis at a purely linguistic level could not possibly account for the quite idiosyncratic performance of the speakers in the study. Two sociolinguistic variables—length of Canadian residence and use of English—were included in the modeling, but neither contributed meaningfully to the outcome. One might propose including additional psycho-social variables (such as aptitude or motivation) to explain some of the individual differences. While doing so might improve the predictive power of the model, many aspects of the data could not possibly be accounted for in such a way.

In the results seen here, the most serious roadblock to understanding the production difficulties lies in the nearly opposite patterns of difficulty seen in the performance of individual speakers. For instance, even if it were found that differences in proficiency, aptitude or English use could explain why some speakers had difficulty with the vowel in *heat*, while others did not, one is still left to explain why some speakers found the *sit* vowel more difficult than the *hit* vowel and others showed the reverse ordering. It seems far-fetched to suppose that *any* quantitative dimension, such as amount of language use, could have differential effects on different speakers, making one word easier for some and a different word easier for others. Instead, the reasons for the differential performance appear to be idiosyncratic. Since the reasons might be tied to quality, rather than quantity, of language experience, understanding them might require examination of individual learners' experience with English in great detail to pinpoint the time at which particular lexical items were acquired, the interlocutors who modeled them and the speakers' frequency of exposure and use over months or years. Such a pursuit it is out of the question in a retrospective study.

Implications for Teaching

The central question posed in this study—whether or not a pedagogically useful hierarchy could be established for vowel intelligibility—must be answered mainly in the negative. In order to be useful to a classroom teacher, a difficulty hierarchy would need to pinpoint common difficulties that all or most learners in a given group of students would experience. Hypothetically, the most useful piece of information on relative difficulty would be that all, or virtually all, Cantonese learners of English have trouble with structure *x*, where *x* refers to a particular vowel, rhyme, or word. Such a finding might immediately suggest a focus of attention for the classroom, assuming that the structure is important for intelligibility. Next on the list of desiderata might be a finding that Cantonese speakers fall into perhaps two or three groups on the basis of whether they have trouble with structure *x*, structure *y* or structure *z*, perhaps determined by their overall proficiency level. The teacher's strategy might then be to determine which category each student in the class belongs to, and to provide exercises tailored to the difficulties of each group of learners. It is obvious, however, that as the patterns of difficulty become more idiosyncratic (and therefore more complex), the less useful any hierarchy becomes. In this study, the degree of idiosyncrasy is high.

In terms of group patterns, a few generalizations can be offered from the present data, though their value is limited. For instance, /ɪ/ was the most difficult vowel. However, that difficulty has already been observed in other L1 groups, such as Mandarin and Slavic speakers (Munro and Derwing, 2008), and may even be a widespread difficulty for ESL learners in general. It is unlikely that teachers would be unaware of the common difficulties with English /ɪ/ after a modest amount of experience with learners. And despite the group-level performance on /ɪ/, its difficulty was inconsistent across rhymes and words, and highly idiosyncratic for individual learners. A useful message for teachers, then, is that a blanket strategy of "teaching /ɪ/" in all contexts to all learners would be inefficient.

The second most difficult vowel was /ʊ/, but here again idiosyncratic performance requires that the generalization be qualified. Looking more closely at a specific rhyme, the nearly universal difficulty with /ʊk/ seems to meet the criterion for a useful finding. All but one of the speakers in the study performed quite poorly on /ʊk/ items, producing this rhyme most often with /o/ and occasionally with /u/. However, an additional concern has to do with the functional load of the English /ʊ/-/oʊ/ and /u/-/oʊ/distinctions, particularly in the /k/ context. In fact, English has only a handful of minimal pairs involving these vowels and thy generally do not entail confusable words (e.g., *cook-Coke*, *took-toke*, *tuque-took*, *Luke-look*, *kook-cook*). The low frequency of many of these items further reduces the functional load of the contrasts. In short, mispronouncing /ʊk/ with one of the adjacent vowels would entail minimal costs in terms of intelligibility and /ʊk/ would rank low on a list of difficulties requiring attention in an English pronunciation class.

A useful recommendation that can be made on the basis of these data is for instructors to lower their expectations of L1-based error hierarchies and instead focus on identifying and addressing individual learner needs. In the context of an entire class of Cantonese ESL learners in Canada, it would probably be counterproductive to target the other rhymes or words found to be statistically difficult in this study. In each case, a sizeable proportion of the class could be expected to perform well, and time would be wasted on unneeded instruction. At the same time, items generally found to be easy, would not be easy for *all* learners, so failure to attend to "minority difficulties" would disadvantage some class members. Therefore, an individualized program tailored to individual needs is preferable to one based on group-based data.

The implementation of individualized instruction requires a suitable assessment tool to pinpoint problem sounds, rhymes and words with high functional loads and then provide learners with appropriate help. While more research is needed to determine how best to carry out pronunciation needs assessments, current trends in technology are helping to bring a needs-based approach within the reach of instructors. Evidence shows that the benefits of training in L2 speech perception can transfer to production (Sakai and Moorman, 2018). Furthermore, identification of individual difficulties in perception can be achieved to some extent with currently available software, and computer-based perceptual instruction can be provided through high variability phonetic training (HVPT). The effectiveness of HVPT has been demonstrated in several studies (e.g., Thomson, 2012, 2018; Iino and Thomson, 2018). (For a discussion of its implementation for teaching purposes see Barriuso and Hayes-Harb, 2018). Also on the horizon are new applications of automatic speech recognition for pronunciation instruction (García et al., 2020) that offer a great deal of promise for individualized instruction that includes direct feedback on production.

This study was not designed to address the question of why performance on vowels is so idiosyncratic. In fact, this finding should interest theorists because of the challenge it poses to modeling that treats individual variability as "noise." In this case, some findings do not seem explicable on the basis of language experience, aptitude, age of learning or any other quantitative variable because such dimensions are usually assumed to have uniform effects for all learners. Rather, idiosyncratic difficulties may arise because of very specific details of a language learner's experience. It is possible, for instance, that a fossilized mispronunciation of particular word might result from encountering and using it in the very early stages of SLA when L2 perceptual and production processes are undeveloped. Yet the same mispronunciation might not occur in a rhyming word acquired later on. Perhaps further research will shed light on this issue. In the meantime, an advisable practice is to ensure that learners receive pronunciation instruction right from the beginning of the acquisition process.

In considering the above recommendations, it must be noted that the current work has a number of shortcomings, chief among them the limited number of speakers surveyed and the narrowness of the focus (i.e., four vowels). Obviously it cannot be assumed that the data are a complete representation of how Cantonese speakers in general perform on English vowels. Although the elicitation procedure was intended to encourage speakers to produce representative exemplars of their typical pronunciation, it must be noted that a word production task cannot be assumed to capture the nuances that would appear in the full range of contexts in which speakers might communicate. In future work, investigators may wish to expand their focus to encompass productions of a wider range of speech sounds in a variety of speaking situations.

DATA AVAILABILITY STATEMENT

The data sets presented in this study are available via email from the author: mjmunro@sfu.ca.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by SFU Research Ethics Board, Simon Fraser University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This research was funded by a grant from the Social Sciences and Humanities Research Council of Canada.

ACKNOWLEDGMENTS

Preliminary findings from this study were presented at the Annual Meeting of the Canadian Acoustical Association in Vancouver in 2008. I thank Herman Li, Susan Morton, and Natasha Penner for their important contributions to data collection, and Ian Bercovitz and Barinder Thind for their invaluable work on the data analysis. Also appreciated are many discussions with my colleagues Tracey Derwing and Ron Thomson.

REFERENCES

Archibald, J. (2017). "Transfer, contrastive analysis and interlanguage phonology," in *The Routledge Handbook of Contemporary English Pronunciation*, eds O. Kang, R. I. Thomson, and J. M. Murphy (London; New York, NY: Routledge), 9–24. doi: 10.4324/9781315145006-2

Barriuso, T. A., and Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *CATESOL J.* 30, 177–194.

Bohn, O.-S., and Garibaldi, C. (2017). "Production and perception of Danish front rounded/y: a comparison of ultimate attainment in native Spanish and native English speakers," in *Romance-Germanic Bilingual Phonology*, eds M. Yavaş, M. Kehoe, and W. Cardoso (Sheffield: Equinox), 121–136.

Brière, E. J. (1966). An investigation of phonological interference. *Language* 42, 768–796. doi: 10.2307/411832

Brinton, D. M. (2018). Reconciling theory and practice. *CATESOL J.* 30, 283–300.

Burgos, P., Cucchiarini, C., van Hout, R., and Strik, H. (2014). Phonology acquisition in Spanish learners of Dutch: error patterns in pronunciation. *Lang. Sci.* 41, 129–142. doi: 10.1016/j.langsci.2013.08.015

Carlisle, R. S. (1998). The acquisition of onsets in a markedness relationship: a longitudinal study. *Stud. Second Lang. Acquis.* 20, 245–260. doi: 10.1017/S027226319800206X

Catford, J. C. (1987). "Phonetics and the teaching of pronunciation: a systemic description of English phonology," in *Current Perspectives on Pronunciation: Practices Anchored in Theory*, TESOL, ed J. Morley (Washington, DC), 87–100.

Chan, A. Y. W., and Li,. D. C. S. (2000). English and Cantonese phonology in contrast: explaining Cantonese ESL learners' English pronunciation problems. *Lang. Cult. Curri.* 13, 67–85. doi: 10.1080/07908310008666590

Darcy, I., and Krüger, F. (2012). Vowel perception and production in Turkish children acquiring L2 German. *J. Phonet.* 40, 568–581. doi: 10.1016/j.wocn.2012.05.001

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. Indivi. Diff.* 40, 63–72. doi: 10.1016/j.lindif.2015.04.005

Derwing, T. M., and Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research*. Amsterdam: Benjamins. doi: 10.1075/lllt.42

Derwing, T. M., Munro, M. J., and Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Lang. Learn.* 48, 393–410. doi: 10.1111/0023-8333.00047

Derwing, T. M., and Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Appl. Lang. Learn.* 13, 1–17.

Eckman, F. R. (2004). From phonemic differences to constraint rankings: research on second language phonology. *Stud. Second Lang. Acquis.* 26, 513–549. doi: 10.1017/S027226310404001X

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

Edwards, J. G. H. (2017). "Pronunciation and individual differences," in *The Routledge Handbook of Contemporary English Pronunciation*, eds O. Kang, R. I. Thomson, and J. M. Murphy (London; New York, NY: Routledge), 385–398. doi: 10.4324/9781315145006-24

Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. *Speech Percept. Linguis. Exp. Issues Cross Lang. Res.* 92, 233–277.

García, C., Nickolai, D., and Jones, L. (2020). Traditional versus ASR-based pronunciation instruction: an empirical study. *Calico J.* 37, 213–232. doi: 10.1558/cj.40379

Gatbonton, E., Trofimovich, P., and Magid, M. (2005). Learners' ethnic group affiliation and L2 pronunciation accuracy: a sociolinguistic investigation. *TESOL Q.* 39, 489–511. doi: 10.2307/3588491

Hung, T. T. N. (2000). Towards a phonology of Hong Kong English. *World English.* 19, 337–356. doi: 10.1111/1467-971X.00183

Iino, A., and Thomson, R. I. (2018). "Effects of web-based HVPT on EFL learners' recognition and production of L2 sounds," in *Future-Proof CALL: Language Learning as Exploration and Encounters-Short Papers From EUROCALL* 2018, eds P. Taalas, J. Jalkanen, L. Bradley, and S. Thouësny (Research-publishing.net), 106–111. doi: 10.14705/rpnet.2018.26.821

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310

Larson-Hall, J. (2015). *A Guide to Doing Statistics in Second Language Research Using SPSS and R.* Long Grove, IL: Routledge. doi: 10.4324/9781315775661

Levis, J. (2020). Revisiting the intelligibility and nativeness principles. *J. Second Lang. Pronunc.* 6, 310–328. doi: 10.1075/jslp.20050.lev

Levis, J., and Cortes, V. (2008). "Minimal pairs in spoken corpora: Implications for pronunciation assessment and teaching," in *Towards Adaptive CALL: Natural Language Processing for Diagnostic Language Assessment*, 197208.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *Tesol Q.* 39, 369–377. doi: 10.2307/3588485

Liu, D. (2021). Prosody transfer failure despite cross-language similarities: Evidence in favor of a complex dynamic system approach in pronunciation teaching. *J. Second Lang. Pronunc.* 7, 38–61. doi: 10.1075/jslp.18047.liu

Meng, H., Lo, Y. Y., Wang, L., and Lau, W. Y. (2007). "Deriving salient learners' mispronunciations from cross-language phonological comparisons," in *2007 IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU)* (New York, NY: IEEE), 437–442. doi: 10.1109/ASRU.2007.4430152

Moulton, W. G. (1962). Toward a classification of pronunciation errors. *Mod. Lang. J.* 46, 101–109. doi: 10.1111/j.1540-4781.1962.tb01773.x

Munro, M. J. (2018). How well can we predict second language learners' pronunciation difficulties? *CATESOL J.* 30, 267–281.

Munro, M. J., and Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* 45, 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., and Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: an exploratory study. *System* 34, 520–531. doi: 10.1016/j.system.2006.09.004

Munro, M. J., and Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: a longitudinal study of vowel production. *Lang. Learn.* 58, 479–502. doi: 10.1111/j.1467-9922.2008.00448.x

Munro, M. J., and Derwing, T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Lang. Teach.* 44, 316–327. doi: 10.1017/S0261444811000103

Munro, M. J., and Derwing, T. M. (2020). Foreign accent, comprehensibility and intelligibility, redux. *J. Second Lang. Pronunc.* 6, 283–309. doi: 10.1075/jslp.20038.mun

Munro, M. J., Derwing, T. M., and Thomson, R. I. (2015). Setting segmental priorities for english learners: evidence from a longitudinal study. *Int. Rev. Appl. Linguist. Lang. Teach.* 53, 39–60. doi: 10.1515/iral-2015-0002

Munro, M. J., Flege, J. E., and MacKay, I. R. A. (1996). The effects of age of second language learning on the production of English vowels. *Appl. Psycholinguist.* 17, 313–334. doi: 10.1017/S0142716400007967

Nagle, C. (2018). Motivation, comprehensibility, and accentedness in L2 Spanish: investigating motivation as a time-varying predictor of pronunciation development. *Mod. Lang. J.* 102, 199–217. doi: 10.1111/modl.12461

Nagle, C. L., and Huensch, A. (2020). Expanding the scope of L2 intelligibility research: intelligibility, comprehensibility, and accentedness in L2 Spanish. *J. Second Lang. Pronunc.* 6, 329–351. doi: 10.1075/jslp.20009.nag

Nilsen, D. L. F., and Nilsen, A. P. (2010). *Pronunciation Contrasts in ENGLISH*. Amsterdam: Waveland Press.

Nimz, K., and Khattab, G. (2020). On the role of orthography in L2 vowel production: the case of Polish learners of German. *Second Lang. Res.* 36, 623–652. doi: 10.1177/0267658319828424

O'Brien, M. G., and Smith, C. (2010). Role of first language dialect in the production of second language German vowels. *Int. Rev. Appl. Linguist. Lang. Teach.* 48, 297–330. doi: 10.1515/iral.2010.013

Okuno, T., and Hardison, D., M. (2016). Perception-production link in L2 Japanese vowel duration: training with technology. *Lang. Learn. Technol.* 20, 61–80.

Saito, K., Dewaele, J.-M., and Hanzawa, K. (2017). A longitudinal investigation of the relationship between motivation and late second language speech learning in classroom settings. *Lang. Speech* 60, 614–632. doi: 10.1177/0023830916687793

Sakai, M., and Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Appl. Psycholingu.* 39, 187–224. doi: 10.1017/S0142716417000418

Sewell, A. (2017). Functional load revisited: reinterpreting the findings of 'lingua franca' intelligibility studies. *J. Second Lang. Pronunc.* 3, 57–79. doi: 10.1075/jslp.3.1.03sew

Swan, M., and Smith, B. (2001). *Learner English: A Teacher's Guide to Interference and Other Problems.* Cambridge: CUP. doi: 10.1017/CBO9780511667121

Sweet, H. (1900). *The Practical Study of Languages: A Guide for Teachers and Learners.* Glasgow: H. Holt.

Sypiańska, J. (2016). Multilingual acquisition of vowels in L1 Polish, L2 Danish and L3 English. *Int. J. Multilingual.* 13, 476–495. doi: 10.1080/14790718.2016.1217606

Thomson, R. I. (2011). Computer assisted pronunciation training: targeting second language vowel perception improves pronunciation. *Calico J.* 28, 744–765. doi: 10.11139/cj.28.3.744-765

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: a computer-mediated approach. *Lang. Learn.* 62, 1231–1258. doi: 10.1111/j.1467-9922.2012.00724.x

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT) A proven technique about which every language teacher and learner ought to know. *J. Second Lang. Pronunc.* 4, 208–231. doi: 10.1075/jslp.17038.tho

Thomson, R. I., and Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: a narrative review. *Appl. Linguis.* 36, 326–344. doi: 10.1093/applin/amu076

Wade, L., Lai, W., and Tamminga, M. (2020). The reliability of individual differences in VOT imitation. *Lang. Speech* 23830920947769. doi: 10.1177/0023830920947769

Walz, J. (1980). An empirical study of pronunciation errors in French. *French Rev.* 53, 424–432.

Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Q.* 4, 123–130. doi: 10.2307/3586182

White, L. (2018). "What is easy and what is hard," in *Meaning and Structure in Second Language Acquisition: In honor of Roumyana Slabakova,* eds J. Cho, M. Iverson, T. Judy, T. Leal, and E. Shimanskaya (Benjamins), 263–282. doi: 10.1075/sibil.55.10whi

Wong, J. W. S. (2013). "The effects of perceptual and/or productive training on the perception and production of English vowels /ɪ/ and /iː/ by Cantonese ESL learners," in *Interspeech*, 2113–2117.

Wong, J. W. S. (2015). "Comparing the perceptual training effects on the perception and production of English high-front and and high-back vowel contrasts by Cantonese ESL learners," in *Proceedings of the 18th International Congress of Phonetic Sciences,* ed The Scottish Consortium for ICPhS (University of Glasgow).

Zee, E. (1991). Chinese (Hong Kong Cantonese). *J. Int. Phonet. Assoc.* 21, 46–48. doi: 10.1017/S0025100300006058

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Communication

GENERAL COMMENTARY
published: 12 October 2021
doi: 10.3389/fcomm.2021.748991

Commentary: When the Easy Becomes Difficult: Factors Affecting the Acquisition of the English /iː/-/ɪ/ Contrast and On the Difficulty of Defining "Difficult" in Second-Language Vowel Acquisition

*Ron I. Thomson**

Department of Applied Linguistics, Brock University, St. Catharines, ON, Canada

Keywords: second language vowel learning, cross-linguistic similarity, L2 phonology, L2 pronunciation, L2 perception, L2 production, individual differences, L2 instruction

A Commentary on

When the Easy Becomes Difficult: Factors Affecting the Acquisition of the English /iː/-/ɪ/ Contrast
by Cebrian J., Gorba C., and Gavaldà N. (2021). Front. Commun. 10:660917. doi: 10.3389/fcomm.2021.660917

On the Difficulty of Defining "Difficult" in Second-Language Vowel Acquisition
by Munro M. J. (2021). Front. Commun. 10:639398. doi: 10.3389/fcomm.2021.639398

OPEN ACCESS

Edited by:
William Choi,
The University of Hong Kong, Hong Kong, SAR China

Reviewed by:
John Archibald,
University of Victoria, Canada
Kakeru Yazawa,
University of Tsukuba, Japan

****Correspondence:***
Ron I. Thomson
rthomson@brocku.ca

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *28 July 2021*
Accepted: *27 September 2021*
Published: *12 October 2021*

Citation:
Thomson RI (2021) Commentary: When the Easy Becomes Difficult: Factors Affecting the Acquisition of the English /iː/-/ɪ/ Contrast and On the Difficulty of Defining "Difficult" in Second-Language Vowel Acquisition. Front. Commun. 6:748991. doi: 10.3389/fcomm.2021.748991

INTRODUCTION

Investigating adult second language (L2) speech learning is difficult, and interpreting results is often a challenge. This is in part because a satisfactory method for measuring interactions between first language (L1) and L2 sound categories remains elusive (Flege and Bohn, 2021). Cebrian et al. (2021) and Munro's (2021) contributions to the Frontiers' Research Topic "L2 Phonology Meets L2 Pronunciation" evidence the effect of this persistent methodological concern. Both also reveal that generalizations based on group means are problematic, and that many complexities that emerge in L2 speech research are attributable to individual differences across learners, independent of their L1.

THEORETICAL AND APPLIED ORIENTATIONS

L2 phonologists, laboratory phoneticians and applied linguists too often work within sub-disciplinary silos. Cebrian et al. and Munro's studies model how to incorporate concepts from each sub-discipline, allowing for richer insights.

Cebrian et al.'s study is contextualized within the Speech Learning Model (SLM) (Flege, 1995), explicitly testing some of its claims. The study is also influenced by theoretical phonology, treating L2 speech categories as phonemic (e.g., Archibald, 1998) rather than phonetic (e.g., Kohler, 1981), in contradiction of the SLM. Borrowing from applied linguistics, Cebrian et al. use the notion of functional load (FL) to justify a focus on the English /iː/-/ɪ/ contrast (Catford, 1987; Munro and Derwing, 2006; Sewell, 2021). Broadly speaking, FL refers to the communicative weight that a phonological contrast carries within a language, based upon its frequency of occurrence in minimal pairs.

While Munro does not explicitly follow a theoretical framework, his search for an implicational hierarchy of English vowel learning by Cantonese L1 speakers should interest L2 phonologists (e.g., Major, 1998). Further, Munro's attention to differences in the pronunciation of the same vowels in different phonetic environments (i.e., different rhymes) demonstrates a commitment to the SLM's claim that L2 speech learning occurs at the level of contextually sensitive allophones, rather than phonemic categories (Flege, 1995). Munro's primary concerns are applied. Like Cebrian et al., Munro couches his study in terms of FL, aiming to help learners develop intelligible speech rather than a native accent (Levis, 2005).

METHODOLOGICAL CONCERNS

Cebrian et al. and Munro's studies both recognize that the crosslinguistic similarity of L1 and L2 vowels is a primary determinant of successful L2 vowel acquisition. Yet, their findings do not clearly confirm this influence. While they offer alternative explanations for their mixed results, their operationalizations of crosslinguistic similarity are at least partially to blame. Cebrian et al. use a perceptual mapping task, which requires listeners to identify foreign language vowel tokens as members of their closest L1 target, and to indicate how well each token fits the selected L1 category. Guion et al. (2000) conclude that perceptual mapping may not be sensitive enough to accurately capture crosslinguistic similarity. Another concern is that Cebrian et al. had listeners evaluate the crosslinguistic similarity of English and Spanish vowels in one phonetic context (/bVt/) to predict the learning of the same L2 English vowels in different phonetic contexts (a range of/CVC/s). It is well-established that the acquisition of an L2 sound in one context rarely generalizes to other contexts (Thomson, 2016; Mitterer et al., 2018; Thomson, 2018; Flege and Bohn, 2021). While Munro took care to account for this fact, he relied upon Chan and Li's (2000) secondary description of English and Cantonese vowel similarity, the empirical basis for which is unknown.

Cebrian et al. report differential mismatches between their measures of crosslinguistic similarity and learners' perception versus learners' productions of the same English vowels. While this may well reflect real differences in the rate with which each skill develops, incommensurable techniques for evaluating each skill makes direct comparisons impossible (see Nagle and Baese-Berk, 2021; Thomson, 2021).

While unsatisfactory methods do not prove Cebrian et al. and Munro's conclusions are inaccurate, it is reasonable to conclude that imprecise methodology partially explains their confusing results.

THE IMPORTANCE OF INDIVIDUAL DIFFERENCES

Cebrian et al. and Munro's most important insight is the extent to which there exist between-subject differences among matched-L1 learners of L2 English vowels. Their results point to a need for greater attention to individual differences, rather than assuming that all learners from the same L1 background will develop along the same path. Cebrian et al. found a weak relationship between the perceived crosslinguistic similarity of English-Spanish vowels and learners' ability to discriminate between those English vowels. Munro's study determined that there is no implicational hierarchy by which contextually-sensitive allophones of the same phoneme are learned. Individual learners acquired allophones of the same phoneme in no consistent order. While there is a growing recognition that individual differences play a substantial role in ultimate attainment for L2 pronunciation (Darcy et al., 2015; Suzukida, 2021), factors such as aptitude, motivation, and quality of experience with the target language have long played a subordinate role to L1 effects in L2 speech research.

DISCUSSION

Cebrian et al. and Munro's tentative conclusions concerning the role of crosslinguistic similarity in L2 speech learning reinforces the necessity to improve how we measure L2 speech perception and production across languages. Thomson et al. (2009) effectively demonstrate that a statistical pattern recognition model of crosslinguistic similarity, incorporating multiple sources of phonetic information, leads to more accurate predictions for both L2 perception and production. Unfortunately, its labor-intensive nature seems to present an obstacle to its wider adoption.

One gap in both Cebrian et al. and Munro's interpretation of their results is that neither considers the concept of markedness in determining what categories are most learnable (see Archibald, 2021). In both studies, some sounds with which learners had the most difficulty were, in fact, marked (e.g., lax vowels and vowels in checked syllables). While markedness has long been a prominent topic among L2 phonologists, the concept appears to be overlooked by most phoneticians. In the Revised SLM (SLM-r) Flege and Bohn (2021) hypothesize that more input is needed for learners to establish more complex sound categories, which they operationalize as how rare particular sounds are across languages. This new hypothesis suggests that they may have (re)discovered markedness.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

Funding for this publication was provided by the Brock Library Open Access Publishing Fund.

ACKNOWLEDGMENTS

The author is grateful to Terrance Nearey for intellectual inspiration on this topic, and for feedback from two reviewers, which helped to strengthen the manuscript.

REFERENCES

Archibald, J. (2021). Ease and Difficulty in L2 Phonology: A Mini-Review. *Front. Commun.* 6, 626529. doi:10.3389/fcomm.2021.626529

Archibald, J. (1998). Second Language Phonology, Phonetics, and Typology. *Stud. Second Lang. Acquis* 20 (2), 189–211. doi:10.1017/s0272263198002046

Catford, J. C. (1987). Phonetics and the Teaching of Pronunciation: A Systemic Description of English Phonology, *Current Perspectives on Pronunciation: Practices Anchored in Theory*. Alexandria, VA: TESOL, 87–100.

Cebrian, J., Gorba, C., and Gavaldà, N. (2021) When the Easy Becomes Difficult: Factors Affecting the Acquisition of the English Contrast. *Front. Commun.* 6, 660917. doi:10.3389/fcomm.2021.660917

Chan, A. Y. W., and Li, D. C. S. (2000). English and Cantonese Phonology in Contrast: Explaining Cantonese ESL Learners' English Pronunciation Problems. *Lang. Cult. Curriculum* 13, 1 67–85. doi:10.1080/07908310008666590

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. Individual differences* 40, 63–72. doi:10.1016/j.lindif.2015.04.005

Flege, J. E. (1995). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, 233–277.

Flege, J. E., and Bohn, O.-S. (2021). "The Revised Speech Learning Model (SLM-R)," in *Second Language Speech Learning: Theoretical and Empirical Progress*. Editor W. Ratree (Cambridge, UK: Cambridge University Press), 3–83. doi:10.1017/9781108886901.002

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (2000). An Investigation of Current Models of Second Language Speech Perception: The Case of Japanese Adults' Perception of English Consonants. *The J. Acoust. Soc. America* 107 (5), 2711–2724. doi:10.1121/1.428657

Kohler, K. J. (1981). Contrastive Phonology and the Acquisition of Phonetic Skills. *Phonetica* 38 (4), 213–226. doi:10.1159/000260025

Levis, J. M. (2005). Changing Contexts and Shifting Paradigms in Pronunciation Teaching. *TESOL Q.* 39 (3), 369–377. doi:10.2307/3588485

Major, R. C. (1998). Interlanguage Phonetics and Phonology. *Stud. Second Lang. Acquis* 20 (2), 131–137. doi:10.1017/s0272263198002010

Mitterer, H., Reinisch, E., and McQueen, J. M. (2018). Allophones, Not Phonemes in Spoken-word Recognition. *J. Mem. Lang.* 98, 77–92. doi:10.1016/j.jml.2017.09.005

Munro, M. J. (2021). On the Difficulty of Defining "Difficult" in Second-Language Vowel Acquisition. *Front. Commun.* 6, 639398. doi:10.3389/fcomm.2021.639398

Munro, M. J., and Derwing, T. M. (2006). The Functional Load Principle in ESL Pronunciation Instruction: An Exploratory Study. *System* 34 (4), 520–531. doi:10.1016/j.system.2006.09.004

Nagle, C. L., and Baese-Berk, M. M. (2021). Advancing the State of the Art in L2 Speech Perception-Production Research: Revisiting Theoretical Assumptions and Methodological Practices. *Stud. Second Lang. Acquis*, 1–26. Advanced online access. doi:10.1017/S0272263121000371

Sewell, A. (2021). Functional Load and the Teaching-Learning Relationship in L2 Pronunciation. *Front. Commun.* 6, 627378. doi:10.3389/fcomm.2021.627378

Suzukida, Y. (2021). The Contribution of Individual Differences to L2 Pronunciation Learning: Insights from Research and Pedagogical Implications. *RELC J.* 52 (1), 48–61. doi:10.1177/0033688220987655

Thomson, R. I. (2018). High Variability [Pronunciation] Training (HVPT). *Journal of Second Language Pronunciation* 4 (2), 208–231. doi:10.1075/jslp.17038.tho

Thomson, R. I., Nearey, T. M., and Derwing, T. M. (2009). A Modified Statistical Pattern Recognition Approach to Measuring the Crosslinguistic Similarity of Mandarin and English Vowels. *J. Acoust. Soc. America* 126 (3), 1447–1460. doi:10.1121/1.3177260

Thomson, R. I. (2016). Does Training to Perceive L2 English Vowels in One Phonetic Context Transfer to Other Phonetic Contexts? *Can. Acoust.* 44 (3), 198–199.

Thomson, R. I. (2021). "The Relationship between L2 Speech Perception and Production," in *The Routledge Handbook of Second Language Acquisition and Speaking*. Editors T. M. Derwing, M. J. Munro, and R. I. Thomson (London: Routledge). to appear.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers
in Communication

ORIGINAL RESEARCH
published: 18 August 2021
doi: 10.3389/fcomm.2021.639390

The Southwestern Mandarin /n/-/l/ Merger: Effects on Production in Standard Mandarin and English

Wei Zhang[1] *and John M. Levis*[2]*

[1]*School of Translation, Qufu Normal University, Qufu, China,* [2]*Department of English, Iowa State University, Ames, IA, United States*

Southwestern Mandarin is one of the most important modern Chinese dialects, with over 270 million speakers. One of its most noticeable phonological features is an inconsistent distinction between the pronunciation of (n) and (l), a feature shared with Cantonese. However, while /n/-/l/ in Cantonese has been studied extensively, especially in its effect upon English pronunciation, the /l/-/n/ distinction has not been widely studied for Southwestern Mandarin speakers. Many speakers of Southwestern Mandarin learn Standard Mandarin as a second language when they begin formal schooling, and English as a third language later. Their lack of /l/-/n/ distinction is largely a marker of regional accent. In English, however, the lack of a distinction risks loss of intelligibility because of the high functional load of /l/-/n/. This study is a phonetic investigation of initial and medial (n) and (l) production in English and Standard Mandarin by speakers of Southwestern Mandarin. Our goal is to identify how Southwestern Mandarin speakers produce (n) and (l) in their additional languages, thus providing evidence for variations within Southwestern Mandarin and identifying likely difficulties for L2 learning. Twenty-five Southwestern Mandarin speakers recorded English words with word initial (n) and (l), medial <ll> or <nn> spellings (e.g., swallow, winner), and word-medial (nl) combinations (e.g., only) and (ln) combinations (e.g., walnut). They also read Standard Mandarin monosyllabic words with initial (l) and (n), and Standard Mandarin disyllabic words with (l) or (n). Of the 25 subjects, 18 showed difficulties producing (n) and (l) consistently where required, while seven (all part of the same regional variety) showed no such difficulty. The results indicate that SWM speakers had more difficulty with initial nasal sounds in Standard Mandarin, which was similar to their performance in producing Standard Mandarin monosyllabic words. For English, production of (l) was significantly less accurate than (n), and (l) production in English was significantly worse than in Standard Mandarin. When both sounds occurred next to each other, there was a tendency toward producing only one sound, suggesting that the speakers assimilated production toward one phonological target. The results suggest that L1 influence may differ for the L2 and L3.

Keywords: /l/-/n/ merger, Southwestern Mandarin, English pronunciation, production, Standard Mandarin, L3 production, lateral nasal

OPEN ACCESS

Edited by:
John Archibald,
University of Victoria, Canada

Reviewed by:
John H. G. Scott,
University of Calgary, Canada
Andrew Sewell,
Lingnan University, Hong Kong, SAR China

****Correspondence:***
John M. Levis
jlevis@iastate.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *08 December 2020*
Accepted: *09 August 2021*
Published: *18 August 2021*

Citation:
Zhang W and Levis JM (2021) The Southwestern Mandarin /n/-/l/ Merger: Effects on Production in Standard Mandarin and English. Front. Commun. 6:639390. doi: 10.3389/fcomm.2021.639390

INTRODUCTION

When the novel coronavirus that produces COVID-19 was first identified in the city of Wuhan, the Chinese government closed off the city of 11 million and a wider area with a population of 60 million. New hospitals were built, and doctors and other medical workers came from all over China to treat victims of the disease. One of the outcomes of this unprecedented lockdown was dialect/language contact between the Southwestern Mandarin (SWM) speakers of Wuhan and its related dialect areas and speakers of other dialects of Mandarin, resulting in the development of apps that translate the pronunciation of SWM speakers into Standard Mandarin for outside medical workers (Li et al., 2020). Without these technological solutions, there would have been widespread difficulty in communication, necessitating a larger number of interpreters who could bridge the gap between SWM patients and healthcare workers from other areas of China. Among the most challenging pronunciation differences for outsiders was the non-distinction between /l/ and /n/, a common marker of SWM.

The lack of an /l/-/n/ contrast in SWM, which is also found among Cantonese speakers (e.g., Hung, 2000; Ng, 2017) and in some other Mandarin-speaking areas of China such as Gansu province (e.g., Zhang, 2012; Han, 2014) is less well-known than some other highly studied pronunciation problems, especially the /l/-/ɹ/ contrast. Difficulty with /l/-/ɹ/ (e.g., lice-rice, collect-correct, fell-fair) is characteristic of Japanese English pronunciation, and it was heavily studied because of the ubiquity of Japanese learners of English starting in the mid 20th century. The mispronunciation of these sounds continues to stereotype Japanese English.

Both the /l/-/ɹ/ and /l/-/n/ contrasts involve the lateral /l/, and both involve another consonant similar to it in sonority (Yip, 2011), one a (coronal) approximant and the other a nasal. Phonologically, Japanese has neither /l/ nor /ɹ/, but rather has an apico-alveolar tap which shares some aspects of its articulation with both English /l/ and /ɹ/. Similarly, many SWM speakers appear not to have a contrast between the /l/ and /n/, which can cause challenges in being able to perceive and produce such a distinction in another language. Brown (1998) argues that when L2 speakers try to learn a phonemic contrast for which their L1 has no equivalent feature contrast (e.g., the coronal feature of English /l-r/ is not employed in Japanese), their perception of the novel contrast is unlikely to improve with increased proficiency. In contrast, learning a novel contrast when the L1 employs the same feature but not the same contrast leads to increased accuracy with increasing proficiency (e.g., English /l-ɹ/ for Chinese speakers who are familiar with the coronal feature from Chinese, or evidence that Japanese speakers improve on English /b-v/ because Japanese exploits the stop-continuant contrast in other sound contrasts). In regard to our study, it is possible that the /l/-/n/ merger in SWM speakers' phonological systems may lead them to produce their sounds with overlapping features of both /l/ and /n/. In other words, both alveolar laterals and alveolar nasals are featurally ambivalent and ambiguous. Mielke (2005) provides evidence that both sounds are ambiguous with regard to the feature (continuant) because in some languages laterals and nasals pattern similarly to (–continuant) sounds and in other languages they pattern similarly to (+continuant) sounds. Mielke suggests that the feature (continuant) be underspecified for certain sounds (especially nasals and laterals) to account for the patterns evident in different languages. Similarly, Ladefoged and Maddieson (1996) demonstrate that laterals occur with multiple types of articulation, and that a posited feature (lateral) may or may not have a central closure. For nasals, Ladefoged and Maddieson (1996) document a variety of partially nasalized consonants, including nasalized approximants.

Besides featural differences, an inconsistent distinction between (l) and (n) may be the result of a laterally articulated nasal (Soejima et al., 1990; Yip, 2011), which would have aspects of both (l) and (n). Without an acoustic analysis of SWM speakers' productions, which is beyond the scope of this paper, it is not clear whether SWM (l) and (n) articulations can be described similarly to the sounds produced in English and Standard Mandarin. Unlike the more extensive research on (l) and (n) in Cantonese learners of English, there is little research on SWM speakers' pronunciation of (l) and (n) in either standard Mandarin or in English. The goal of this study is to provide a beginning to such research.

The study uses production data to describe the distribution of SWM speakers' production of /l/ and /n/ in two additional languages, Standard Mandarin and English. Because we have no equivalent perception data, we cannot make claims for why there are variations in the production of (l) and (n). Goto (1971) reminds us that adult learners may be more successful in producing sounds that they do not perceive, and that production data may overrepresent L2 learners' perception. It is also possible that the sound produced by SWM speakers may use a variable articulation that has both nasal and lateral elements (Soejima et al., 1990) that are heard by listeners with an /l/-/n/ contrast (such as the researchers) as fitting into one category better than another. Such variable articulations are common in languages. The /s/ in English, for example, may be articulated with the tongue tip up toward the alveolar ridge or the tongue tip down behind the bottom teeth and the tongue blade raised toward the alveolar ridge. This variation is rarely noticed by English speakers, but the tongue tip up articulation has been argued to be the cause of Japanese speakers' difficulty producing differences between /s/ and /ʃ/ in English words (Raver-Lampman et al., 2015).

SOUTHWESTERN MANDARIN

Southwestern Mandarin, the most widely used dialect of Mandarin (Li, 2009) with over 270,000,000 speakers (Chinese Academy of Social Sciences, 2012), is spoken mainly in nine provinces and regions including all of Sichuan, Chongqing, Guizhou, and Yunnan, and some areas in Hubei, Guangxi, Hunan, Shanxi, and Jiangxi (**Figure 1**).

Southwestern Mandarin was probably originally centered in Sichuan, then spread to Hubei, Guizhou, Yunnan, Guangxi and other provinces with Sichuan (including Chongqing municipality) as the center (Qian, 2010). Southwestern Mandarin has been classified into six dialect varieties, that is Chuanqian, Xishu, Chuanxi, Yunnan, Huguang and Guiliu. Each variety has several sub-varieties, with a total of 22 sub-varieties identified (Li, 2009; Qian, 2010; Chinese Academy of Social

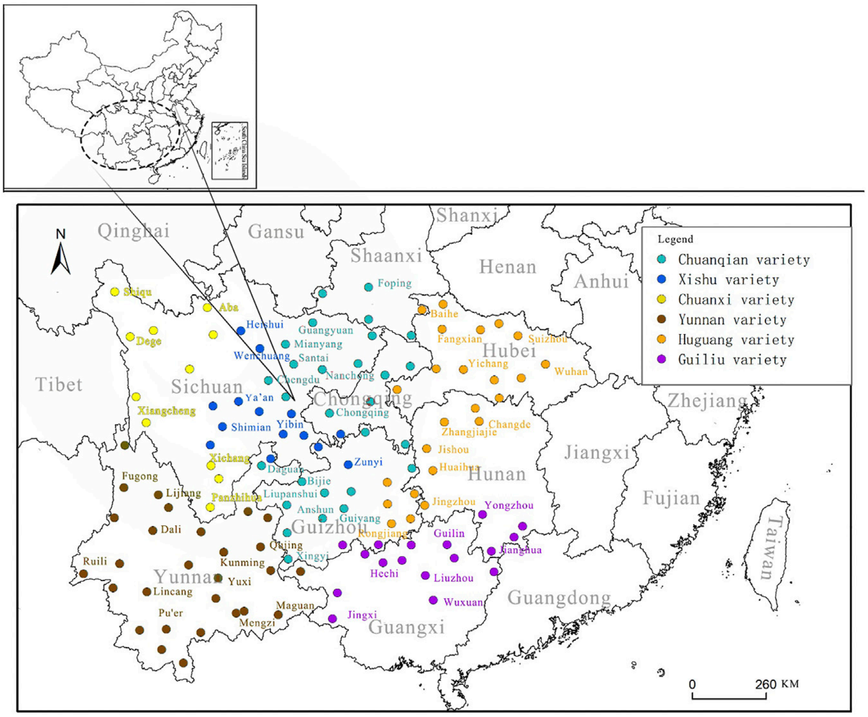


FIGURE 1 | Southwestern Mandarin location and six dialect varieties.

Sciences, 2012). Others have classified it into more dialects (Baker, 1993; Kupaska, 2010). Different dialect varieties have distinctive distributions of phonological features.

For this study, we look at the pronunciation of alveolar nasal and lateral consonants in SWM speakers' L2 and L3 production (Li, 2004; Sun, 2011). We first look at the production of all speakers to determine whether they represent different sub-dialects of SWM in their production of (l) and (n) (Qian, 2010). Like any large dialect area, Southwestern Mandarin has variation in how particular phonological features are realized. For example, the Chuanqian variety is reported to widely confuse /n/ and /l/. Only the alveolar nasal exists in the Chengyu sub-variety of the Chuanqian variety (Zeng, 2009; Zhou, 2014; Li, 2017), but in the Qianzhong sub-variety of Chuanqian, the opposite is true, with only a lateral evident but no nasal (Wang, 1981; Yuan, 1996). In the Xishu variety, the confusion of /n/-/l/ also exists, but the phonology of the sounds in the variety is not established (Luo, 2016). In contrast, the Yunnan variety is considered part of SWM because it shares other phonological features with other sub-varieties of SWM, but speakers of this variety appear to distinguish /l/ and /n/ (Wang, 1986; Li, 2004; Qian, 2010; Chen, 2013).

LITERATURE REVIEW

The /l/-/n/ Contrast in Cantonese and Southwest Mandarin

Nasal and lateral sounds are common across languages of the world, with only a few languages lacking nasals (Mielke, 2005). Ladefoged and Maddieson (1996) also show that both nasals and laterals occur with multiple types of articulation (e.g., approximants, fricatives, clicks) and in different places of articulation (e.g., labial, velar, alveolar, palatal). Nasal consonants are produced by lowering the soft palate so air flows out through the nasal cavity, and the articulators completely stop the oral airflow. A lateral consonant is often but not always produced with a central obstruction, and air passes through one or both sides of the tongue (Ladefoged and Johnson, 2011). The two sounds thus differ mainly in the presence or absence of nasal airflow or in whether the air comes out of the nose or from the sides of the tongue. Acoustically, English (n) and (l) differ in their F2 (Koffi, 2019). In Standard Mandarin, the F2 of the Standard Mandarin (n) is often weak and sometimes even disappears, while the F2 of (l) sometimes shows only a low F2 value (Lin et al., 2013) but is higher when preceding a high vowel and lower when preceding a low vowel (Bao and Lin, 2014).

There are at least two Chinese languages in which /l/ and /n/ do not always contrast. The first is Southwestern Mandarin, and the second is Cantonese (Deterding, 2010). Of these two, Cantonese is far more extensively studied. Historically, Cantonese had both /l/ and /n/ phonemes, but the two sounds have merged over time toward becoming one phoneme with two allophones in complementary distribution (Zee, 1999): (l) occurs in syllable-initial environments, while (n) occurs syllable finally (Ng, 2017), a similar distribution to the light and dark /l/ sounds in English. As a result, Cantonese speakers do not produce (l) in coda position, and the dark (ɫ) found in English is commonly ignored altogether (Wong, 2008). Younger speakers in Hong

Kong have led this merger (Yeung, 1980; Tong and James, 1994; Chan and Li, 2000). Ng (2017) showed an incomplete merger of these two sounds. Initial /n/ was pronounced as (l) 72.3% of the time in Standard Mandarin words whereas initial /l/ was pronounced as (n) only about 4% of the time. In perception, Cantonese listeners misidentified (n) as (l) in 30.3% of items, while (l) was misidentified as (n) in 24.8% of items.

Because of the historical connections between Britain and Hong Kong, much of the research on the /l/-/n/ merger among Hong Kong Cantonese speakers has looked at the effect of the merger on their English pronunciation (e.g., Deterding, 2006; Deterding et al., 2008). Au (2011) looked at how HK Cantonese speakers produced initial /n/ and /l/ in English syllables. The research reported that words with initial (n) were produced invariably with (l). Chan (2014) does not mention /l/-/n/ as a problem for HK Cantonese EFL learners, indicating that initial /l/ is more likely to be confused with /ɹ/ or /w/, but this may be because Chan (2011), in a study of perception of English sounds by Cantonese EFL learners, found that their production was almost completely accurate (97%), and that learners were largely able to distinguish /l/ and /n/ initially (96% accurate). In an earlier contrastive analysis, Chan and Li (2000) characterized initial laterals in Cantonese in this way: initial (l) as in 'love' /lʌv/ may sometimes be pronounced with some "n" quality, giving the impression of a nasalized /l/ sound, viz. (l̃) (p. 80). Another study suggests that their perception may not be so accurate. Li and Hua (2014) studied the effects of visual cues on the ability of both groups to perceive /l/ and /n/ accurately. While they found that audiovisual input (audio supplemented by facial cues of native speakers) significantly helped Cantonese speakers to better distinguish the two sounds, the Mandarin speakers who served as a comparison group found the visual cues unnecessary because they already had a phonemic contrast between /l/ and /n/.

The non-final /l/ and /n/ of SWM speakers seems to follow different patterns from those attested for Cantonese, but there are many questions about /l/-/n/ in SWM that remain unanswered. [In coda position, SWM, like other Mandarin varieties, does not license the lateral but allows (n), although the realization of (n) in our data (Zhang and Levis, 2020) varied between a (+consonantal) pronunciation and a (−consonantal) sound, i.e., a nasalized vowel.] Although the historical loss of a distinction between these sounds in SWM is likely related to a historical process of denasalization (Soejima et al., 1990; Hu, 2007), different sub-dialects in SWM have different realizations of the historical change, as Soejima et al. (1990), p. 131 say:

> In some Chinese dialects, mainly in southern China, (including southern Mandarin dialects), the opposition between /n/ and /1/ has been lost which existed in so-called Ancient Chinese . . . However, the reflexes of these phonemes in modern Chinese dialects and the environmental conditions of this merger are not uniform. That is, in some dialects these phonemes coalesced into /n/ (e.g., Changsha in Hunan, Chengdu in Sichuan), and in some other dialects into /1/ (e.g. Nanjing in Jiangsu). In some dialects, that coalescence occurred only in syllables without the medial front glide (j) (e.g. Nanchang in Jiangxi . . .), while in some other dialects this coalescence occurred spontaneously.

In another possible example of a production that combines both lateral and nasal features Chan (1987) reports that younger Southern Min speakers produce (n) as (n̥l). Some sub-dialects of SWM are reported to have /l/ but not /n/ (Zhang, 2007 for Sichuan English), others /n/ but not /l/, and yet others have both phonemes yet fail to always contrast the sounds (Ao and Low, 2012). It is also clear that Cantonese and SWM pattern differently. In SWM, initial /n/ seems to be more stable than initial /l/ (Koffi, 2019), and there is little evidence of a complementary distribution in the two sounds, but neither do the sounds seem to be in free variation (Zhang, 2007). Ultimately, there is too little information about the distribution of the two sounds in the L1, and not enough is known about how the two sounds occur in L2 and L3 perception and production. Zhang (2007), for example, presented only a single case study of a speaker from Sichuan. Pennington and Saunders (2013) developed a pilot corpus of two speakers from Guiliu, a subdialect of SWM, in which they assert that "/l/ merges with /n/", yet without evidence for the assumption that /n/ is the phoneme that is more stable. Koffi (2019) provided an acoustic analysis of only five mispronunciations in three words from an online database. Ao and Low (2012), in a study of the English spoken in Yunnan province (part of the larger SW Mandarin dialect region) did not identify /l/ and /n/ as a likely problem for this sub-dialect, perhaps because Yunnan is sometimes reported to speak Northwestern Mandarin (Chan, 1987).

The /l/-/ɹ/ and /l/-/n/ Contrasts

We can hypothesize patterns of production of (l) and (n) from other research, especially the /l/-/ɹ/ contrast. There is strong research evidence from Japanese learners of English that learning to perceive and produce the /l/-/ɹ/ contrast is extremely challenging. A number of studies have documented significant improvements from training in both perception and production (Hazan et al., 2005; Aoyama et al., 2008), and there is also evidence that production accuracy can be better than perception accuracy (Goto, 1971), and that perception is a more reliable measure of whether an L2 speaker's production difficulties are based on fundamental difficulties in perceiving differences between two L2 sounds. Brown (1998) and Brown (2000) demonstrates that difficulties with nonnative phonemic contrasts are not all equivalent, proposing that when features in a learner's L1 sound system match features in the L2 (e.g., continuant), L2 phonological perception is likely to be more successful, but that when a feature important in distinguishing phonemes in the L2 system is not present in the L1 system [e.g., the non-use of the (coronal) feature in Japanese and Korean], L2 perception will be impaired regardless of increased L2 proficiency. Brown (2000) provides an example of the English /s/-/θ/ contrast, which because of the feature (distributed), was equally difficult to perceive for Japanese, Korean and Chinese learners of English whose L1s do not require a distinction between (distributed) and (anterior). Applied to the /l/-/n/

contrast, SWM may not have a featural distinction between (n) and (l) even though childhood classroom learning of Standard Mandarin could call attention to the contrast, attention that could be later reinforced by learning of English. Yip (2011), p. 12 argues that although there is conflicting evidence for a feature called (lateral), there is compelling indirect evidence for its reality:

> In Eastern Catalan (and Sanskrit), for example, (lateral) spreads onto nasals to create a lateral nasal: /nl/ → (l͠l) in /son les tres/ → (sol͠les tres) . . . There are well-known phonological processes that involve only (l) and (r), and in which they either dissimilate, as in Latin, where the suffix /-alis/ surfaces as (-aris) after a lateral root: nav-alis *vs.* milit-aris . . . or assimilate, as in Sundanese, where the infix /-ar-/ surfaces as (-al) after a preceding /l/: (k-ar-usut) *vs.* (l-al- əga).

In the absence of perception data, previous research on other contrasts offers hints about possible challenges faced by SWM speakers in producing a distinction between (l) and (n). Goto (1971) shows that Japanese speakers who cannot reliably distinguish /l/-/ɹ/ nevertheless may be able, because of varied spoken strategies, to produce the sounds correctly at a much higher rate. Brown (1998) explains that L2 learners often can “accurately produce a nonnative contrast even though the same learners are unable to distinguish the two sounds perceptually” (p. 156). This happens because they already have fully developed motor control, allowing them to often carry out necessary articulatory movements to be heard as they intend. For example, SWM speakers may be able to produce differences between (l) and (n) when they read words because they know that spelling differences in the written input reflect different sounds.

If the challenges in pronouncing this contrast come from SWM lacking a featural distinction between the two sounds, we would expect that production errors would be common in all environments, with some environments being more challenging than others. We would also expect that speakers from different varieties of SWM would have different frequencies of errors depending on whether the variety has been described as having an /n/, an /l/, or both phonemes (Qian, 2010). Ultimately, however, a production study cannot provide evidence for whether SWM speakers perceive the differences between the sounds when pronouncing an L2.

/l/-/n/, Functional Load and Speech Intelligibility

Many Chinese students study abroad in English-speaking countries, with around 370,000 in the United States alone (IIE, 2019). Of these, a substantial number likely come from the Southwestern Mandarin dialect area, which is also a destination for foreign students (Leung and Sharma, 2020). A non-distinction between /l/ and /n/ in English, although an identifying characteristic of their Chinese variety, may create difficulties in outsiders understanding their Chinese speech (as in the story at the beginning of this paper), but it is likely to be highly damaging to the intelligibility of their English speech. L2 pronunciation teachers who have not encountered SWM students may find themselves puzzled by this little discussed error (Richards, 2012). This can be illustrated by an experience of the second author. Once, a friend from Sichuan was going to go the supermarket to buy “wallets.” When he was successfully helped by his housemate (after initial confusion and many corrections) to say “walnuts” instead, he immediately reverted to “wallets” thinking he was finally saying it correctly, but the production would likely create confusion in a supermarket.

In English, lack of a distinction between /l/ and /n/ is rated at the highest level of the functional load (FL) scale for onset consonants (Catford, 1987; Brown, 1988), a measure of how likely two contrasting sounds are to confuse listeners. (There is no equivalent measure for contrasts in Standard Mandarin, but in the view of the first author, /l/ and /n/ do not often contrast in Standard Mandarin.) High FL sound contrasts have many minimal pairs (as in low-know, light-night) and a correspondingly greater chance that mispronunciations will be heard as a different word. In comparison, low FL sound contrasts (such as thought-fought) have few minimal pairs, a lower likelihood of confusion for listeners, and a greater likelihood that listeners will be able to successfully understand an utterance even when there is a pronunciation error. Functional load for onsets is particularly important because such consonants are most likely to lead listeners to expect a particular cohort of words (Bent et al., 2007).

Functional load is a way to measure error gravity in regard to segmental mispronunciations. Errors at the segmental level are frequent because they are unavoidable in speech; they are also conspicuous signals of nonnativeness or dialect, and they correlate with how well speakers are understood (Zielinski, 2006; Munro et al., 2015). In one investigation of the effect of segmental accuracy on intelligibility judgments, Bent et al. (2007) found a strong correlation between judgments of intelligibility and segmental pronunciation accuracy of vowels and word-initial consonants. Zielinski (2008) similarly showed that errors in stressed vowels and consonants that occurred in stressed syllables were likely to damage intelligibility. These findings have one thing in common: each segment is not equally important for intelligibility, and thus teachers should prioritize teaching certain segments. This view finds its most complete explanation in the concept of FL (Catford, 1987; Brown, 1988; Sewell, 2017), which offers a reason why some segments are more important for intelligibility.

Although discussions of FL stretch back over 100 years, FL became widely noticed in the 1980s in regard to L2 pronunciation teaching (Catford, 1987; Brown, 1988; Catford, 1988). The FL model presents the relative importance of segments in terms of how much work two phonemes do in communicating meaning differences in a language (Sewell, 2017). When words differing in one sound (i.e., minimal pairs) are more frequent, the corresponding sound contrast has a higher functional load in the language. More complex measures of FL can be made by taking various criteria into account, such as the number of minimal pairs that a particular phonemic distinction differentiates at the beginning of a word and end of a word (Catford, 1987) and the frequency of occurrence of each word in a minimal pair; pairs of words that are both frequently encountered have higher FL than those that are infrequent. Part of speech in a minimal pair is another contributing factor. There is a higher value when two words share the same part of speech, as listeners are thought not to confound words as easily if the words are, for example, a noun and a pronoun (Brown, 1988). Brown (1988) also presents a hierarchy of the English phonemic contrasts ranging from 1 (low) to 10 (high). There is empirical evidence to support the use of FL in predicting judgments of comprehensibility and accentedness. Munro and Derwing (2006), using read-aloud sentences, explored the difference between high FL and low FL consonantal errors (including /l/-/n/) on ratings of comprehensibility and accentedness. They found that high FL errors had a significantly larger effect on judgments of comprehensibility and accentedness, and they also found an effect of frequency for ratings of accentedness, that is, two high FL errors caused greater severity in judgments of accentedness, but this was not true for multiple low FL errors.

Kang and Moran (2014) correlated high and low FL errors with test scores from the Cambridge ESOL General English examinations. Low proficiency scores correlated with more FL errors, while higher scores did not have many high FL errors. Low FL errors, in contrast, though more frequent than high FL errors, were similar in frequency for most levels. Suzukida and Saito (2019), in another study looking at the correlation of high FL errors with scores on two tasks with Japanese English learners in Canada, found that segmental errors with high FL values were more detrimental to judgments of comprehensibility than were low FL errors.

Because the /l/-/n/ contrast is likely to affect listener judgments of how comprehensible SWM speakers are and because it has been insufficiently studied in previous research, we want to describe the patterns of pronunciation for initial and medial /l/ and /n/ as well as to look at how speakers pronounce the two sounds when they occur together in the same word (e.g., walnut, only).

L2 and L3 Phonological Acquisition

This study looks at the effects of L1 on /l/-/n/ production in two additional languages, Standard Mandarin and English. For SWM speakers, Standard Mandarin is learned earlier (from the beginning of formal schooling) than English, and Standard Mandarin is typologically similar while English is not. There is conflicting evidence about the effects of the L1 or L2 on L3 pronunciation. Hammarberg (2001) says that "there appears to be a general tendency to activate an earlier secondary language in L3 performance rather than L1" (p. 23). The L3 can also influence the L2, as Cabrelli Amaro (2017) points out that "an ostensibly native-like L2 is more vulnerable to L3 influence than an L1" (p. 699). Llama et al. (2010) point out that the two primary factors involved in cross-linguistic influence for L3 acquisition are language distance (i.e., whether they two languages are typologically similar or dissimilar) and the language status of the L2—that is, the proficiency strength of the L2. Llama et al. (2010), who studied the voice onset time (VOT) of French-English and English-French bilinguals learning Spanish as an L3, found that the L2 was the stronger influence on VOT production of L3 Spanish. In another study, however, Wrembel (2013) found that Polish L1 speakers (with French as their L2 and English as their L3) showed the strongest influence of their L1 on L3 accentedness. Their L2 French, in contrast, showed much less Polish influence. She suggests that L2 influence is likely to be strongest in the earliest stages of proficiency. She also points out that even though their French L2 was more advanced, the participants had learned English as a second L2 but were now coming back to it after some years. As a result, the stronger influence of language distance may have occurred because both French and English were simultaneous L2s.

Our study is not able to state unambiguously whether the L1 or L2 is the stronger influence on /l/ and /n/ production because we have no data on the participants' L1 production. Cabrelli Amaro (2012) makes clear that the features being examined must be represented by equivalent data from all three languages. However, Standard Mandarin is both typologically closer and likely to be more advanced in proficiency because of earlier learning. This would indicate that accurate pronunciation of the English /l/-/n/ contrast, if influenced by the L2, would have a similar error rate. However, if the error rate is much larger than that in the L2, it would suggest that L1 influence is stronger.

This Study

Our exploratory study examines how common /l/ and /n/ pronunciation deviations are in the L2 speech of SWM speakers. Although /l/ and /n/ substitutions are noticeable in SWM L2 speech, and those substitutions are likely to affect intelligibility, SWM speech is understudied in comparison to /l/-/n/ in Cantonese L2 speech or /l/-/ɹ/ in Japanese L2 speech. Because of this, our study describes what kinds of mispronunciations are likely in SWM speakers' production in other languages, whether their L2 pronunciations can shed light on their L2 phonology, and what information it provides for L2 pronunciation teaching. We look at five research questions that focus on the effect of the L1 phonological system on errors, error frequency for /l/ and /n/ production in both Standard Mandarin and English, and the influence of linguistic environment on production accuracy.

1) In regard to the production of /l/ and /n/, is SWM a consistent dialect, that is, do all subjects have difficulty producing a distinction between the two sounds?

> Because there is evidence that different dialects of SWM have differing patterns of (l) and (n) production, we first established whether all participants have difficulty in producing the distinctions. Those who had no difficulty in producing the target distinction would be excluded from further analysis so as to provide more accurate results for those who did not distinguish the two sounds.

2) Does (l) and (n) production vary by Tone and Rhyme in Standard Mandarin? Does accuracy change when (l) and (n) are pronounced in Standard Mandarin disyllabic words?

TABLE 1 | Subjects' dialect variety and hometowns.

Dialect variety	City
Subjects whose dialect/city is attested as having /l/ but not /n/ (*n* = 14)	
Xishu	Yibin; Zunyi
Huguang	Jishou; Changde
Chuanqian	Anshun; Bijie; Liupanshui; Guiyang; Xingyi
Subjects whose dialect/city is attested as having /n/ but not /l/ (*n* = 4)	
Chuanqian	Mianyang; Nanchong
Subjects whose dialect/city is attested as having /l/ and /n/ (*n* = 7)	
Yunnan	Qujing; Kunming; Dali; Mengzi; Hechi

The purpose of this question is to explore whether linguistic environment affects accuracy of /l/-/n/ production in Standard Mandarin reading. Environment here involves Tone and Rhyme (especially the vowel following the /l/ or /n/).

3) Is there an effect of linguistic environment on the accuracy of production in English words? To what extent do subjects mispronounce initial (n) or (l) in English both when the sounds occur alone and when there are competitor sounds at the end of the word?

This question addresses whether the presence of a competitor sound in the coda of the word (lemon, label, nine, nail) as opposed to the target sound being alone in the word (light, night). If participants make more errors in production with a competitor sound, this would suggest that the accuracy of initial (l) and (n) are affected by the presence of a similar sound.

4) Do SWM speakers produce /n/ and /l/ significantly differently in Standard Mandarin than they do in English?

This question asks whether the L1 shows different effects on the production of (l) and (n) in the L2 (Standard Mandarin) and L3 (English). It is possible for the L1 to similarly affect both the L2 and L3, or for the L2 to affect the L3 more than the L1. If L3 English has a higher rate of production errors than L2 Standard Mandarin, this would suggest that the L1 has a different impact on the additional languages, and that the L2 does not impact the L3.

5) When /l/ and /n/ occur word medially in English, are there differences in accuracy when the sounds occur alone or in combination with the other sound?

This question looks at potentially dissimilar phenomena. Medial (l) and (n) productions occur singly, in this case with words with orthographic doubling in which the orthography represents a single sound, or together, that is, when both sounds are pronounced (medial <ln> and <nl>). We expect that words with both sounds (walnut, only) will show a higher error rate while words with orthographically doubled letters (swallow, winner) will have error rates similar to initial (l) and (n).

METHOD

Subjects

Twenty-five native speakers of Southwestern Mandarin, all current undergraduate students of Qufu Normal University, China, were recruited in a manner consistent with the institutional research guidelines of the university. All were compensated for taking part in the study. All grew up within the SWM area before attending college outside of the SWM area. A questionnaire was used to obtain their biographical information. All subjects indicated SW Mandarin was their first oral language, and it was also their dominant language for daily communication. Their hometowns and dialect areas are listed in **Table 1**. The subjects were divided into three groups according to their SWM dialect varieties and their first language phonological system: the /l/ group were subjects whose dialect is attested as having /l/ but not /n/ (Wang, 1981; Xiao, 1996; Ming, 1997, 2005; Zheng, 1999; Li, 2002; Wang et al., 2009; Xin, 2013; Luo, 2016); the /n/ group were those whose dialect was attested as having /n/ but not /l/ (Zeng, 2009; Zhou, 2014; Li, 2017) and the /l/-/n/ group were those whose dialect is reported as having /l/ and /n/ (Wang, 1986; Lan, 1995; Su, 2010; Chen, 2013; Wei, 2018). Note that all varieties except the Chunqiang variety are connected to a single pattern of /l/-/n/ production. All subjects learned Standard Mandarin and English when they started their formal education. They started to learn Standard Mandarin when they were in kindergarten at the age of 4 or 5 years old, and to learn English at the third year of primary school (around 9 years old). Their mean age was 19.5 years (range 18–21), their mean length of formal English instruction was 10 years (range 7–12 years), all of which was conducted in China (English is a core subject in the curriculum of elementary, secondary and tertiary education in China). They were not majoring in language and had not had formal phonetic training, although phonetic symbols are part of normal English instruction. All reported normal speaking and hearing abilities.

Stimuli

Subjects read aloud a word list which included English and Standard Mandarin words. The English word list included words with /n/ and /l/ in onset, medial and coda positions (**Table 2**). (Codas were included to examine the effect of competitor sounds within the same word, but they were not analyzed for this study.) For initial /n/, there were words with only initial /n/ (night), words with initial and final /n/ (nation), and words with initial /n/ and final /l/ (nail). Words with initial /l/ followed the same pattern. Example words include light, label, and lemon. The three different types of words for initial /l/ and

TABLE 2 | English word list numbers for initial and middle /l/ and /n/ recording.

Word type (n)	Words/Total	Word type (l)	Words/Total
Initial (n)		**Initial (l)**	
#n (night)	15/372	#l (light)	15/375
#n___n# (nation)	9/225	#l__l# (label)	9/225
#n___l# (nail)	10/246	#l___n# (lemon)	10/249
Medial (n)		**Medial (l)**	
#__nl__# (only)	10/250	#__ln__#(walnut)	7/174
#__<nn>__# (winner)	6/150	#__<ll>__#(swallow)	6/148

/n/ were used to explore whether the presence of a competitor sound at the end of the word would affect the accuracy of initial /l/ and /n/ production. Evidence from L1 acquisition shows that children may assimilate initial consonants to the articulation of following consonants (e.g., dog pronounced as *gog, see Pater and Werle, 2003). This led us to include words with word-final /n/ and /l/ to determine whether these expected sounds influenced accuracy of initial /l/ and /n/.

> The words with /l/ and /n/ in medial environments included words with <nl> or <ln> (only or walnut), words with <ll> (swallow), and words with <nn> (winner). The list included 97 English words two times each, 34 with word-initial /n/ and 34 with word-final /l/, 10 with word-medial /nl/, seven with word-medial /ln/, and six each with word-medial /l/ and word-medial /n/. There were no distractor items. All English words were presented using both normal orthography and phonetic symbols based on British phonetic symbols used in Chinese EFL textbooks. (Chinese English learners start to learn English phonetic symbols when they are in junior high, and it was our expectation that using phonetic symbols could help them to understand the pronunciation of the words.)

Also included in the reading task were 50 Standard Mandarin monosyllabic words, all spoken with each of the four tones. All Standard Mandarin words were presented in Pinyin, which is the official romanization system for Chinese in China. In some cases, there was a gap and no actual Standard Mandarin word existed, resulting in a nonsense word, and subjects were told that some of the words they read would not be real words. They were nonetheless able to complete the task without undue difficulty. The total tokens were therefore 400, included 200 initial nasal words and 200 initial lateral words. Subjects read each word two times; the total analyzed initial /l/ tokens were 5,000 (25 initial l-words × 4 tones × 25 subjects × 2 times), with the same number for initial /n/ tokens.

The syllable structure of SW Mandarin is the same as Standard Mandarin and maximally consists of onset (consonant), (glide) and rhyme (with an optional nasal coda), as specified in Třísková (2011). For most words, the effective syllable structure is CV, far simpler than English. There are also four tones in Standard Mandarin, i.e., Tone 1 (level), Tone 2 (rising), Tone 3 (falling-rising) and Tone 4 (falling). Rhymes include vowel (plus glide) with an optional final nasal. Standard Mandarin has a Sihu rhyming system with four kinds of rhymes, called Kaikou Hu, Hekou Hu, Qichi Hu and Cuokou Hu. Linguistically, Kaikou Hu is the rhyme that begins with non-high vowels, that is, not /i u y/, Hekou Hu begins with the /u/ sound, Qichi Hu begins with /i/, and Cuokou Hu begins with /y/. The Standard Mandarin words we used in our study were formed with an initial nasal or lateral with each of the four kinds of rhyme. According to the Mandarin Phonetic Alignment Chart Huang and Liao (2011), initial nasals and laterals can be followed by all four Sihu rhymes, but not with all vowels in each rhyme. In our words, we used only phonotactically-licensed rhymes, even though these did not always create an actual word in Standard Mandarin (**Table 3**). In the first word type, the vowels in pinyin were <a> (a), <e> (ɤ), <o> (o) or rhyming units that began with these vowels, such as <ai>, <ei>, <ou>, etc. In the second word type, the vowels were the high front vowel (i) or rhyming units beginning with <i>, such as <ie> (iɛ), <ing> (iŋ), etc. The third word type preceded the high back vowel (u) or rhyming units beginning with (u), such as <uo> (uo), <uan> (uan), etc. In the last word type, /n/ or /l/ preceded the high front vowel (y) or rhyming units starting with <ü>, like <üe> (yɛ).

Chinese words can be divided into monosyllabic words, disyllabic words (two monosyllabic words that function as a single word), trisyllabic words and tetrasyllabic words according to the number of syllables, and the syllables that make up a word are generally non-variable (Modern Chinese Teaching and Research Program, Chinese Department, Peking University, 2020). We used CNCORPUS (www.cncorpus.org) to identify whether the Chinese disyllabic stimuli were a (phonological) word or a phrase. Standard Mandarin disyllabic words have the features of word stress (Duanmu, 2000; Feng, 2016; Zhou, 2018; Zhang, 2021). Of the sixty disyllabic words, 56 were considered to be phonological words rather than phrases. The sixty Standard Mandarin disyllabic words with nasals and laterals in combination with other nasals and laterals (combo-type) or alone (non-combo type) were also part of the reading. Non-combo disyllabic words had one word with an initial nasal or lateral, while the other word did not, like nan fang (nan faŋ). In combo-type disyllabic words, both words had an initial nasal or lateral, such as nie lian (niɛ lɛn) (**Table 4**). The total recorded tokens for non-combo Standard Mandarin disyllabic words were 1,000, and for combo 2,000. In combo-type words, the second sound was a competitor (as in the words like lemon and label for English). In non-combo type words, only the word with the targeted sound was analyzed. Standard Mandarin disyllabic words were presented in simplified Chinese script, which is the norm in China. The Standard Mandarin reading items were also presented without distractors.

Procedures

The recordings were collected in a professional recording studio at Qufu Normal University. The productions were recorded to a computer *via* a directional microphone using Audacity software,

TABLE 3 | Standard Mandarin word list with initial /l/ and /n/.

Type	Word
Non-high vowels rhymes	na, la, ne, le, nai, lai, nei, lei, nao, lao, nou, lou, nan, lan, nang, lang, neng, leng, nong, long
/i/ rhyme	ni, li, nia, lia, nie, lie, niao, liao, niu, liu, nian, lian, nin, lin, ning, ling, niang, liang
/u/ rhyme	nu, lu, nuo, luo, nun, lun, nuan, luan
/y/ (ü) rhyme	nü, lü, nüe, lüe

Note: <ü> represents (y).

TABLE 4 | Standard Mandarin disyllabic words list numbers for initial /l/ and /n/ recordings.

Combo phrase type (n)	Words/Total	Combo phrase type (l)	Words/Total
#n+X (nan fang)	5/125	#l+X (lan fang)	5/125
X+#n (huang ni)	5/125	X+ #l (huang li)	5/125
#n+#n (niu nai)	10/250	#l+#l (liu lliang)	10/250
#n+#l (nie lian)	10/250	#l+#n (liu nian)	10/250

Note: "X" refers to an onset consonant that is neither (n) nor (l). These are classified as non-combo type phrases.

with a sampling rate set to 44.1 kHz at 16 bits per sample on one channel. Before recording, all speakers were given sufficient time to practice all words. Speakers were asked to read each word two times. The duration of the reading was about 15–20 min for both English and Mandarin. The mono recordings were saved as individual WAV files for evaluation. Recordings that were problematic (e.g., from subject noise or from sitting too far from the microphone) were excluded from the analysis.

Subjects first read the Standard Mandarin monosyllabic words, each spoken with all of the four tones, followed by the sixty Standard Mandarin disyllabic words. Finally, subjects also read all the English words two times each. Certain tokens were excluded because three subjects missed some words while reading or because subjects sat too far from the microphone or made other noises.

Data Analysis

All English sound files were evaluated by the researchers, and all Standard Mandarin sound files were evaluated by the first author, a native speaker of Standard Mandarin. Both researchers in this study have linguistic training and are trained in listening to second language pronunciation learners. One researcher is a native speaker of American English with 35 years of experience. The other researcher is a native speaker of Standard Mandarin and is an advanced L2 speaker of English who teaches English in China. Errors in the production of initial /l/ and /n/ were identified as a substitution. We evaluated the second production of each pair. This allowed subjects to read each word once before producing words that were evaluated. As a result, if the first reading was correct and the second one was incorrect, we rated it as incorrect. If the first was incorrect and the second was correct, we rated it as correct. If the researchers were not in agreement, we consulted Praat, especially looking at the F2 measures (Koffi, 2019) and discussed our decisions until we agreed.

RESULTS

The first research question asked whether our subjects all had difficulty producing the distinction between (l) and (n). **Figure 2** shows the percentage of errors from the original 25 SWM speakers in this study. Seven, all from the Yunnan variety, were found to have no difficulty producing a difference between the two sounds in Standard Mandarin, in accord with some descriptions of Yunnan speakers of English (Deterding, 2006; Ao and Low, 2012). Their mean error rates for (n) production and (l) production were 0.009 and 0.014%, respectively. As a result, we report results only from the remaining 18 subjects, 14 from the /l/ group and four from the /n/ group. Because of the disparity in numbers of subjects, any comparisons between these two varieties can only be reported descriptively. It appears that the /n/ group had a greater difficulty with the production of (l), as would be expected, and that the /l/ group's errors for (n) and (l) had more errors for (n) readings. Mean error rates for the /l/ group's production of (n) and (l) were 14.61 and 7.25%, respectively, while the /n/ group's mean error rates for production of (n) and (l) were 7.63 and 16.25%, respectively. These proportions are mirror images of each other and indicate an influence of the L1 phonological systems.

Mean Error Rate of Initial /l/ and /n/ in Standard Mandarin

Figure 3 shows the mean error rate of initial /n/ production for the four types of rhymes in Mandarin. SWM speakers had the largest number of errors for initial nasals in Standard Mandarin (25%) when the rhyme started with /y/ and with /u/ (23.05%). The other rhymes, in descending order, were rhymes starting with non-high vowels and with /i/ at 20.98 and 18.05%, respectively. When SWM speakers pronounced initial laterals in Standard Mandarin, the rhymes may have influenced the accuracy of their production, but the variations were not as noticeable. In descending order, the mean error rates for initial /l/ were at 15.74% for Kaikou Hu (non-high vowels rhyme), 18.05% for Hekou Hu (/u/-rhyme), 21.45% for Qichi Hu (/i/-rhyme), and 16.66% for Cukou Hu (/y/-rhyme) (**Figure 3**). These results suggest that SWM speakers may be generally more accurate with initial nasals and laterals when the rhymes start with a nuclear non-high vowel.

We also examined SWM speakers' performance on disyllabic words in Standard Mandarin. The results indicate that SWM speakers had more difficulty in producing initial nasal sounds

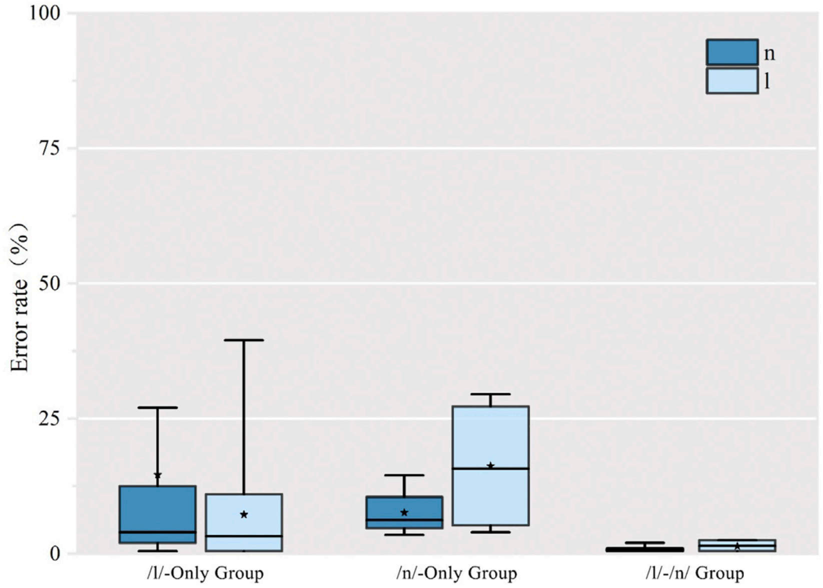


FIGURE 2 | Error rate of production of Standard Mandarin words with initial /n/ and /l/ by subjects' variety.

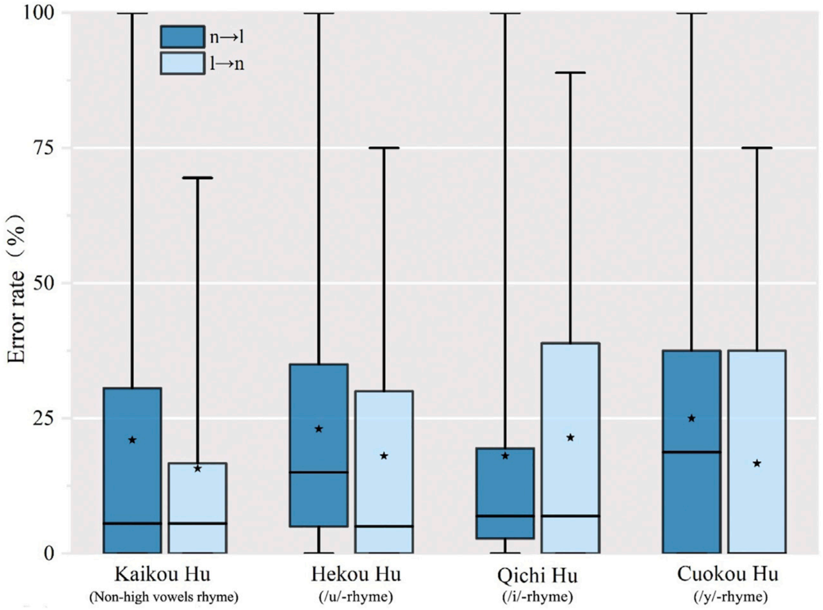


FIGURE 3 | Error rate of production of Standard Mandarin words with initial /n/ and /l/ by word types.

than lateral sounds in Standard Mandarin (see **Figure 4**), which was similar to their performance in producing Standard Mandarin monosyllabic words. There were two types of Standard Mandarin disyllabic words analyzed. **Figure 5** shows SWM speakers had more trouble producing initial nasal and lateral sounds in non-combo Standard Mandarin disyllabic words. The mean error rate for producing initial nasals in non-combo disyllabic words in Standard Mandarin was 22.77%, and the mean error rate for initial laterals in non-combo disyllabic words was 15.55%. However, the mean error rate of production for combo disyllabic words was 18.05% for initial nasals and only 7.5% for initial laterals, which indicates, somewhat surprisingly, that the

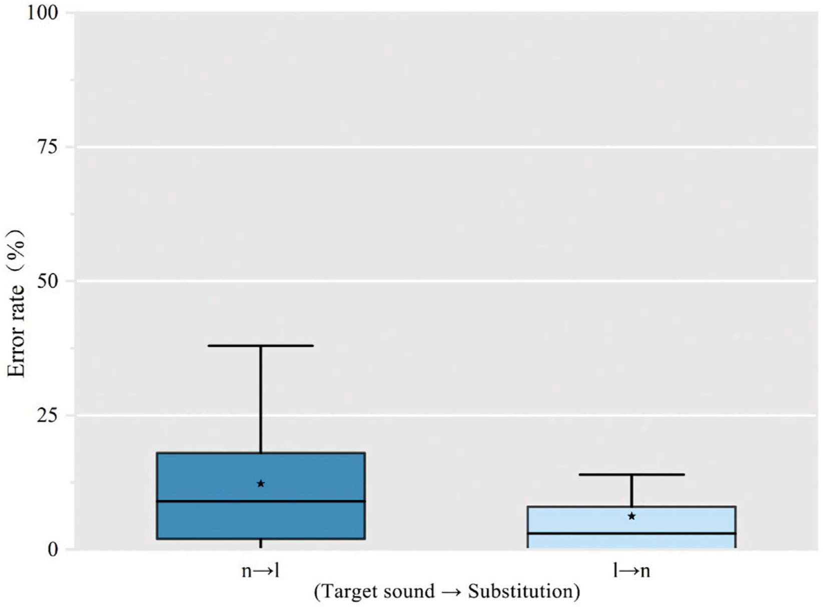


FIGURE 4 | Error rate of production of Standard Mandarin disyllabic words with initial /n/ and /l/.

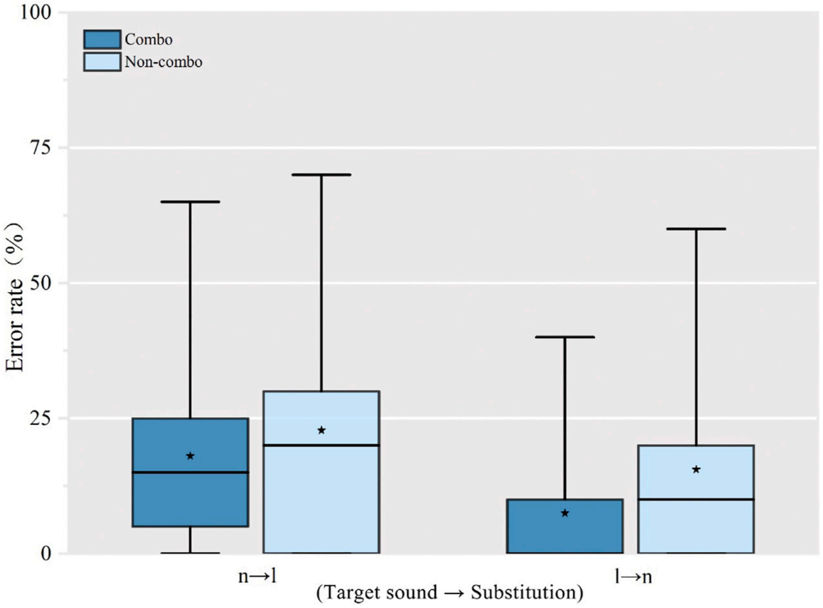


FIGURE 5 | Error rate of production of Standard Mandarin disyllabic words by formation.

presence of a competitor sound facilitated accuracy for both initial nasals and laterals. Rather than making the articulation more challenging, subjects appeared to do better. When the competitor sound was identical, this makes sense as both can more easily assimilate to the same articulation. When the competitor was different (as in lemon, nail), this greater accuracy suggested greater care in distinguishing the sounds.

A mixed-effects logistic regression analysis using lme4 in R was used to examine the effect of tone on production of initial nasals and laterals in Standard Mandarin (R Development Core

TABLE 5 | Descriptive results for tone effects.

		β	*SE*	*z*	*p*
NL1	(Intercept)	2.936	0.800	3.673	0.000
	Tone 2	−1.027	0.881	−1.166	0.243
	Tone 3	−0.766	0.908	−0.844	0.398
	Tone 4	−1.248	0.863	−1.446	0.148
Tone 1	NL2	0.139	0.528	0.263	0.792

Notes: NL1 = lateral, NL2 = nasal.

TABLE 6 | Tone comparison results under the condition of lateral.

Tone pairwise	*β*	*SE*	*z*	*p*
1–2	1.203	1.210	0.994	0.320
1–3	0.732	1.274	0.575	0.565
1–4	1.221	1.208	1.01	0.312
2–3	−0.47	0.998	−0.471	0.637
2–4	0.017	0.911	0.019	0.985
3–4	0.488	0.995	0.490	0.624

Notes: 1 = tone 1,2 = tone 2, 3 = tone 3, 4 = tone 4

TABLE 7 | Tone comparison results under the condition of nasal.

Tone pairwise	*β*	*SE*	*z*	*p*
1–2	0.801	1.288	0.622	0.534
1–3	0.789	1.290	0.612	0.540
1–4	1.267	1.227	1.033	0.301
2–3	−0.012	1.069	−0.011	0.991
2–4	0.466	0.991	0.470	0.638
3–4	0.478	0.994	0.481	0.631

Notes: 1 = tone 1,2 = tone 2, 3 = tone 3, 4 = tone 4

Team, 2009) with four tones, and two consonants (nasal and lateral) as fixed factors, subject and item number as random factors. The results (**Tables 5–7**, **Figure 6**) showed that there was no significant difference between subjects' error rates for lateral and nasal production for the four tones. Tone did not affect mispronunciations of initial /l/ or /n/ for SWM speakers.

Mean Error Rate of Initial and Medial /l/ and /n/ Production in English

Figure 7 shows that SWM speakers had consistent difficulties in the production of syllable initial /l/ with all three types of initial /l/ English words. Descriptively, the mean error rate for each was high, with 30% or more productions being wrong for each environment. Words with initial lateral and final alveolar nasal (lemon) were mispronounced 37.77% of the time, followed by the words with initial and final lateral (label) at 35.18%. The lowest mean error rate was for words with only an initial lateral (light), which was 31.11%. SWM speakers had much less trouble pronouncing nasals. The mean error rate was around 15% with modest variation between environments. The mean for the words only with initial nasal (night), the words with initial nasal and final lateral (nail), and initial and final nasals (nation) were 21.48, 15 and 14.51%, respectively. These descriptive results suggest that environments may affect initial /n/ and /l/ differently. While the absence of a competitor sound meant worse accuracy for /l/, it meant better accuracy for initial /n/.

English is the third language for all subjects; we also wanted to know how different groups of SWM speakers pronounced English words with initial /l/ or /n/. As with Standard Mandarin, the effect of the participants' sub-variety is

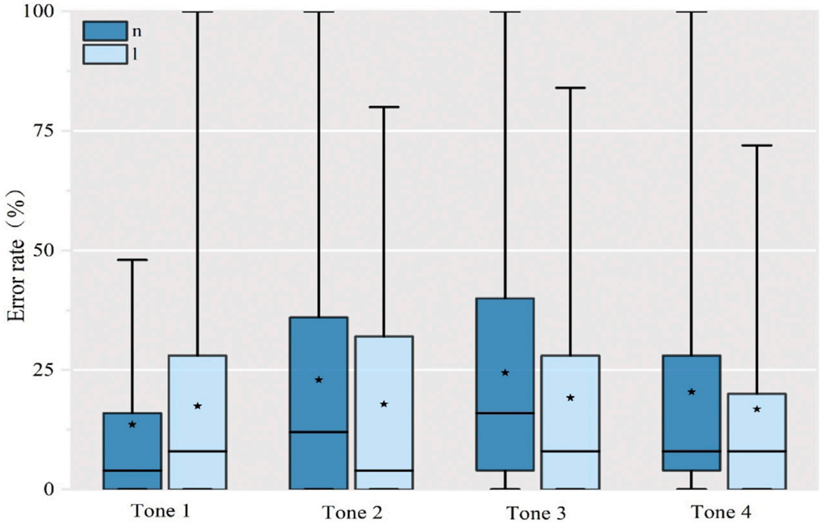


FIGURE 6 | Error rate of production of Standard Mandarin words by tone.

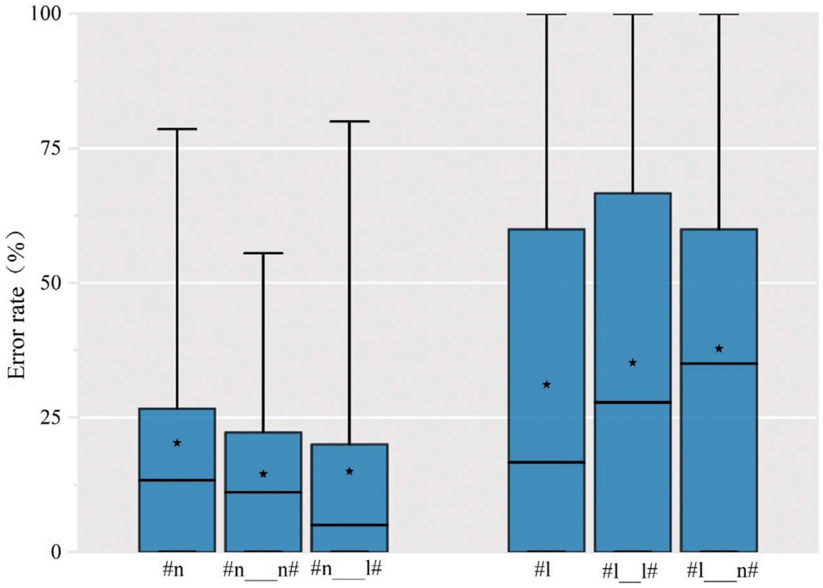


FIGURE 7 | Error rate of production for the syllable initial /n/ and /l/ in English words by word types.

evident (**Figure 8**). For both the /l/-only group and /n/-only group initial (n) was much more accurately produced than initial (l), suggesting that even for speakers for whom initial /n/ is not phonemic, the sound is more likely to be pronounced accurately. This may be because most Mandarin varieties license nasals as codas. In other words, the feature (nasal) exists in Mandarin even when it is not phonemic in a particular environment, and the /l/-only group is therefore as successful as the /n/-only group in using this feature in their L2 and L3 nasal production. It is in the production of initial /l/ that we see differences, with the /n/ group being much less accurate than the /l/ group. In light and lemon type words, the /n/ group was far less accurate, but in label type words, the /l/ and /n/ groups showed similar accuracy. This may be because words with initial and final /l/ helped them to focus on the production of initial /l/, whereas /l/ words with final /n/ resulted in a much higher mean error rate for the /n/ group, suggesting that the presence of both sounds made it more challenging for them to

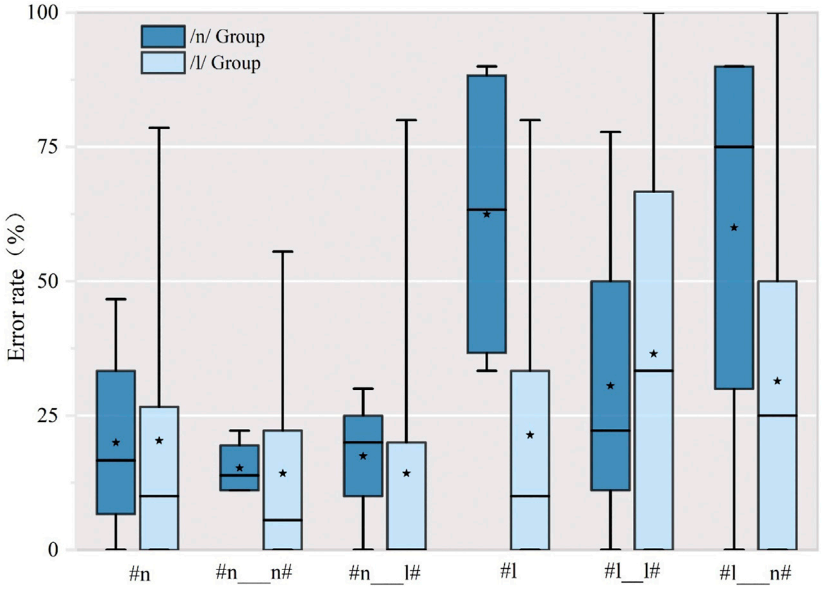


FIGURE 8 | Error rate of production for the syllable initial /n/ and /l/ in English words by subjects' variety.

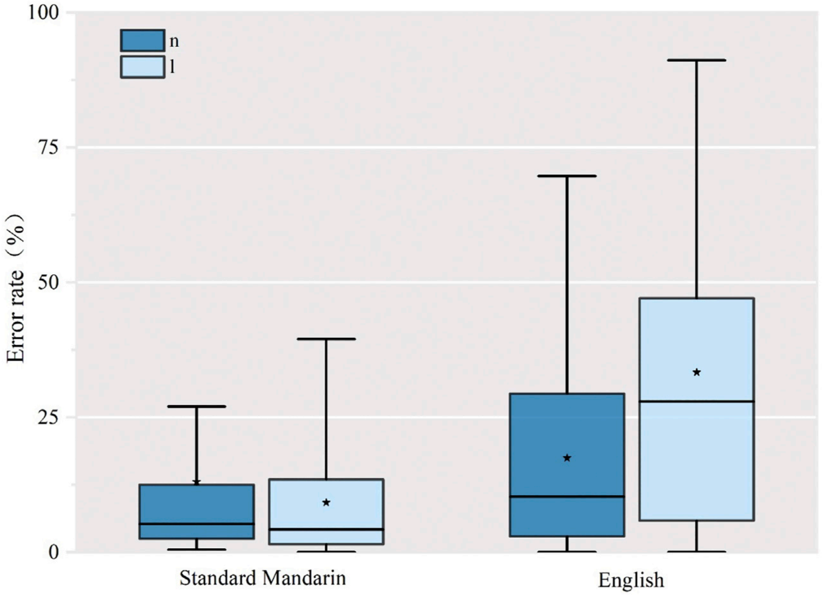


FIGURE 9 | Error rate of production of words with initial /n/ and /l/ in Standard Mandarin and English.

articulate the competing sounds in the same word. The /n/ group had a similar pattern for initial /n/ words, with n_l words being more difficult than n_n words, albeit with a smaller difference.

Influence of L1 on L2 (Standard Mandarin) and (L3) English Production

SWM speakers demonstrate production problems with the initial /l/ and /n/ when they speak in both Standard Mandarin and English (see **Figure 9**). In Standard Mandarin, the mean error rate for initial /n/ was slightly higher than for initial /l/. In English, their performance was different. The mean error rate for syllable initial /l/ in English (33.37%) was much higher than in the syllable initial /n/ (17.49%), with greater variability in performance.

A 2 × 2 within-subjects ANOVA with language (English and Standard Mandarin) and consonants /n l/ as factors showed a significant interaction [$F(1, 17) = 11.916$, p =0.003]. Simple effects indicated that English productions had a greater number of errors than Standard Mandarin both in nasal and lateral sounds. For lateral sounds, the errors in English were significantly higher than those in Standard Mandarin (M =0 .238, SD =0.049, p =0.000). There was no significant difference between production of the nasal in Standard Mandarin and English.

We computed the Spearman correlation of SWM speakers' production of Mandarin and English. The correlation showed that the accuracy of initial nasal production for English was highly correlated to the initial nasal production of Standard Mandarin (r =0.898, p <0.001) as was the production of initial lateral sounds in Standard Mandarin and English (r =0.778, p <0.001). This means that if our SWM subjects had more accurate pronunciation for initial nasals and laterals in Standard Mandarin, they were likely to also be more accurate in English.

Medial (l) and (n) in English, Alone and in Combination

In English, SWM speakers demonstrated serious difficulty when both nasal and lateral sounds occurred together in medial position (**Figure 10**), with #_ln_# words mispronounced at a higher rate than #_nl_# These words were mispronounced in a large majority of all productions. SWM speakers consistently produced only a single nasal or lateral rather than two distinct sounds. It appears that the /n/-only group, small though it was, had greater trouble. A paired t-test showed a significant effect of segment for both #_ln_# and #_nl_# type words, with subjects more frequently using a nasal sound than a lateral (**Table 8**). This production of only one sound where English native speakers would use two may indicate an assimilation to the manner of articulation of the other segment [cf. Gordon, 1957, pp. 280–82, in which (n) was deleted before other sonorants, including (l) and the presence of (r) with (l) or (n) led to (ll) and (nn)]. Alternatively, SWM speakers' tendency to produce only a single sound in English words may be due to phonotactic constraints from Mandarin syllabic structure which does not allow (l) in a coda nor (l) and (n) to occur next to each other. English words with a double spelled <nn> or <ll>, which required only a single sound, were produced much more accurately, suggesting that double spellings which do not indicate two different sounds may promote more accurate production for /l/ and for /n/.

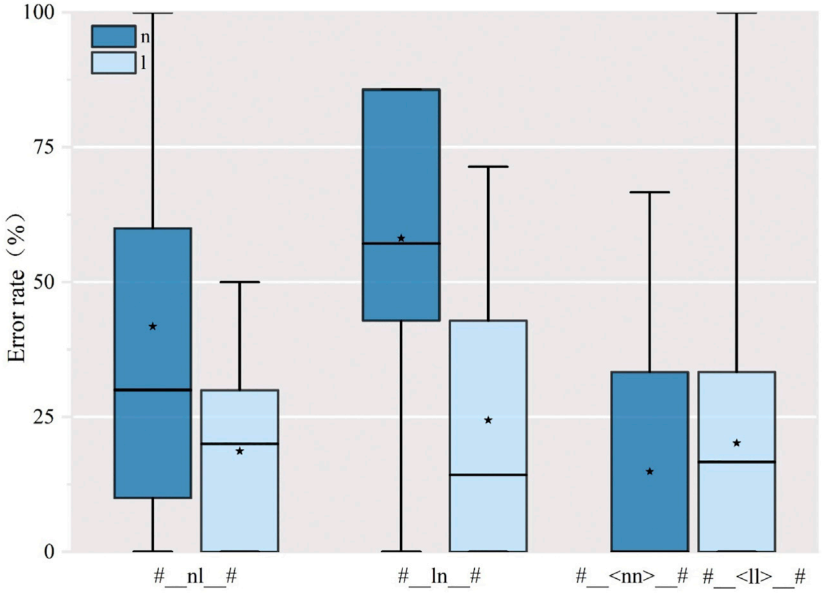


FIGURE 10 | Error rate of production for the English words with the syllable /-nl-/, /-ln-/ and <ll>, <nn>.

TABLE 8 | *t*-test results comparing the type of #_nl_# and #_ln_# words.

Comparison	*M*	*SD*	*t*	*df*	*p*
#_nl_#→l	0.197	0.169	−2.177	17	0.044
#_nl_#→n	0.441	0.340			
#_ln_#→l	0.230	0.219	−4.408	17	0.000
#_ln_#→n	0.612	0.207			

Notes: *#_nl_#→l means the type of #_nl_# words only produce a single lateral sound. #_nl_#→n means the type of #_nl_# words only produce a single nasal sound. Same pattern with the type of #_ln_# words.*

DISCUSSION

Research Questions

The first research question looked at the effect of L1 dialect on production of /l/ and /n/ in Standard Mandarin. Descriptions of SWM recognize difference within the variety regarding the phonological status of /l/ and /n/ (Li, 2004; Qian, 2010). Some areas have /n/ but not /l/, others the opposite, and yet others have a phonological contrast between the two sounds. Our findings confirm these descriptions of /l/-/n/ variation in SW Mandarin. All subjects who were from Yunnan province showed no difficulty in producing both (n) and (l). In contrast, the subjects who were from the /n/ dialect area showed more difficulty producing (l) than (n), as should be expected. Similarly, the subjects from an attested /l/ dialect area had the opposite pattern, with (n) showing a higher error rate than (l). This indicates that subjects' L1 phonological system influenced their productions in their additional languages. Those with a single category were more accurate in the pronunciation of items that matched the category in their dialect. However, they were also quite accurate overall, and their mispronunciations of /l/ and /n/ were a small minority of the total productions.

The second research question asked about the effects of environment on /l/ and /n/ production in Standard Mandarin. Environment was defined as differences in tone, whether the words were pronounced with or without a competitor sound in phrases, and the effects of the Sihu rhyming system in Mandarin, whose rhyme differences include four categories: rhymes beginning with (i), with (u), with (y), and with non-high vowels. The findings showed that the tone associated with the production of each word of did not affect (l) and (n) production accuracy. The presence or absence of a competitor sound in a phrase showed greater accuracy with /n/ when there was a competitor, and less accuracy for /l/. Finally, the effect of following vowel created different error patterns for (n) and (l) production. For a rhyme beginning with (i), both sounds had similar error rates [18.05% for (n), 21.45% for (l)], but for the other three rhymes, the error rates for (n) production were considerably higher than for (l). This suggests there may be some speakers for whom environment affects their production, and some studies of SWM have indicated that speakers have a phonological contrast between (n) and (l) before high front vowels but not elsewhere (Qian, 2010). Our data do not allow us to confirm this.

The third research question asked about the phonological environments most likely to elicit /n/ and /l/ pronunciation confusions in English. There was no clear effect of environment, but because our subjects included speakers from two phonological systems in uneven numbers, we can only offer descriptive data for the full group. The three initial /l/ environments had error rates near 30% or higher, while all of the /n/ environments had error rates at half the frequency of the pronunciation of /l/. This indicates that environment did not clearly affect the accuracy of the pronunciation of either sound in English. This is perhaps not surprising since our competitor sounds were in coda position and often in unstressed syllables. Mandarin varieties allow /n/ in codas [though their realization varies between (+consonantal) and (−consonantal)] but do not allow /l/ in coda position (Zhang and Levis, 2020). Child language acquisition shows a tendency to assimilate onset consonants to following consonants (e.g., yellow pronounced as *lellow, as in, for example, Gillespie and Greenberg, 2017). However, studies on tone perception House (1996) have shown that there are different types and amounts of information available from onsets and codas and the two do not consistently affect each other. For future research, initial /n/ and /l/ words could likely include all environments without affecting findings.

It should be pointed out that our SWM subjects correctly produced (n) and (l) most of the time. However, when listening to such speakers, it is our experience that confusion between /l/ and /n/ seems extremely common in spoken production. This may be because both /n/ and /l are frequently-occurring consonants in English (Hayden, 1950), and that listeners notice deviations not correct productions. An analogy to overregularization in child language is instructive. In a corpus of child language, Marcus et al. (1992) found that overregularized forms (e.g., flied for the past tense flew) occurred in only five percent of possible environments. Adult speakers notice these forms in child language and believe they are far more common than they are. (When our linguistics students are asked to estimate the frequency of such forms, giving them four answers, 5, 25, 45 and 65%, they usually pick the two highest frequencies.) Similarly, difficulties in distinguishing /l/ and /n/ in English speech may seem more common than they are. It may be that SWM speakers are using knowledge that /l/ and /n/ are different sounds in other languages to provide imperfect production of both sounds. Archibald (2005) argues that L2 learners can redeploy phonological knowledge from the features of their L1 for more accurate L2 production. For example, Japanese speakers do not have a distinction between /b/ and /v/ but do distinguish stops from continuants (Brown, 2000). As a result, they can use this knowledge of featural distinctions from their L1 to learn an unfamiliar distinction in another language using the same features. The overall accuracy of production suggests that SWM speakers may be able to use some knowledge from their L1 to produce the sounds in their L2 and L3. We cannot know whether SWM speakers similarly make use of their language learning experience (and perhaps the knowledge that their variety is marked in its pronunciation of /l/ and /n/) to pronounce phones more accurately in their L2 and L3. Brown (1998) suggests as much for Japanese speakers when she says that Japanese speakers' performance on the picture task (roughly 60% correct) is, on the whole, better than their performance on the auditory task (roughly 30% correct) (p. 169). If this is the case, it would suggest that in a task in which attention to their production is maximized (as in the reading tasks we used, see Brown, 1998), language learners may be able to demonstrate greater production accuracy than perception accuracy. This question must await perceptual studies.

The fourth research question examined the comparison of SWM speakers' /l/ and /n/ mispronunciations when reading in Standard Mandarin and English. This question sought to establish how the speakers' L1 affected the accuracy of their production in their other languages which had both phonemes. The results showed that the frequency of errors for Standard Mandarin was never more than 15%. Somewhat surprisingly, the error rate increased when the subjects read phrases with only one /l/ or /n/ word, while it decreased when they read phrases with two target sounds. Thus, phrases with competitor sounds were associated with greater accuracy. We expected that such competition would make mispronunciations more common, but this was not supported by the data.

For the English word reading, the /l/ and /n/ mispronunciation rate was higher than for Standard Mandarin, with a significant difference between the accuracy of production of /l/ in English and in Standard Mandarin. This is somewhat surprising. For Standard Mandarin words, both the /l/-only group and /n/-only group (refer to **Figure 2**) demonstrated greater difficulty with the production of (l), but in English, the number of /l/ errors was greater than in Mandarin while the number of /n/ errors remained similar for the L2 and L3 production. This suggests that even for /l/-only subjects, the production of (l) was more difficult.

Finally, we looked at the production accuracy for medial /l/ and /n/, both in words where the sounds were represented by orthographically doubled letters with a single sound and when both sounds were expected to be produced separately in sequence. Words with orthographic doubling of /l/ were produced more accurately than initial /l/ words, and those with medial <nn> were produced with accuracy similar to words with initial (n). In contrast, words with medial <ln> and /<nl> (e.g., walnut, only) were almost never correctly pronounced, with walnut type words being more difficult. In most mispronunciations, subjects produced only one of the sounds, more commonly (n). These words were different from the other words in that their syllable structure included a coda consonant in the first syllable followed by an onset consonant in the second syllable. When the coda was (l) (as in walnut), it involved a violation of the phonotactics of SWM. Zhang and Levis (2020) showed that final /l/ in English is almost never pronounced by SWM speakers, which may be the reason that word internal coda (l) would also be mispronounced. Although coda (n) is licensed in Mandarin, it is often produced as a nasalized vowel, and its production as a consonant occurs only about half the time. In general, Mandarin most commonly is produced with CV syllables, and in this study the two separate consonants in the middle of the word seemed to be assimilated to a single consonant realization in line with Mandarin syllable structure. In our results, this medial consonant is most likely to be (n). Although most of our subjects came from the /l/-only dialect area, and we would expect that they would favor medial (l) rather than (n), this was not the case. More importantly, because these medial combinations of (n) and (l) are almost certain to be mispronounced by SWM speakers, it would be valuable to know whether the mispronunciations affect how English listeners understand the words. The initial /l/-/n/ contrast has a high functional load (Catford, 1987), but we do not have equivalent measure for loss of a segment in the middle of a word. The earlier anecdote about pronouncing walnuts as wallets creates a different word that caused confusion for the listeners, but not all examples are so clear. Pronouncing only as either (oni) or (oli) while preserving the original stress pattern may be fully intelligible to a listener, and there may be no need in most cases to pronounce both sounds. Instead, it may be better to focus attention on particular words that cause loss of understanding.

/l/-/n/ in Cantonese and in Southwestern Mandarin

There is a substantially larger amount of research on /l/ and /n/ in Cantonese speakers' (l)-(n) production than in SWM, but our findings make clear that the lack of an /l/-/n/ contrast in the two Chinese languages follow different patterns. Cantonese, especially Hong Kong Cantonese, is in the process of a merger between the two sounds that suggests a complementary distribution, with (l) occurring in onset and (n) in coda position (Ng, 2017). This merger is more evident in younger speakers (Yeung, 1980; Tong and James, 1994), but HKC speakers perceive the differences between the two sounds, indicating that this advanced merger remains incomplete (Chan, 2011). Additionally, Cantonese speakers may infrequently pronounce (l) as (ɹ) (Chan, 2010), an error we found even more rarely in our SWM data. In SWM, any merger of the two sounds (l) and (n) took place in more distant history; some varieties of SWM currently have /l/ and others /n/, but the /l/ was a more difficult sound in English for all SWM speakers. There was also difficulty in our agreeing on identification of some English productions because the subjects sometimes appeared to produce a co-articulated sound with both nasal and lateral elements, sometimes leading to confusion for us as researchers.

Indeed, some productions defied easy categorization in the English data. One example is the word "nock" produced by subject D15 (provided in the **Supplementary Material** on the journal site). One author was confident that the initial sound was a lateral, but the other was certain it was a nasal. The acoustic measurements could not completely clarify our disagreement as the productions had acoustic evidence of both phonetic features. This indicates the need for an analysis of SWM (l) and (n) in terms of phonological features, especially because both /n/ and /l/ are ambiguous in regards to the feature (continuant), and their specific feature settings may be different in different languages (Mielke, 2005). It may also be the case that the features (nasal) and (lateral) are not distinguished for SWM speakers, causing us as listeners to hear their productions in terms of our own phonemic systems, which may not be equivalent in how the features distinguishing /n/ and /l/ are realized. As mentioned earlier, some researchers have argued that hybrid productions that include aspects of both lateral and nasal sounds occur within some Mandarin varieties (Chan, 1987; Soejima et al., 1990). How this would be described in terms of features is not clear, but it is likely that such productions would include both (nasal) and (lateral) features. This intriguing possibility must, however, be left for future research.

Influences on L2 and L3 Pronunciation

Although we do not have evidence of L1 patterns of /l/-/n/ productions for SW Mandarin itself, there is suggestive evidence that the L1 system affects production in both the L2 (Standard Mandarin) and L3 (English). The first piece of evidence is that the effect of the sub-variety was similar for both L2 and L3. Speakers with only /n/ in their L1 had higher numbers of /l/ errors in both the L2 and L3, and those with only /l/ in their phonological inventory had more /n/ errors in L2 and L3. This indicates a consistent effect of the L1 on both additional languages. A second piece of evidence is that /l/ appeared to be more challenging than /n/ for both L2 and L3 production, suggesting the same influence on both additional languages. Despite differences in frequency of errors, there was a strong correlation between the performance in both the L2 and L3. Those subjects with better accuracy in Standard Mandarin were likely to have better performance in English, and those with worse performance in English were likewise less accurate in Standard Mandarin. Evidence against the influence of the L2 on L3 can be seen in how English error rates for /l/ were higher than the frequency of Standard Mandarin errors. If L2 production had been influential on L3 production, we would expect fewer English errors than we in fact found. Instead, the language distance (Llama et al., 2010) between SWM and Standard Mandarin, two historically related varieties, was associated with greater accuracy in Standard Mandarin production than in English, which is more typologically distant.

The differing accuracy rates may also occur because of proficiency differences and different phonotactic constraints in the L2 and L3. SWM speakers learn Standard Mandarin from an earlier age and the language is presented consistently in the school setting. English is introduced later and is used primarily within English classes. SWM and Standard Mandarin share the same syllable structures, the same possible environments for both (n) (onset and coda) and (l) (onset only), and little difference in basic vocabulary, making the words produced more familiar. In contrast, the production of (n) and (l) in English occurs in unfamiliar linguistic environments (initial, in clusters, medially, and final) especially when the sounds were produced in clusters. The syllable structure and phonotactics of English are more complex than those of Mandarin, and this may make the production of (l) and (n) more challenging overall because /l/ and /n/ get pronounced in more environments and include different allophones.

Finally, SWM learners of English also have a higher learning burden for English, not only phonologically but also lexically and orthographically, contending with both more complex phonotactics and a more challenging knowledge of vocabulary. This suggests that when speaking English, SWM speakers may pay less attention to particular segments because of the extra attention needed to attend to meaning. In addition, English words are represented with a different writing system, which may add to the burden of accurate pronunciation. Latin orthography in the form of pinyin is used for elements of Mandarin reading, but its orthographic system does not have the same sound-spelling correspondence used in English. Even though <l> and <n>

often reflect the corresponding phonemes, English's notably opaque orthography adds to the burden of pronouncing words that learners may or may not have ever heard before.

Implications for Pronunciation Teaching

The /l/-/n/ contrast, as mentioned earlier, is similar in functional load to contrasts such as /l/-/ɹ/ and /p/-/b/, which are at the highest level of functional load in English (Catford, 1987; Brown, 1988). This means that it is very likely that confusions of these two phonemes in English will lead to misunderstandings, to challenges in processing speech, and to increased perceptions of accentedness (Munro and Derwing, 2006). Higher functional load is well-established as an important criterion for comprehensibility and accentedness in English (Munro and Derwing, 2006; Suzukida and Saito, 2019), but it is also likely that this contrast is equally important for learning other major languages such as French, Swedish, Russian, and Spanish, all of which have the /l/-/n/ contrast. Thus this challenge for SWM speakers (and Cantonese speakers) in speaking English may also be relevant for a variety of other additional languages.

Our findings indicate that /l/ may be a more challenging sound for SWM speakers in English as /l/ words had consistently higher error rates than /n/ words, and in words with both sounds (e.g, only, walnut), subjects were more likely to delete (l) and pronounce the (n). Nonetheless, the degree of difference between /n/ and /l/ accuracy for different subjects may not be relevant from a teacher's point of view since both the /n/ group and /l/ group had a similar combined accuracy on /n/ and /l/ words together. Although they had different proportions of /l/ and /n/ mispronunciation, from the viewpoint of a teacher, they would have similar numbers of errors. In English, /n/ and /l/ are two of the five most common consonants (Hayden, 1950), which means that listeners will be regularly confronted by word identification decisions when speaking with a SWM speaker who does not consistently produce a difference between the sounds. For SWM speakers, this may indicate a need for not only production but also perception training. If the difficulties that SWM speakers have in production are related to perception, then robust perceptual training using High Variability Phonetic Training (HVPT) may be required to help develop or strengthen distinct phonetic categories for /l/ and /n/. Although there is evidence that Cantonese speakers perceive the differences between (l) and (n) (Chan, 2011), because of the phonological distribution of the two sounds, there is no reason to assume this is the case for SWM speakers. Qian (2018) looked at the /l/-/n/ contrast in her HVPT study of Chinese learners. For her subjects, which included some SWM speakers, HVPT training appeared to help the few speakers who had this challenge. A study involving only SWM speakers would be valuable in determining the extent to which SWM speakers can improve their perception of the two sounds.

The results suggest that there are easier and more difficult environments for teaching. Single medial /l/ and /n/ seemed to be the most successfully pronounced and may be a good place to start with production and perception practice, while medial <ln> and <nl> were almost always mispronounced. These frequent mispronunciations of medial sound pairs may suggest that these should be a priority, but an intelligibility-based approach to teaching (Levis, 2018) asks for evidence that errors affect understanding. However, we have no evidence that producing only one sound where two are expected affects understanding in the way that functional load predicts problems for initial /l/ and /n/, which are likely to be more important because initial mispronunciations can lead listeners to access the wrong cohort of potential words (Marslen-Wilson and Welsh, 1978; Zielinski, 2008).

Limitations

A clear limitation of our study is that we had uneven numbers of subjects from different sub-varieties. It would be helpful to have larger and more equal numbers of subjects controlled for sub-variety. The exploratory nature of our study and the uneven numbers only made it possible for us to suggest trends in the production of /l/ and /n/. Another limitation was in the nature of the data we collected. There was no free speech, which may have provided different frequencies for confused sounds. This could also make it possible to quantify the comprehensibility of SWM speech by allowing listeners to rate the speech. Third, in collecting our recordings, we did not include distractor items though it is not clear that this affected how our subjects attended to the task. Even though all the words included (l) and (n), several subjects still asked the first author about the purpose of the study after reading the items. In addition, the uneven numbers of items produced added complications to our ability to compare frequencies. Some had 34 items (initial /n/ and /l/ for English) while other had as few as seven items or as many as 400 (Standard Mandarin words). Larger numbers of words almost certainly guaranteed greater numbers of unknown items, perhaps causing unforeseen difficulties in lexical access. Finally, our study did not include perception data. Because perception data is a better measure of whether there are intractable limitations in the L2 phonological system (Brown, 1998; Brown, 2000), and production data may not be directly related to perception, perception data would have allowed more confident interpretations of the production data.

Future Research

As mentioned above, it is very important that there be perception studies about how well SWM speakers hear (l) and (n). We know that the perception of even very difficult L2 contrasts can be improved (e.g., Bradlow et al., 1999), and that improved perception can also lead to improved production (Sakai and Moorman, 2018), although this is not always the case (Kartushina et al., 2015). If /l/-/n/ difficulties in production have a perceptual component, then high variability phonetic training approaches can be employed to build new category boundaries (Barriuso and Hayes-Harb, 2018) and complement production practice.

Another type of study that would be valuable would be an acoustic analysis of /l/ and /n/ production by SWM speakers. There are suggestions in previous research (Soejima et al., 1990) that SWM speakers may produce a sound that has both nasal and lateral realizations (e.g., a lateral nasal), indicating that a better analysis of the /l/-/n/ merger would be in terms of features rather than segments, especially for items that seemed to have features associated with phonologically distinct segments in languages like English. This would require a careful acoustic analysis that looked at productions that were in-between the category boundaries of the researchers. We identified quite a few potential tokens in our own listening, and they invariably led us as English-L1 and Standard Mandarin-L1 listeners to disagree on their categorization. If SWM speakers indeed have a sound that has characteristics of both nasals and laterals, it would suggest that /l/-/n/ problems are more like that attested for Japanese speakers with /l/-/ɹ/. Alternatively, we could be seeing pronunciations that result from speakers with one category speaking a language with a new category. Additionally, in regard to only and walnut type words, It may be that the subjects produced the singleton sound with greater duration, that is, as a geminate, to acoustically signal a difference between words with a single sound and these words. This would require production data for words like whining and whaling, in which there was only one medial consonant letter.

It would also be interesting to examine the L2 speech of different age groups. The learning of Standard Mandarin has become more standardized in China's education system over many decades, and our subjects began learning Standard Mandarin at the beginning of their schooling. Subjects who are older (e.g., over 40 years old) are less likely to have received early training in Standard Mandarin or English, and the effects of their L1 may be greater.

CONCLUSION

The SWM lack of contrast between /l/ and /n/ is a problem that helps explore the interface between L2 phonology and L2 pronunciation and is important for a number of reasons. First, from a point of view of L2 phonology, this study addresses a phonemic contrast that is common in a wide variety of languages (Yip, 2011) and is thus relevant to a variety of L2 learning contexts. SWM is unusual in not having such a distinction. Since our production data do not make clear the extent to which (l) and (n) represent two allophones of one phoneme in SWM, or whether they represent a sound in SWM sub-varieties that is neither /l/ nor /n/ (Soejima et al., 1990), data that bears on this question would be valuable for studying the phonetics and phonology of SW Mandarin.

From an L2 pronunciation point of view, the sounds in this study are important to L2 learners from a large population. More than 85,000,000 people speak Cantonese around the world, most of them L1 speakers in China and Hong Kong (Ethnologue, 2021), and up to 270,000,000 speakers speak some variety of SW Mandarin (Chinese Academy of Social Sciences, 2012). The production of these sounds in English thus becomes important because English is a required subject throughout their education beginning at age nine.

Our starting point for this study came from our interests as language teachers, which led us to examine how SWM speakers produced words with /l/ and /n/ in their additional languages. We were especially interested in the frequency and distribution of the subjects' /l/-/n/ errors in English because of the likelihood of such errors leading to unintelligibility in SWM speakers' English. It is our hope that this study provides a beginning that will lead to a more accurate understanding of where and how often such errors occur, and that our results may inform English pronunciation teachers' work with SWM learners.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Qufu Normal University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

WZ collected the recordings for the study, did a large part of the data analysis, and was equally involved in writing the paper. JL contributed to the data analysis, and was equally involved in writing the paper.

FUNDING

Thanks for the International Cooperation Program for Key Professors Funding by Shandong Provincial Education Department and the Funding from Shandong Social Science Project (No.11CWZZ16).

ACKNOWLEDGMENTS

We would like to thank our universities for supporting our research at Iowa State University. Thanks to the subjects and students who had helped to complete data-collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.639390/full#supplementary-material

REFERENCES

Amaro, J. C. (2017). Testing the Phonological Permeability Hypothesis: L3 Phonological Effects on L1 versus L2 Systems. *Int. J. Bilingualism* 21 (6), 698–717. doi:10.1177/1367006916637287

Ao, R., and Low, E. L. (2012). Exploring Pronunciation Features of Yunnan English. *English Today* 28 (3), 27–33. doi:10.1017/s0266078412000284

Aoyama, K., Guion, S. G., Flege, J. E., Yamada, T., and Akahane-Yamada, R. (2008). The First Years in an L2-Speaking Environment: A Comparison of Japanese Children and Adults Learning American English. *IRAL - Int. Rev. Appl. Linguistics Lang. Teach.* 46 (1), 61–90. doi:10.1515/iral.2008.003

Archibald, J. (2005). Second Language Phonology as Redeployment of L1 Phonological Knowledge. *The Can. J. Linguistics / La revue canadienne de linguistique* 50 (1), 285–314. doi:10.1353/cjl.2007.0000

Au, Y. N. A. (2011). The Markedness Theories and the Relationship between /n/ and /l/ in the English Syllable of Cantonese Speakers. Proceedings of the 16th Conference of Pan-Pacific Association of Applied Linguistics, August 2011, pp. 140–141.

Baker, H. D. R. (1993). Language atlas of China. Gen Ed. (Australia), S. A. Wurm, B. T'sou, D. Bradley; (China) Li Rong, Xiong Zhenghui, Zhang Zhenxing. [In two parts, boxed]. (Pacific Linguistics, Series C, no. 102). 36 maps, 22 sheets text material + addenda et corrigenda. Canberra: Australian Academy of the Humanities and the Chinese Academy of Social Sciences [and] Department of Linguistics, Research School of Pacific Studies, ANU; Hong Kong: Longman Group (Far East), 1987, 1989. *Bull. Sch. Oriental Afr. Stud.* 56 (2), 398–399. doi:10.1017/S0041977X0000598X

Bao, H. Q., and Lin, M. C. (2014). *ShiYan YuYingXue GaiYao (Revised Edition)*. Beijing: Peking University Press.

Barriuso, T. A., and Hayes-Harb, R. (2018). High Variability Phonetic Training as a Bridge from Research to Practice. *CATESOL J.* 30 (1), 177–194.

Bent, T., Bradlow, A. R., and Smith, B. (2007). "Phonemic Errors in Different Word Positions and Their Effects on Intelligibility of Non-native Speech," in *Second-language Speech Learning: The Role of Language Experience in Speech Perception and Production: A Festschrift in Honour of James E. Flege*. Editors O.-S. Bohn and M. J. Munro (Amsterdam: John Benjamins), 331–347. doi:10.1075/lllt.17.28ben

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. i. (1999). Training Japanese Listeners to Identify English /r/and /l/: Long-Term Retention of Learning in Perception and Production. *Perception & Psychophysics* 61 (5), 977–985. doi:10.3758/bf03206911

Brown, A. (1988). Functional Load and the Teaching of Pronunciation. *TESOL Q.* 22 (4), 593–606. doi:10.2307/3587258

Brown, C. A. (1998). The Role of the L1 Grammar in the L2 Acquisition of Segmental Structure. *Second Lang. Res.* 14, 136–193. doi:10.1191/026765898669508401

Brown, C. (2000). "The Interrelation between Speech Perception and Phonological Acquisition from Infant to Adult," in *Second Language Acquisition and Linguistic Theory*. Editor J. Archibald (Malden, MA: Blackwell), 4–63.

Cabrelli Amaro, J. (2012). "L3 Phonology," in *Third Language Acquisition in Adulthood*. Editors J. Cabrelli Amaro, S. Flynn, and J. Rothman (Amsterdam/ Philadelphia: John Benjamins), 33–60. doi:10.1075/sibil.46.05ama

Catford, J. C. (1988). "Functional Load and Diachronic Phonology," in *The Prague School and its Legacy*. Editor Y. Tobin (Amsterdam: John Benjamins), 3–19. doi:10.1075/llsee.27.04cat

Catford, J. C. (1987). "Phonetics and the Teaching of Pronunciation: a Systemic Description of English Phonology," in *Current Perspectives on Pronunciation: Practices Anchored in Theory*. Editor J. Morley (Washington, DC: TESOL Press), 87–100.

Chan, A. Y. W. (2010). Advanced Cantonese ESL Learners' Production of English Speech Sounds: Problems and Strategies. *System* 38 (2), 316–328. doi:10.1016/j.system.2009.11.008

Chan, A. Y. W., and Li, D. C. S. (2000). English and Cantonese Phonology in Contrast: Explaining Cantonese ESL Learners' English Pronunciation Problems. *Lang. Cult. Curriculum* 13 (1), 67–85. doi:10.1080/07908310008666590

Chan, A. Y. W. (2014). The Perception and Production of English Speech Sounds by Cantonese ESL Learners in Hong Kong. *Linguistics* 52 (1), 35 – 72. doi:10.1515/ling-2013-0056

Chan, A. Y. W. (2011). The Perception of English Speech Sounds by Cantonese ESL Learners in Hong Kong. *TESOL Q.* 45 (4), 718–748. doi:10.5054/tq.2011.268056

Chan, M. K. M. (1987). Post-stopped Nasals in Chinese: An Areal Study. *UCLA Working Pap. Phonetics* 68, 73–120.

Chen, X. (2013). *Study on the Origin and Change of the Phonetic System of Yunnan Dialect*. Dissertation. Tianjin, China: Nankai University.

Chinese Academy of Social Sciences (2012). *Zhōngguó Yǔyán Dìtú Jí (Dì 2 Bǎn): Hànyǔ Fāngyán Juǎn [Language Atlas of China*. in *Chinese Dialect Volume]*. 2nd edition (Beijing: The Commercial Press).

Deterding, D. (2010). ELF-based Pronunciation Teaching in China. *Chin. J. Appl. Linguistics* 33 (6), 3–15.

Deterding, D. (2006). The Pronunciation of English by Speakers from China. *Eww* 27 (2), 175–198. doi:10.1075/eww.27.2.04det

Deterding, D., Wong, J., and Kirkpatrick, A. (2008). The Pronunciation of Hong Kong English. *Eww* 29 (2), 148–175. doi:10.1075/eww.29.2.03det

Duanmu, San. (2000). *The Phonology of Standard Chinese*. New York, Oxford: Oxford University Press.

Ethnologue (2021). Chinese, Yue. Availableat: https://www-ethnologue-com.eu1.proxy.openathens.net/language/yue (Accessed June 1, 2021).

Feng, S. L. (2016). Mandarin Chinese Is a Stress Language. *YUYAN KEXUE* (05), 449–473.

Gillespie, L. G., and Greenberg, J. D. (2017). Empowering Infants' and Toddlers' Learning through Scaffolding. *YC Young Child.* 72 (2), 90–93.

Gordon, E. V. (1957). *An Introduction to Old Norse*. Oxford, UK: Oxford University.

Goto, H. (1971). Auditory Perception by normal Japanese Adults of the Sounds "L" and "R". *Neuropsychologia* 9, 317–323. doi:10.1016/0028-3932(71)90027-3

Hammarberg, B. (2001). "Chapter 2. Roles of L1 and L2 in L3 Production and Acquisition," in *Cross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. Editors J. Cenoz, B. Hufeisen, and U. Jessner (Tonawanda, NY, United States: Multilingual Matters Ltd), 21–41. doi:10.21832/9781853595509-003

Han, S. X. (2014). A Research into the Negative Transfer of Gansu Dialect on the English Phonetic Learning. *J. Longdong Univ.* 25 (6), 45–47.

Hayden, R. E. (1950). The Relative Frequency of Phonemes in General-American English. *WORD* 6 (3), 217–223. doi:10.1080/00437956.1950.11659381

Hazan, V., Sennema, A., Iba, M., and Faulkner, A. (2005). Effect of Audiovisual Perceptual Training on the Perception and Production of Consonants by Japanese Learners of English. *Speech Commun.* 47 (3), 360–378. doi:10.1016/j.specom.2005.04.007

House, D. (1996). Differential Perception of Tonal Contours through the Syllable. Proceedings of Fourth International Conference on Spoken Language Processin. *ICSLP*, 96. IEEE, 2048–2051.

Hu, F. (2007). *International Congress of Phonetic Sciences XVI*, 1405–1408. Availableat: http://www.icphs2007.de/conference/Papers/1127/1127.pdf .Post-oralized Nasal Consonants in Chinese Dialects—Aerodynamic and Acoustic Data

Huang, B. R., and Liao, X. D. (2011). *Modern Chinese*. Beijing, China: Higher Education Press.

Hung, T. T. N. (2000). Towards a Phonology of Hong Kong English. *World Englishes* 19 (3), 337–356. doi:10.1111/1467-971x.00183

IIE (2019). The Power of International Education: Number of International Students in the United States Hits All-Time High. Availableat: https://www.iie.org/Why-IIE/Announcements/2019/11/Number-of-International-Students-in-the-United-States-Hits-All-Time-High (Accessed March 2, 2020).

Kang, O., and Moran, M. (2014). Functional Loads of Pronunciation Features in Nonnative Speakers' Oral Assessment. *Tesol Q.* 48 (1), 176–187. doi:10.1002/tesq.152

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., and Golestani, N. (2015). The Effect of Phonetic Production Training with Visual Feedback on the Perception and Production of Foreign Speech Sounds. *The J. Acoust. Soc. America* 138 (2), 817–832. doi:10.1121/1.4926561

Koffi, E. (2019). An Acoustic Phonetic Account of the Confusion between [l] and [n] by Some Chinese Speakers. *Linguistic Portfolios* 8 (6), 64–73.

Kurpaska, M. (2010). *Chinese Language(s): A Look through the Prism of the Great Dictionary of Modern Chinese Dialects*. Berlin, New York: De Gruyter Mouton.

Ladefoged, P., and Johnson, K. (2011). *A Course in Phonetics*. 6th Ed.. Boston: Wadsworth.

Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Languages*. MA: Blackwell: Cambridge.

Lan, Q. Y. (1995). Hechi Dialect Phonology. *Acad. Forum* 5, 51–56.

Leung, M., and Sharma, Y. (2020). International Student Flows Extend Fallout from Coronavirus Outbreak. Availableat: https://www.universityworldnews.com/post.php?story=20200129174911175 (Accessed to March 2, 2020).

Levis, J. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. New York: Cambridge University Press.

Li, B., and Hua, C. C. (2014). Effects of Visual Cues on Perception of Non-native Consonant Contrasts by Chinese EFL Learners. *Asian EFL J.* 16 (1), 271–295.

Li, L. (2009). Classification/distribution of Southwest Mandarin. *Dialects* 1, 72–87.

Li, M. (2017). *Study on the Phonetic System of the Dialects in the Area of Nanchong, Sichuan Province*. Dissertation. Chengdu, China: Sichuan Normal University.

Li, Q. Q. (2002). *Jishou Fāng Yán Yán Jiū [Study of Jishou Dialect]*. Beijing: The Ethnic Publishing House.

Li, X. (2004). *Comprehensive Phonetic Study of Southwestern Mandarin*. Dissertation. Shanghai, China: Shanghai Normal University.

Li, Y. M., Zhao, S. J., and He, L. (2020). The Practice of and Reflections on "Epidemic Language Service Corps". *Chin. J. Lang. Pol. Plann.* 27 (3), 23–30.

Lin, T., Wang, L. J., and Wang, Y. J. (2013). YuYinXue JiaoCheng *(Revised Edition)*. Beijing: Peking University Press.

Llama, R., Cardoso, W., and Collins, L. (2010). The Influence of Language Distance and Language Status on the Acquisition of L3 Phonology. *Int. J. Multilingualism* 7 (1), 39–57. doi:10.1080/14790710902972255

Luo, Y. H. (2016). Phonology of Guizhou Zunyi Dialect. *Mod. Chin.* 05, 43–45.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., et al. (1992). Overregularization in Language Acquisition. *Monogr. Soc. Res. Child Development* 57, i–178. doi:10.2307/1166115

Marslen-Wilson, W. D., and Welsh, A. (1978). Processing Interactions and Lexical Access during Word Recognition in Continuous Speech. *Cogn. Psychol.* 10 (1), 29–63. doi:10.1016/0010-0285(78)90018-x

Mielke, J. (2005). Ambivalence and Ambiguity in Laterals and Nasals. *Phonology* 22 (2), 169–203. doi:10.1017/s0952675705000539

Ming, S. R. (2005). A Study on the Law of Correspondence between the Initial Consonants in Bijie Dialect and the Initial Consonants of Beijing Pronunciation. *J. Guizhou Education Inst. (Social Science)* 01, 75–77+103.

Ming, S. R. (1997). The Vowels, Rhymes and Tones of the Bijie Dialect. *J. Guizhou Education Univ.* 02, 82–84.

Modern Chinese Teaching and Research Program, Chinese Department, Peking University (2002). *Modern Chinese*. Beijing: The Commercial Press.

Munro, M., Derwing, T., and Thomson, R. (2015). Setting Segmental Priorities for English Learners: Evidence from a Longitudinal Study. *Int. Rev. Appl. Linguistics Lang. Teach.* 53 (1), 39–60. doi:10.1515/iral-2015-0002

Munro, M. J., and Derwing, T. M. (2006). The Functional Load Principle in ESL Pronunciation Instruction: An Exploratory Study. *System* 34 (4), 520–531. doi:10.1016/j.system.2006.09.004

Ng, C. L. C. (2017). Merger of the Syllable-Initial [n-] and [l-] in Hong Kong Cantonese. (Outstanding Academic Papers by Students (OAPS), City University of Hong Kong). Availableat: http://lbms03.cityu.edu.hk/oaps/lt2017-4235-ncl190.pdf.

Pater, J., and Werle, A. (2003). Direction of Assimilation in Child Consonant harmony. *The Can. J. Linguistics/La revue canadienne de linguistique* 48 (2), 385–408. doi:10.1353/cjl.2004.0033

Pennington, R., and Saunders, J. (2013). *Introduction to the Documentary Corpus of Guiliu, a Southwestern Mandarin Dialect*. Course paper.

Qian, M. (2018). *An Adaptive Computational System for Automated, Learner-Customized Segmental Perception Training in Words and Sentences: Design, Implementation, Assessment*. Doctoral Dissertation. Ames, IA, United States: Iowa State University

Qian, Z. Y. (2010). *A Study of Mandarin Chinese Dialect*. Jinan, China: Qilu Publishing House.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Computer Program. Version 2.9.2. Vienna, Austria: R Foundation for Statistical Computing. Availableat: http://www.R-project.org.

Raver-Lampman, G., Toreno, F., and Bing, J. (2015). An Acceptable Non-alveolar Japanese /s/. *Jslp* 1 (2), 238–253. doi:10.1075/jslp.1.2.05rav

Richards, M. (2012). "Helping Chinese Learners Distinguish English /l/ and /n/," in Proceedings Of the 3rd Pronunciation In Second Language Learning And Teaching Conference, *Sept. 2011*. Editors J. Levis and K. LeVelle (Ames, IA: Iowa State University), 161–167.

Sakai, M., and Moorman, C. (2018). Can Perception Training Improve the Production of Second Language Phonemes? A Meta-Analytic Review of 25 Years of Perception Training Research. *Appl. Psycholinguistics* 39 (1), 187–224. doi:10.1017/s0142716417000418

Sewell, A. (2017). Functional Load Revisited. *Jslp* 3 (1), 57–79. doi:10.1075/jslp.3.1.03sew

Soejima, K., Imagana, H., and Kiritani, S. (1990). Alternation between Stop Nasal and (Nasalized) Flap or Lateral. *Ann. Bull. RILP* 24, 131–144.

Su, P. (2010). Photosynthesis of C4 Desert Plants. *Anhui Lit.* 8, 243–259. doi:10.1007/978-3-642-02550-1_12

Sun, Y. (2011). *Studies on Dialects of Southwestern Mandarin in Sichuan Province - Their Phonological Systems and Historical Development*. Dissertation. Hangzhou, China: Zhejiang University.

Suzukida, Y., and Saito, K. (2019). Which Segmental Features Matter for Successful L2 Comprehensibility? Revisiting and Generalizing the Pedagogical Value of the Functional Load Principle. *Lang. Teach. Res.* 136216881985824.

Tong, K. S., and James, G. (1994). *Colloquial Cantonese*. New York: Psychology Press.

Třísková, H. (2011). Chapter Four. The CVX Theory of Syllable Structure. *Archiv orientální* 79 (1), 99–127. doi:10.1163/ej.9789004187405.i-464.24

Wang, H., Wu, D., Chen, H., and Chen, J. C. (2009). Research on Yibin Dialect and Mandarin Chinese in Initial Part and Final Part of a Syllable. *J. Chongqing Univ. Arts Sci. (Social Sci. Edition)* 1, 86–88.

Wang, P. (1981). The Phonological System of Guiyang Dialect. *Dialect* 02, 122–130.

Wang, S. J. (1986). The Phonetic Differences between Qujing Dialect and Mandarin Chinese. *J. Qujing Teach. Coll.* 00, 75–78.

Wei, J. (2018). A Comparative Analysis of Consonants between Mengzi Dialect and Standard Mandarin. *Data Cult. Education* 15, 24–26.

Wong, M. (2008). Can Consciousness-Raising and Imitation Improve Pronunciation?. *Int. J. Learn. Annu. Rev.* 15 (6), 43–48. doi:10.18848/1447-9494/cgp/v15i06/45808

Wrembel, M. (2013). "Foreign Accent Ratings in Third Language Acquisition: The Case of L3 French," in *Teaching and Researching English Accents in Native and Non-native Speakers*. Editors E. E. Waniek-Klimczak and L. Shockey (Berlin: Springer), 31–47. doi:10.1007/978-3-642-24019-5_3

Xiao, Y. F. (1996). Anita Hill Resigns Her Professorship at the University of Oklahoma College of Law. *J. Blacks Higher Education* 01, 69–81. doi:10.2307/2962837

Xin, M. (2013). The Phonetic System of Guizhou Xingyi Dialect. *J. Xingyi Normal Univ. Nationalities* 01, 57–61.

Yeung, S. W. (1980). *Some Aspects of Phonological Variations in the Cantonese Spoken in Hong Kong*. Dissertation. [Hong Kong]: University of Hong Kong.

Yip, M. (2011). "Lateral Consonants," in *The Blackwell Companion to Phonology*. Editors M. van Oostendorp, C. Ewen, E. Hume, and K. Rice (Wiley Blackwell), 1–26. doi:10.1002/9781444335262.wbctp0031

Yuan, B. L. (1996). Anshun Urban Dialect Phonological System. *Anshun Shizhuan Xuebao* 01, 61–73.

Zee, E. (1999). *Chinese (Hong Kong Cantonese). The Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, 58–60.

Zeng, X. G. (2009). Nanchong Fāng Yán Yán Jiū *[Study of Nanchong Dialect]*. Chengdu: Sichuan People's Publishing House.

Zhang, J. S. (2021). Also On Chinese Word Stress. *Chin. Lang.* 01, 43–55+127.

Zhang, W. (2007). Alternation of [n] and [l] in Sichuan Dialect, Standard Mandarin and English: a Single-Case Study. *Leeds Working in Linguistics and Phonetics* 12, 156–173.

Zhang, W., and Levis, J. (2020). "Southwestern Mandarin Speakers' Production of English Word-Final /l/ and /n/," in Proceedings Of the 11th Pronunciation In Second Language Learning And Teaching Conference, *ISSN 2380-9566, Northern Arizona University, September 2019*. Editors O. Kang, S. Staples, K. Yaw, and K. Hirschi (Ames, IA: Iowa State University), 292–303.

Zhang, Z. R. (2012). Analysis of the Influence of Gansu Dialect on English Pronunciation. *J. Gansu Lianhe Univ. (Social Sciences)* 28 (2), 73–75.

Zheng, Q. J. (1999). *Changde Fāng Yán Yán Jiū [Study of Changde Dialect]*. Zhangsha: Hunan Education Publishing House.

Zhou, R. (2018). A Review on Mandarin Word Stress Study in 60 Years. *Lang. Teach. Linguistic Stud.* (06), 102–112.

Zhou, Y. Y. (2014). *The Experimental Analysis of Dialectal Phonetic Systems and Research on Their Linguistic Geography in Mianyan Sichun*. Dissertation. Chengdu, China: Sichuan Normal University.

Zielinski, B. (2006). The Intelligibility Cocktail: An Interaction between Speaker and Listener Ingredients. *Prospect: Aust. J. TESOL* 21 (1), 22–45.

Zielinski, B. W. (2008). The Listener: No Longer the Silent Partner in Reduced Intelligibility. *System* 36 (1), 69–84. doi:10.1016/j.system.2007.11.004

Conflict of Interest: 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.

The reviewer, AS, declared a past co-authorship with one of the authors, JL.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

GENERAL COMMENTARY
published: 13 December 2021
doi: 10.3389/fpsyg.2021.774721

Commentary: L2 Phonology: Where Theory, Data, and Methods Meet

Christine E. Shea[1,2]*

[1] *Department of Spanish and Portuguese, University of Iowa, Iowa City, IA, United States,* [2] *Department of Linguistics, University of Iowa, Iowa City, IA, United States*

Keywords: L2 phonology acquisition, L2 speech acquisition, phonology, L2 phonetics, phonetics and phonology

A commentary on selected articles from the Research Topic "L2 Phonology Meets L2 Pronunciation"

Developmental Sequences in Second Language Phonology: Effects of Instruction on the Acquisition of Foreign sC Onsets
by Cardoso, W., Collins, L., and Cardoso, W. (2021). Front. Commun. 6:662934. doi: 10.3389/fcomm.2021.662934

The Effects of L1 English Constraints on the Acquisition of the L2 Spanish Alveopalatal Nasal
by Stefanich, S., and Cabrelli, J. (2021). Front. Psychol. 12:640354. doi: 10.3389/fpsyg.2021.640354

The Southwestern Mandarin /n/-/l/ Merger: Effects on Production in Standard Mandarin and English
by Zhang, W., and Levis, J. M. (2021). Front. Commun. 6:639390. doi: 10.3389/fcomm.2021.639390

OPEN ACCESS

Edited by:
Juhani Järvikivi,
University of Alberta, Canada

Reviewed by:
John Archibald,
University of Victoria, Canada

****Correspondence:***
Christine E. Shea
christine-shea@uiowa.edu

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

Received: *12 September 2021*
Accepted: *05 October 2021*
Published: *13 December 2021*

Citation:
Shea CE (2021) Commentary: L2 Phonology: Where Theory, Data, and Methods Meet. Front. Psychol. 12:774721. doi: 10.3389/fpsyg.2021.774721

The field of L2 phonology studies the abstract representations created by L2 learners over the course of acquisition. L2 pronunciation, on the other hand, addresses concrete aspects of L2 speech, related to primarily to intelligibility, comprehensibility, and accentedness (Munro and Derwing, 1995). L2 phonology can be conceptualized as the frame in which L2 pronunciation develops or, where theory, data, and methods meet. In other words, pronunciation does not happen without phonology.

In their contribution "Developmental sequences in second language phonology: Effects of instruction on the acquisition of foreign sC onsets," Cardoso, Collins and Cardoso examine how three different types of instruction interact with the production of /s/ + /l n t/ clusters by L1 Brazilian Portuguese/L2 English speakers (Cardoso et al., 2021). Previous research suggests that the order of acquisition of these non-native clusters will be affected by their internal sonority profile, or markedness, understood as relative degree of linguistic complexity [see Gass and Ard (1980) for early work on L2 syntax using universal hierarchies in the acquisition of relative clauses; (Cardoso and Liakin, 2009)]. From that perspective, /st/ clusters will be more difficult to acquire because there is minimal sonority fall from /s/ to /t/, compared to /s/ + /n/ or /l/ [see Yavaş (2010) for a similar analysis in L1 speech development]. Cardoso et al. ask whether the design of L2 pronunciation teaching should follow this natural, or universal order of acquisition and focus first on clusters that are predicted to be more easily acquired (/s/ + /n/ or /l/). Alternatively, research from L1 phonological development in disordered populations has shown that focusing first on more complex structures cascades down to the acquisition of less-marked structures (Gierut, 1999). A third possibility is also considered by Cardoso et al., where both complex and simple structures are taught together. Their results show that the pedagogical intervention emphasizing the most marked structure led to the greatest improvement in production of that cluster (st/) and also benefitted production of the other two clusters. Thus, the authors conclude that teaching oriented to more complex structures can lead to the acquisition and development of less marked structures as well. The results from Cardoso et al. show that abstract universal concepts such as sonority and phonotactic constraints influence L2 phonological acquisition and crucially, interact with explicit classroom learning. This has important implications for classroom pronunciation materials design, which typically does not take into account phonological universals of this type.

Speech-sound complexity is also at play in the contribution from Stefanich and Cabrelli (S&C) "The Effects of L1 English Constraints on the Acquisition of the L2 Spanish Alveopalatal Nasal" (Stefanich and Cabrelli, 2021). The authors examine the acquisition of the Spanish sound /ɲ/ by beginner and advanced L1 English speakers. The sound /ɲ/ occurs in syllable-onset position in Spanish words such as *montaña* "mountain" and *año* "year" and is not part of the English phonemic inventory. The closest English sound is the heterosyllabic sequence [n.j] in words such as *canyon*[1]. Their findings show that while the new contrast is learnable, neither group produced the target forms in the same way as the L1 Spanish speakers. The authors examined the production of F1 and F2 formant cues across the different groups. A rather unexpected finding was that the advanced group relied upon L1 representations to a greater extent than the beginner group. S&C speculate that advanced L2 learners' representations may, in fact, reflect a U-shaped development pattern; that is, the advanced learners have realized that /ɲ/ is not /nj/, but they are still not able to produce [ɲ] in target-like fashion and revert to the closest L1 sound (Tessier, 2019).

Zhang and Levis in their article "The Southwestern Mandarin /n/-/l/ merger: Effects on production in Standard Mandarin and English" examine how speakers of Southwestern Mandarin, a dialect of Mandarin in which the contrast between [l] and [n] is merged, produce these segments in initial and word-medial positions in L3 English and L2 Standard Mandarin. The results from a reading task revealed an interesting asymmetry with respect to the /n/-/l/ merger whereby in English, participants neutralized in favor of [l] while in Standard Mandarin, neutralization was in favor of [n], suggesting an interaction between the order of language acquisition (L2 vs. L1) and potentially, the role of each sound in the phonological system of each language.

These three studies make interesting and important contributions to the field of L2 phonological development in terms of how abstract structure affects the order of acquisition (Cardoso et al., 2021), syllable structure (S&C), and L2 vs. L3 neutralization asymmetries (Zhang and Levis, 2021). Moving forward, I would encourage researchers in the field of L2 phonology to consider two important additional factors related to (a) the type of data collected and (b) the pool of participants.

In terms of the type of data collected, the field of L2 phonology would benefit greatly from more studies focused on longitudinal development. While recent studies have examined longitudinal phonetic development (Nagle, 2019; Casillas, 2020), there are relatively few that use empirical, quantitative data to analyze phonological development in this way (e.g., changes over time in syllable structure representation, such as that examined by Cardoso et al. and S&C). As well, researchers should consider the longitudinal development of perception and production together [Nagle, 2021; see Nagle and Baese-Berk (2021) for an overview] to determine how tightly coupled these may be at the phonological level.

Related to this is a call for a clearer definition of what is meant by "phonological representations." S&C provide a good example of how future researchers may go about this in their contribution. These authors establish clear hypotheses regarding changes in the representation of the alveopalatal nasal and how, as the learner gains experience with the target language, each change in representation might manifest in production. Furthermore, S&C also characterize what learners needed to adjust in terms of their syllable representations to successfully produce the alveopalatal nasal. Not every study that purports to examine L2 phonology is as clear on the learning task, possible outcomes and implications for representational claims.

Both Cardoso et al. and S&C address issues related to complexity and re-alignment of L1 phonological representations as part of the task facing their participants. I would encourage future researchers to consider these issues as they relate to Heritage Speakers (HS). While HS are (of course) distinct from L2 learners, who are the focus of this Research Topic, data from this group of speakers can contribute meaningfully to the development of phonological representations because (a) HS are naturalistic learners and (b) HS undergo a shift in language dominance at some point during childhood, due to a decline in the amount of input received. The study of HS phonological development in both the heritage and dominant language would allow researchers to consider age and input as separate factors (Flege, 2018; Flege and Bohn, 2021; p.c. B. McMurray) and help understand the way phonological development is affected by each.

Finally, future research should also consider how phonology relates to the real-time unfolding of speech cues in lexical recognition. Crucially, L2 (and L3) phonological representations are only "functional" in so far as they encode lexical items, or, at the very least, can serve as potential representations for the target language (Darcy et al., 2013; Cook et al., 2016). Recent work has shown that both phonology and phonetics play a key role in lexical encoding of contrasts. Combining this with studies examining how cues to, say, syllable structure unfold in real time and affect lexical competition in bilinguals (Sarrett et al., 2021) is necessary to gain a fuller picture of L2 phonological development.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

[1] Phonotactic constraints in English prohibit homosyllabic *[.nj] sequences in most dialects (Kulikov, 2010). Words such as *news* [njuz] can be produced with the [nj] sequence in palatizing dialects.

REFERENCES

Cardoso, W., Collins, L., and Cardoso, W. (2021). Developmental sequences in second language phonology: effects of instruction on the acquisition of foreign sc onsets. *Front. Commun.* 6:662934. doi: 10.3389/fcomm.2021.662934

Cardoso, W., and Liakin, D. (2009). "When input frequency patterns fail to drive learning: The acquisition of sC onset clusters," in *Recent Research in Second Language Phonetics/Phonology: Perception and Production*, 174–202.

Casillas, J. V. (2020). The longitudinal development of fine-phonetic detail: stop production in a domestic immersion program. *Lang. Learn.* 70, 768–806.

Cook, S. V., Pandža, N. B., Lancaster, A. K., 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

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/ml.8.3.06dar

Flege, J. (2018). It's input that matters most, not age. *Biling. Lang. Cogn.* 21, 919–920. doi: 10.1017/S136672891800010X

Flege, J. E., and Bohn, O. S. (2021). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress*, ed R. Wayland (Cambridge : Cambridge University Press), 3–83.

Gass, S., and Ard, J. (1980). L2 data: their relevance for language universals. *TESOL Q.* 443–452.

Gierut, J. (1999). Syllable onsets: clusters and adjuncts in acquisition. *J. Speech Lang. Hear. Res.* 42, 708–726. doi: 10.1044/jslhr.4203.708

Kulikov, V. (2010). "Features, cues, and syllable structure in the acquisition of Russian palatalization by L2 American learners," in *Achievements and Perspectives in SLA of Speech: New Sounds*, Vol. 1, eds M. Wrembel, M. Kul, and K. Dziubalska-Kolaczyk (Frankfurt am Main: Peter Lang Verlag), 193–204.

Munro, M. J., and Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* 45, 73–97.

Nagle, C. L. (2019). A longitudinal study of voice onset time development in L2 Spanish stops. *Appl. Linguist.* 40, 86–107. doi: 10.1093/applin/amx011

Nagle, C. L. (2021). Revisiting perception-production relationships: Exploring a new approach to investigate perception as a time-varying predictor. *Lang. Learn.* 71, 243–279. doi: 10.1111/lang.12431

Nagle, C. L., and Baese-Berk, M. M. (2021). Advancing the state of the art in L2 speech perception-production research: Revisiting the theoretical assumptions and methodological practices. *Stud. Second. Lang. Acquis.* 1–26. doi: 10.1017/S0272263121000371

Sarrett, M. E., Shea, C., and McMurray, B. (2021). Within-and between-language competition in adult second language learners: implications for language proficiency. *Lang. Cogn. Neurosci.* 1–17. doi: 10.31234/osf.io/bwv7c

Stefanich, S., and Cabrelli, J. (2021). The effects of l1 english constraints on the acquisition of the l2 spanish alveopalatal nasal. *Front. Psychol.* 12:640354. doi: 10.3389/fpsyg.2021.640354

Tessier, A. (2019). U-shaped development in error-driven child phonology. *Wires Cog. Sci.* 10:e1505. doi: 10.1002/wcs.1505

Yavaş, M. (2010). Sonority and the acquisition of/s/clusters in children with phonological disorders. *Clin. Linguist. Phon.* 24, 169–176. doi: 10.3109/02699200903362901

Zhang, W., and Levis, J. M. (2021). The southwestern mandarin /n/-/l/ merger: effects on production in standard mandarin and english. *Front. Commun.* 6:639390. doi: 10.3389/fcomm.2021.639390

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership