

BRIDGING READING ALOUD AND SPEECH PRODUCTION

EDITED BY: Simone Sulpizio and Sachiko Kinoshita
PUBLISHED IN: Frontiers in Psychology



frontiers

Frontiers Copyright Statement

© Copyright 2007-2016 Frontiers Media SA. All rights reserved.

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

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

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

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

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

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

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

ISSN 1664-8714

ISBN 978-2-88919-895-5

DOI 10.3389/978-2-88919-895-5

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

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

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

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

What are Frontiers Research Topics?

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

BRIDGING READING ALOUD AND SPEECH PRODUCTION

Topic Editors:

Simone Sulpizio, University of Trento, Italy

Sachiko Kinoshita, Macquarie University, Australia

For decades, human cognition involved in reading aloud and speech production has been investigated extensively (a quote search of the two in google scholar produces about 83,000 and 255,000 results, respectively). This large amount of research has produced quite detailed descriptions of the cognitive mechanisms that allow people to speak or to read aloud a word. However, despite the fact that reading aloud and speech production share some processes – generation of phonology and preparation of a motor speech response – the research in this two areas seems to have taken parallel and independent tracks, with almost no contact between the two.

The present Research Topic takes an initial step towards building a bridge that will link the two research areas, as we believe that such an endeavour is essential for moving forward in our understanding of how the mind/brain processes words. To this aim, we encourage contributions exploring the relation between speech production and reading aloud.

The questions the Research Topic should address include, but are not limited to, the following: To what extent are speech production and word reading/reading aloud similar? Are there some shared components and/or mechanisms between the two processes? Is the time course of the (supposed) shared mechanisms activation similar in the two processes? How does the different input (conceptual vs. orthographic) interact with the types of information that reading and speaking share (semantic and phonological knowledge, articulatory codes)? How does a difference in the input affect the (supposed) common stages of processing (i.e., phonological encoding, and articulatory planning and execution)?

We welcome any kind of contribution (e.g., original research article, review, opinion) that answers the above or other questions related to the Topic.

Citation: Sulpizio, S., Kinoshita, S., eds. (2016). Bridging Reading Aloud and Speech Production. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-895-5

Table of Contents

04 Editorial: Bridging Reading Aloud and Speech Production

Simone Sulpizio and Sachiko Kinoshita

1. Reading aloud as a speech production task

06 The eye-voice span during reading aloud

Jochen Laubrock and Reinhold Kliegl

25 Literacy transforms speech production

Meredith Saletta

29 The segment as the minimal planning unit in speech production and reading aloud: evidence and implications

Alan H. Kawamoto, Qiang Liu and Christopher T. Kello

35 Lexical frequency effects on articulation: a comparison of picture naming and reading aloud

Petroula Mousikou and Kathleen Rastle

44 Incrementality in Planning of Speech During Speaking and Reading Aloud: Evidence from Eye-Tracking

Lesya Y. Ganushchak and Yiya Chen

52 The segment-to-frame association in word reading: early effects of the interaction between segmental and suprasegmental information

Simone Sulpizio and Remo Job

2. The influence of written distractors on speech production

66 Distinguishing Target From Distractor in Stroop, Picture–Word, and Word–Word Interference Tasks

Xenia Schmalz, Barbara Treccani and Claudio Mulatti

71 What can Written-Words Tell us About Lexical Retrieval in Speech Production?

Eduardo Navarrete, Bradford Z. Mahon, Anna Lorenzoni and Francesca Peressotti

83 “When” Does Picture Naming Take Longer Than Word Reading?

Andrea Valente, Svetlana Pinet, F-Xavier Alario and Marina Laganaro

3. Reading and speech production in bilinguals and dyslexia

94 Asymmetric Switch Costs in Numeral Naming and Number Word Reading: Implications for Models of Bilingual Language Production

Michael G. Reynolds, Sophie Schlöffel and Francesca Peressotti

109 The Emergence of a Phoneme-Sized Unit in L2 Speech Production: Evidence from Japanese–English Bilinguals

Mariko Nakayama, Sachiko Kinoshita and Rinus G. Verdonchot

119 A Principled Relation between Reading and Naming in Acquired and Developmental Anomia: Surface Dyslexia Following Impairment in the Phonological Output Lexicon

Aviah Gvion and Naama Friedmann



Editorial: Bridging Reading Aloud and Speech Production

Simone Sulpizio^{1,2*} and Sachiko Kinoshita³

¹ Department of Psychology and Cognitive Science, University of Trento, Trento, Italy, ² Fondazione Marica De Vincenzi ONLUS, Rovereto, Italy, ³ Department of Psychology, Macquarie University, Sydney, NSW, Australia

Keywords: reading aloud, speech production, phonological encoding, lexical access, planning, bilingualism, ERPs, eye-tracking

The Editorial on the Research Topic

Bridging Reading Aloud and Speech Production

The study of how people can speak or read started from the beginning of the modern era of psycholinguistics and neurolinguistics (e.g., Lichteim, 1885; Huey, 1898) and continues to this day. However, the two lines of research—speech production and reading aloud—have followed two separate and parallel paths: While they both concern language production, they seldom meet. Both have produced detailed descriptions of cognitive mechanisms underlying the processes that goes from the message planning to its articulatory realization, but they generally have little contact with each other.

Given their long and fruitful history, this parallel and separate courses of the two research traditions is quite surprising. Everyday experience suggests that speech production and reading aloud do have something in common: In both cases the endpoint is to utter the same linguistic units (saying the word table and reading aloud TABLE will produce the same acoustic material). This suggests that the processes of speech production and reading aloud may not be totally independent: There must be some shared processes at least in terms of generation of phonology and preparation of a motor speech response. The aim of the present research topic is to point the magnifying glass on this issue and to address questions such as the following: To what extent are speech production and word reading/reading aloud similar? Are there some shared components and/or mechanisms between the two process? Is the time course of the (supposed) shared mechanisms activation similar in the two processes? How does the different input (conceptual vs. orthographic) interact with the types of information that reading and speaking share?

Our call has resulted in 12 excellent articles (9 original research, 1 mini review, 1 opinion, 1 perspective article) that provide a first answer to the above questions and provide the impetus for future research. Three articles address the issue of similarity and differences between the processing stages and components of speech production and reading. Valente et al. inspected the spatio-temporal segmentation of ERPs in response to picture naming and word reading and shown that the two tasks are highly similar from 250 ms onward, which is an index of a shared phonological processing stage. Topographic similarities emerged also between 75 and 150 ms, suggesting similar visual processes, although of variable intensity, between the two tasks. Converging evidence for a shared phonological processing comes from Givon and Friedmann who studied patients with phonological output lexicon anomia: They demonstrate that lexical retrieval and reading are tightly linked processes, and suggest a principled relation between dyslexia and anomia. In contrast, Navarrete et al. highlight a difference between reading aloud and speech production: Testing semantic context effects, the authors shown that these effects can be transferred from pictures to words, but not *vice versa*, since the format of the stimuli affects lexical retrieval. As well as contributing to theoretical advancement, the Navarrete et al.'s findings have important

OPEN ACCESS

Edited and reviewed by:

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

*Correspondence:

Simone Sulpizio
simone.sulpizio@unitn.it

Specialty section:

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

Received: 12 April 2016

Accepted: 21 April 2016

Published: 06 May 2016

Citation:

Sulpizio S and Kinoshita S (2016)
Editorial: Bridging Reading Aloud and
Speech Production.
Front. Psychol. 7:661.
doi: 10.3389/fpsyg.2016.00661

methodological implication for the study of lexical access in speech production, when selecting stimulus materials.

Four original research articles targeted the stage of message planning. Speech production and reading aloud are both incremental processes, in which people tend to plan and articulate chunks that are smaller than the whole message they want to utter. Ganushchak and Chen ran two eye-tracking experiments to investigate how the utterance planning is affected by linguistic information (known vs. new information) in reading (sentence reading) vs. speaking (picture description) tasks and showed that planning is more incremental during reading than speaking and that this difference may be ascribed to conceptual preparation. Focusing on words, Mousikou and Rastle investigated participants' response in reading aloud and picture naming to high- and low-frequency stimuli by examining response latencies, and initial-phoneme and whole-word durations. Response latencies were shorter in reading than in picture naming, but initial-phoneme and whole-word durations were longer in the former than in the latter: These findings indicate that reading aloud, but not picture naming, is initiated on the basis of partial information from the printed word, and, that the effect of higher-level cognitive processes influence, to some extent, lower-level articulatory processes. Similarly, Sulpizio and Job offer evidence for a rapid initiation of articulation in reading aloud: By manipulating segmental and suprasegmental information in a series of masked-priming experiments, the authors shown that articulation planning is addressed through a process that starts as soon as the relevant information about the to-be-planned unit (i.e., stress position and phonemes of the stressed syllable) is active. An even extreme position is offered by the Kawamoto et al.'s mini-review, in which the authors advance the intriguing proposal that the minimal planning unit in reading aloud and speech production is the single initial segment.

Two original research articles looked at reading and speaking in bilinguals. Reynolds et al. investigated asymmetric switch costs in unbalanced bilinguals performing production and comprehension tasks: They shown that switch costs are affected by the task participants perform, highlighting that there exist relevant task-related differences in how different languages are controlled. Nakayama et al. investigated the phonological encoding of low- and high-proficient Japanese-English bilinguals,

two languages that differ in the functional phonological unit used in speech production: Their results show that, when processing L2 English, low-proficient bilinguals keep using the phonological unit of their L1, whereas high-proficient bilinguals can use that of the L2. Moreover, their results suggested that it is the length of exposure to L2, rather than Age of Acquisition or proficiency, that led to the adoption of more native-like unit of phonological encoding.

The article by Laubrock and Kliegl is a substantial investigation of the eye-voice span in reading, and its relation with the eye-movement behavior and the response dynamics: As well as offering a promising direction for the understanding of the eye-voice coordination in reading, the study shows that the eye-voice span is directly related to the process of working-memory buffer updating. Finally, the opinion and the perspective article close the topic advancing intriguing theoretical proposal. Schmalz et al. propose a new solution for the issue of lexical selection that can account for performance in tasks as Stroop, picture-word and word-word interference: In their proposal, the conflict is resolved by linking the stimulus perceptual features with the linguistic information, which allows the system to identify which is the target and which is the distractor. Finally, Saletta proposes that speaking and reading are tightly linked since they share mechanisms of processing and learning; she argues that orthographic input exerts positive effects on speech learning and evocatively links these effects to individual's speech movements and motor control.

Overall, we believe that this Research Topic provides a good start to our initial questions. By means of the use of different techniques and the involvement of different populations, the empirical articles highlights similarities and differences between reading aloud and speech production; at the same time, theoretical articles offer new perspectives that guide the direction for future research. Altogether, the articles offer a solid scaffold for the challenge of bridging reading aloud and speech production. We hope that future research will build on the scaffold and pursue this line by filling the remaining gaps that mutually benefit the research on speech production and reading.

AUTHOR CONTRIBUTIONS

SS and SK have made substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

- Huey, E. B. (1898). Preliminary experiments in the physiology and psychology of reading. *Am. J. Psychol.* 9, 575–586. doi: 10.2307/1412192
- Lichtheim, L. (1885). On aphasia. *Brain* 7, 433–484. doi: 10.1093/brain/7.4.433

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

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



The eye-voice span during reading aloud

Jochen Laubrock* and Reinhold Kliegl

Department of Psychology, University of Potsdam, Potsdam, Germany

OPEN ACCESS

Edited by:

Sachiko Kinoshita,
Macquarie University, Australia

Reviewed by:

Manon Wyn Jones,
Bangor University, UK
Aaron Veldre,
University of Sydney, Australia

*Correspondence:

Jochen Laubrock,
Department of Psychology, University
of Potsdam, Karl-Liebknecht-Straße
24-25, 14476 Potsdam, Germany
laubrock@uni-potsdam.de

Specialty section:

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

Received: 02 August 2015

Accepted: 08 September 2015

Published: 24 September 2015

Citation:

Laubrock J and Kliegl R (2015)
The eye-voice span during reading
aloud. *Front. Psychol.* 6:1432.
doi: 10.3389/fpsyg.2015.01432

Although eye movements during reading are modulated by cognitive processing demands, they also reflect visual sampling of the input, and possibly preparation of output for speech or the inner voice. By simultaneously recording eye movements and the voice during reading aloud, we obtained an output measure that constrains the length of time spent on cognitive processing. Here we investigate the dynamics of the eye-voice span (EVS), the distance between eye and voice. We show that the EVS is regulated immediately during fixation of a word by either increasing fixation duration or programming a regressive eye movement against the reading direction. EVS size at the beginning of a fixation was positively correlated with the likelihood of regressions and refixations. Regression probability was further increased if the EVS was still large at the end of a fixation: if adjustment of fixation duration did not sufficiently reduce the EVS during a fixation, then a regression rather than a refixation followed with high probability. We further show that the EVS can help understand cognitive influences on fixation duration during reading: in mixed model analyses, the EVS was a stronger predictor of fixation durations than either word frequency or word length. The EVS modulated the influence of several other predictors on single fixation durations (SFDs). For example, word-N frequency effects were larger with a large EVS, especially when word N-1 frequency was low. Finally, a comparison of SFDs during oral and silent reading showed that reading is governed by similar principles in both reading modes, although EVS maintenance and articulatory processing also cause some differences. In summary, the EVS is regulated by adjusting fixation duration and/or by programming a regressive eye movement when the EVS gets too large. Overall, the EVS appears to be directly related to updating of the working memory buffer during reading.

Keywords: reading, eye movements, eye-voice span, synchronization, working memory updating, psycholinguistics

Introduction

The pattern of fixations and saccades during reading is arguably one of the most practiced and fastest motor activities humans routinely perform. Eye movements during silent reading are clearly affected by cognitive processing. Both low-level visuo-motor factors and high-level comprehension processes co-determine where the eyes land within a word during reading (see Rayner, 1998, 2009, for reviews). Cognitive modulation of oculomotor control has been incorporated in all successful computational models of eye movements during reading, such as SWIFT (Engbert et al., 2002, 2005), EZ-reader (Reichle et al., 1998, 2003), or Glenmore (Reilly and Radach, 2006). However, almost all of the data on which these models are based originates from studies examining silent reading. Here we argue that, by measuring the dynamics between eyes and voice during oral

reading [i.e., differences between the fixated and pronounced words related to processing difficulty at a given point in time; eye-voice span (EVS)], we obtain information about limits of phonological representations of words in working memory (Inhoff et al., 2004), episodic buffer (Baddeley, 2000), or long-term working memory (Ericsson and Kintsch, 1995), available for cognitive processing of the text. Fixation location approximately tells us which input is processed at any point in time, taking into account the fact that the perceptual span during reading has a maximum extent of 10–15 characters to the right of fixation (Rayner, 1975). Articulatory output of a word presumably tells us that it no longer needs to be buffered in working memory. Note that these limits are obtained during a continuous updating of working memory. Indeed, the regulation of the EVS by local processing difficulty may be the most direct measure of limits associated with these constructs. It may also provide additional constraints for computational models of eye-movement control during reading.

Silent reading is a fairly recent cultural invention, at least in the West, where it was introduced only around the 8th century, following the introduction of word spaces (Manguel, 1996). Even though there are reported instances of reading silently, reading aloud was the default in classical antiquity. Similarly, reading aloud precedes silent reading in individual development, for example, in primary school education. Thus, in addition to developing a mental model of the text, a major goal of the reading process is to prepare the words for pronunciation. Indeed, there is evidence that subvocalization takes place even during silent reading and typically occurs during fixation of the subsequent word (Inhoff et al., 2004; Eiter and Inhoff, 2010; Yan et al., 2014a).

Given the importance of oral reading, the lack of data on the coordination of eye and voice during oral reading is surprising. Most of the available data appear to originate with Buswell's (1920, 1922 seminal work using an early eye tracker (see also Tiffin, 1934, for an early approach at simultaneous recording). Buswell (1920) found that the pattern of eye movements during oral reading, just like the pattern during silent reading, consists of forward saccades, regressions, refixations, and word skipplings. More recent research supports the view that eye movements during silent and oral reading are qualitatively similar, although there are also a number of consistently reported quantitative differences. Due to the additional articulatory demands, the average fixation duration is about 50 ms longer in oral reading, the average saccade length is shorter, and there are more regressions (Rayner et al., 2012, p. 92; Inhoff and Radach, 2014). However, the correlation between eye-movement measures obtained during silent and oral reading is high (Anderson and Swanson, 1937). In essence this suggests that oral reading processes may be essentially the same as silent reading processes, but that readers don't want the eyes to go too far ahead of the voice.

Parafoveal processing of upcoming text is important for efficient silent reading (e.g., Sperlich et al., 2015). Interestingly, although parafoveal processing also plays a role in oral reading, the size of the perceptual span is smaller in this mode, possibly related to the overall decrease in saccade size, (Ashby et al., 2012) or the later use of parafoveally extracted information (Inhoff

and Radach, 2014). Thus although more time is available due to the longer fixations in oral reading, apparently this time is not used in the same way for parafoveal preprocessing. Nevertheless, given that parafoveal processing plays a role in silent reading, the spatial region of information extraction and cognitive processing is somewhat larger than the EVS.

Buswell (1920) defined EVS as the distance that the eye is ahead of the voice during reading aloud. He reported the EVS to be on the order of 15 letters (or two to three words) for college students and as increasing over the course of high-school education (Buswell, 1920, Table 1). Buswell also reported that the EVS is sensitive to local processing difficulty, e.g., he found an increased number of regressions (saccades against the reading direction) following a large EVS (see also Fairbanks, 1937). However, he did not have available the rich set of tools that statistics and psycholinguistics provide us with today. These allow us to examine influences of linguistic word properties (e.g., word length, frequency, and predictability) of the currently fixated word or of its neighbors on eye-movement measures of the currently fixated or the currently spoken word. Linear mixed models (LMMs) allow us to evaluate the degree of parallel processing. For example, we can re-evaluate Buswell's hypothesis that the EVS might be responsible for long fixation durations—a hypothesis he could not confirm with his analysis methods (Buswell, 1920, pp. 80–81).

The empirical database on the EVS during reading aloud is very sparse, and most published articles after Buswell used a rather imprecise offline method, that is without recording of eye movements (e.g., Levin and Buckler-Addis, 1979). The offline method works by switching off the light during reading of a sentence and counting how many words can be articulated after the light was off. Obviously, this “off-line EVS” not only includes parafoveal preview and guessing, but may also depend on task-dependent strategies such as looking at the final part of the sentence before starting to read aloud. For these reasons, the offline EVS typically ranges from 6 to 10 words and, to anticipate one of our results, grossly overestimates the EVS measured with eyetracking equipment. Using an eyetracker, Inhoff et al. (2011) determined the temporal EVS, that is the average time the voice trails behind the eyes. They found an average temporal EVS of about 500 ms, which is in good agreement with Buswell (1920), but certainly too short to process 6–10 words, given an average fixation duration of 250 ms. In the most recent study, De Luca et al. (2013) reported a spatial EVS of about 13.8 letters for normal and of 8.4 letters for dyslexic readers.

What does the EVS measure? Although it is possible that synchronization of the eyes with the speed of articulation is attempted for no particular reason, the EVS is more likely related to updating of working memory. During the time between visual input and speech output, the written text is transformed into a phonological code, which is then buffered in the phonological loop (Baddeley and Hitch, 1974). The need for translation into a phonological code arises from the fact that purely visual short-term memory decays very quickly (Sperling, 1960). Buffering is necessary because the articulatory motor system is just too slow to produce understandable speech at the maximum rate of visual decoding and grapheme-to-phoneme conversion. In support of

this view, Pan et al. (2013) found that the EVS in a RAN task correlated with naming speed only when highly familiar and practiced symbols (digits) were named automatically, but not with naming of less well-practiced items with identical articulatory demands (number of dots on a dice). Moreover, dyslexic readers did not exhibit this correlation between EVS and automatic naming of digits, suggesting that a larger EVS is indicative of buffering of material that can be rapidly decoded and translated from graphemic input into a phonological code. Buffering is followed by selection of and commitment to a single phonological code in order to conduct explicit programming for the articulatory response (Jones et al., 2015). Thus dyslexic readers also exhibit a temporal EVS delay on RAN, which is specific to this measure, i.e., no analogous deficit appears in gaze duration (GD; Jones et al., 2013).

Such a first-in-first-out buffer is conceptualized with a finite and rather limited capacity, that is, it cannot sample input infinitely when no output occurs. In general, as we don't appear to use visual short-term memory for buffering of text, most of the buffering during oral reading is probably on the phonological side of the translation, but before the actual articulatory motor processes start. This is compatible with estimates of the inner voice during silent reading: phonological codes appear to be activated for most words we read and this phonological information is held in working memory and is used to comprehend text (Rayner et al., 2012, chap. 7). These phonological codes lag behind the eyes in reading. The phonological buffer in the Baddeley and Hitch (1974) working-memory model has a special capacity for temporal order information. Thus, one important function of phonological codes is to provide access to the order in which words were read.

Synchronized recordings of eye movements and other motor activity are occasionally reported from other domains (Land and Tatler, 2009, for an overview); for example there are several reports of the eye-hand span during piano playing (Truitt et al., 1997; Furneaux and Land, 1999), writing (Almargot et al., 2007), typewriting (Butsch, 1932; Inhoff et al., 1986; Inhoff and Wang, 1992; Inhoff and Gordon, 1998), or performing sports (Land and Furneaux, 1997; Land and McLeod, 2000). One general finding emerging from these studies is that the eye-hand span increases with expertise if measured in units of information (letters or notes), whereas it appears to be fairly constant at around one second if measured in units of time (e.g., Butsch, 1932; Furneaux and Land, 1999). Although these data are only indirectly related to oral reading because of obvious differences in input information and effector system, they are similar in the need to coordinate fast eye movements and a much slower motor system. In particular, working memory buffering is also needed for other forms of output, but may use different codes depending on the output demands.

The aims of the current study are twofold. First, we measure visual sampling of the input and oral output simultaneously to obtain a precise estimate of cognitive processing times during oral reading. These data yield a description of the EVS that allows us to evaluate Buswell's (1920) findings with state-of-the-art equipment. Second, we investigate the dynamics of the EVS during reading aloud with LMMs for statistical inference

and with reference to the possible role of working memory. In perspective, these analyses are to provide constraints for computational models of eye movement control during reading.

Arguably, during silent reading there are well-documented effects of neighboring words on fixation duration (Kliegl et al., 2006; Kliegl, 2007; Wotschack and Kliegl, 2013; but see Rayner et al., 2007). For example, Kliegl et al. (2006) examined the effect of word frequency of the current, past, and upcoming word on current fixation duration during the reading of German sentences. They reported that the negative linear influence of word frequency of the currently fixated word was weaker than that of the word frequency of its left neighbor, indicating that lagged cognitive processing can directly influence saccade programming (see also Rayner and Duffy, 1986). There was also a weak, but significant negative effect of the word frequency of the right neighbor. Moreover, in the same analyses, the predictability of the upcoming word prolonged fixation durations, as indicated by a significant *positive* effect of the predictability of its right neighbor, suggesting that memory retrieval of the right-parafoveal word is attempted when it is likely to be successful. These effects were obtained across nine independent samples of readers (Kliegl, 2007).

Experimental evidence for preprocessing of the parafoveal word to the right also comes from studies using the gaze-contingent display-change paradigm (Rayner, 1975), in which a preview is replaced by a target word during a saccade to the target; preview benefit is the reduction of target fixation duration as a function of the relatedness of the preview relative to a non-word or unrelated preview word. Orthographic and phonological information has long been known to produce preview benefits (e.g., Rayner, 1978; Rayner et al., 1978; Pollatsek et al., 1992; Henderson et al., 1995), and although overall the data are not completely clear (Rayner, 2009), evidence is accumulating that semantic relatedness can also result in preview benefit (Yan et al., 2009; Hohenstein et al., 2010; Laubrock and Hohenstein, 2012; Schotter, 2013).

In summary, during a fixation on a word, processing of the last and of the upcoming word as well as predictive processes are simultaneously ongoing. Given that not only properties of the current word, but also those of its neighbors influence a fixation duration, the question arises to what extent they also affect the EVS. Conversely, how does the EVS affect the where and when of eye movement programming? Having access to an explicit measure of the EVS allows us to answer these questions in detail. The goal of the present work is to present a rich description of the EVS, its relation to eye-movement behavior and to cognitive demands. In perspective, we aim for a novel, on-line characterization of the working memory buffer during actual reading that we hope stimulates and constrains further modeling attempts.

Materials and Methods

Participants

Thirty-two subjects (12 males, 20 females) received 7 € or course credit for participating in an oral experiment

lasting approximately 40 min. Their mean age was 18 years ($SD = 1.5$ years, range = 16–24 years). An additional 31 subjects (12 males, 19 females; mean age 19 years, $SD = 1.4$ years, range = 16–24 years) read the same sentences in a silent reading experiment. All subjects had normal or corrected-to-normal vision. Experiments comply with the June 1964 Declaration of Helsinki (entitled “Ethical Principles for Medical Research Involving Human Subjects”), as last revised, concluded by the World Medical Association. Our eye-tracking research has been approved by Ethikkommission der DGPs (Registriernummer: JKRKRE19092006DGPS).

Apparatus and Material

Sentences were presented on a 22" Iiyama Vision Master Pro 514 CRT monitor with a resolution of 1280×960 pixels controlled by a custom C++ program running on a standard PC. Voice was recorded to hard disk using a Sennheiser K6 series condensator microphone connected to an ASIO compatible SoundBlaster Audigy sound card inside the PC, ensuring a fixed audio latency of 5 ms. Eye movements were registered using the Eyelink 1000 tower mount (SR Research, Ottawa, ON, Canada). The head was stabilized and a viewing distance of 60 cm was assured with a headrest, but the usual additional chinrest was removed to allow for easy articulation. Eye movements and voice protocols were synchronized by sending trigger signals to the eye tracker at the beginning and end of each sound recording, which were recorded in tracker time in the series of eye tracker time stamps and later adjusted for the audio output delay.

The experimental material was the Potsdam Sentence Corpus 2 (PSC2), consisting of 144 simple, declarative German sentences taken from various newspapers (Poltröck, unpublished Diploma thesis). Word length ranged from 2 to 13 letters ($M = 5.26$, $SD = 2.59$ letters), sentence length ranged from 7 to 13 words ($M = 8.54$, $SD = 1.44$) and from 34 to 84 letters ($M = 54.58$, $SD = 10.67$). Word frequency information for the 1230 words was obtained from the DWDS/dlexdb corpus (Heister et al., 2011) based on ca. 120 Million entries. Median word frequency was 234.2 per Million, and the range was from 0.008 to 26530 per Million (for “Geplänkels” and “der”, respectively). Incremental cloze predictabilities were collected from different 283 participants generating more than 85,000 predictions (mean N of predictions per word 69.6, range from 57 to 84) using an internet-based questionnaire, combined with an ipod lottery to increase motivation. The mean predictability over words in the corpus was 0.188, and the median predictability was 0.042; about 1/3 of all words were completely unpredictable. As usual in single-sentence material, predictability in the PSC2 increases with position of word in the sentence (e.g., mean predictability of 0.063 and 0.435 for sentence-initial and sentence-final words, respectively).

Procedure

The 144 experimental sentences were read in random order after six initial training sentences used to familiarize the participants with the task and to adjust the volume/gain setting of the microphone. One sentence was presented per trial, vertically centered on the screen, in black on a white background, using

a fixed-width Courier New font with a font size of 24 points. A letter subtended 14 pixels or 0.45° of visual angle horizontally. A trial started with a drift correction in the screen center (standard drift correction target), followed by presentation of a gaze-contingent sentence trigger target 18.1° to the left of the screen center, followed by presentation of the sentence. The sentence was only revealed after the gaze-contingent trigger had been fixated for at least 50 ms. Visual properties of the sentence trigger target were identical to those of the drift correction target. Sentences were aligned with the center of the first word positioned slightly to the right of the sentence trigger target; so that the gaze was initially positioned at the first word's optimal viewing position. Sentence presentation ended when subjects fixated a point in the lower right screen corner. To ensure that subjects read the sentences and not just moved their eyes, a randomly determined third of sentences were followed by an easy comprehension question, requiring a three-alternative choice response.

The eye tracker was calibrated at the beginning of the experiment and after every 36th trial or whenever calibration was bad. Bad calibrations were detected at the beginning of each trial: when the gaze was not detected within an area of 1° centered on the sentence trigger target within 1 s from the start of its presentation, a re-calibration was automatically scheduled. A trial ended when subjects fixated another gaze-contingent trigger (150×150 pixels square) in the bottom right corner of the screen for at least 50 ms, which was visually represented by a 5×5 pixel in its center.

Data Analysis

Eye Movement Recordings

The horizontal position of the gaze was mapped to letter positions, and standard measures were determined such as first-fixation duration (FFD; duration of the first fixation on a word in firstpass reading), single fixation duration (SFD; duration of fixations on words that received exactly one first-pass fixation), GD (sum of all first-pass fixations) as well as skipping refixation, regression, and single-fixation probabilities. Trials with eye blinks were removed from the analysis. Also data from the first and last words of each sentence were not included in the analysis.

Voice Recordings

A Praat (Boersma, 2001; Boersma and Weenink, 2010) script was prepared that looped over subjects and sentences and presented each sentence (divided into words) together with its associated sound recording, showing a representation of the waveform together with a spectrogram, formants, and intensity and pitch contours. The script attempted to locate the beginning and end of spoken parts by crossings of an intensity threshold, and initially distributed word boundaries across the spoken part in proportion to word length. Human scorers then manually dragged word boundaries to the subjective real boundary locations by repeatedly listening to stretches of the speech signal. Several zoom levels were available, and scorers were instructed to zoom in so far that only the word in question and its immediate neighbors were visible (and audible) for the ultimate adjustment. In the case of ambiguous boundaries due to co-articulation,

scorers were instructed to locate the boundary in the middle of such ambiguous stretches¹. Only articulated word durations from sentences that were read without error were used in further analyses.

Eye-Voice Span

The 86% of sentences (3938 out of 4608) with correct articulation and without eye blinks were used in analyses of the EVS. The EVS can be defined in either temporal or spatial units, or either relative to the fixated or the articulated word. As temporal measures, we calculated the time difference in milliseconds to articulation onset at the beginning of the first fixation on a word (termed *onset-EVS* below) and at the end of the last fixation on a word (*offset-EVS*). As spatial measures, we calculated the distance in letters of the currently articulated letter relative to each fixation onset and offset.

Labeling word boundaries in the auditory signal is like sampling the signal only at word boundaries. However, the eye and voice are to a certain degree independent of each other, that is fixations usually start during the pronunciation of a word. In an attempt to increase the precision of the position of the voice at fixation onset, we made use of the very high linear correlation between articulated word times and word length in German ($r = 0.86$ in the present data). Specifically, we linearly interpolated letters by assuming that the per-letter duration is given by the word's articulated duration divided by its number of letters to estimate the proportion of a word that was spoken at fixation onset. For most analyses reported below, the spatial distance in letters at first-fixation onset or offset will be used.

(Generalized) Linear Mixed Models

Analyses were performed with the R statistical computing environment (R Development Core Team, 2015) and the packages lme4 (Bates et al., 2015b) and remef (Hohenstein and Kliegl, 2015), using a LMM approach that allows to investigate experimental effects with statistical control due to differences between subjects and sentences as random factors (Bates et al., 2015a). We used two GLMMs and two LMMs. With the two GLMMs we modeled regressive and refixation saccades as a function of either onset EVS and the change in EVS (from onset-EVS to offset-EVS) during a fixation using the logit link, with statistical control for differences between participants and sentences. With the two LMMs we modeled SFDs; to achieve normally distributed residuals, SFDs were log transformed. Both models used the covariates reported in Kliegl et al. (2006) with nine word and three oculomotor variables as a starting point (see Results for details). These covariates are not necessarily in a strict linear relation with the dependent variable. Therefore, to guard against overlooking an important non-linear contribution, we modeled these covariates with quadratic polynomials, except frequency of the fixated word for which we specified a cubic trend (see Heister et al., 2012). To the first LMM, we added EVS (a linear within-subject covariate) and its interactions with all the other covariates as additional fixed effects. Analogously,

we added reading condition (oral vs. silent; a between-subject fixed factor) and its interactions with the other covariates. Thus, the two LMMs were of equal complexity. Moreover, for all models we determined significant variance components for experimental effects and associated correlation parameters. In principle, there is no upper limit to model complexity with 12 quadratic (or higher-order) covariates. Therefore, we built the LMM with the constraint that the model was not overparameterized, following recommendations and procedures in Hohenstein and Kliegl (2014) and Bates et al. (2015b). Data, scripts, and results of all analyses are available as a supplement at Rpubs.com.

Results

General descriptive statistics relating to eye movements and articulation during oral reading are summarized in **Table 1**. For comparison we include also eye movement data from a new sample of 31 readers who read the same material silently. The comprehension questions were accurately answered in both reading modes, with mean accuracies of 97.7% (range 94–100%) for oral and 97.4% (range 94–100%) for silent reading. Fixation durations were longer and saccades were shorter during oral than during silent reading. The probability of refixating a word was higher, whereas the probabilities of word skipping and of regressions were lower in oral than in silent reading. The average spoken word duration in oral reading was similar to the average GD. Notably, the time till pronunciation of the first word was about the duration of three spoken words, suggesting that the eye initially gets a head start before articulation of the sentence starts.

In the following we focus on the dynamic relation of eye and voice. The presentation of results is organized as follows. In the first sections the focus is on active control of EVS by regression, refixation, and fixation durations. The final section informs about whether previously reported effects of distributed processing of words in the perceptual span on fixations during silent reading are also observed during oral reading.

TABLE 1 | Descriptive statistics for oral and silent reading.

	Oral		Silent	
	Mean	SD	Mean	SD
Fixation duration [ms]	253	96	209	81
First fixation duration [ms]	262	96	213	81
Single fixation duration [ms]	273	99	216	82
Gaze duration [ms]	334	162	247	124
Total viewing time [ms]	362	187	288	173
Saccade length [letters]	5.9	2.6	7.0	3.1
Skipping probability	0.14	0.35	0.21	0.41
Single fixation probability	0.51	0.50	0.59	0.49
Refixation probability	0.18	0.38	0.11	0.31
Regression probability	0.06	0.23	0.10	0.30
Time to first-word pronunciation [ms]	877	191		
Spoken word duration [ms]	293	150		

¹Even with this computer-assisted procedure, scoring of word boundaries was rather laborious.

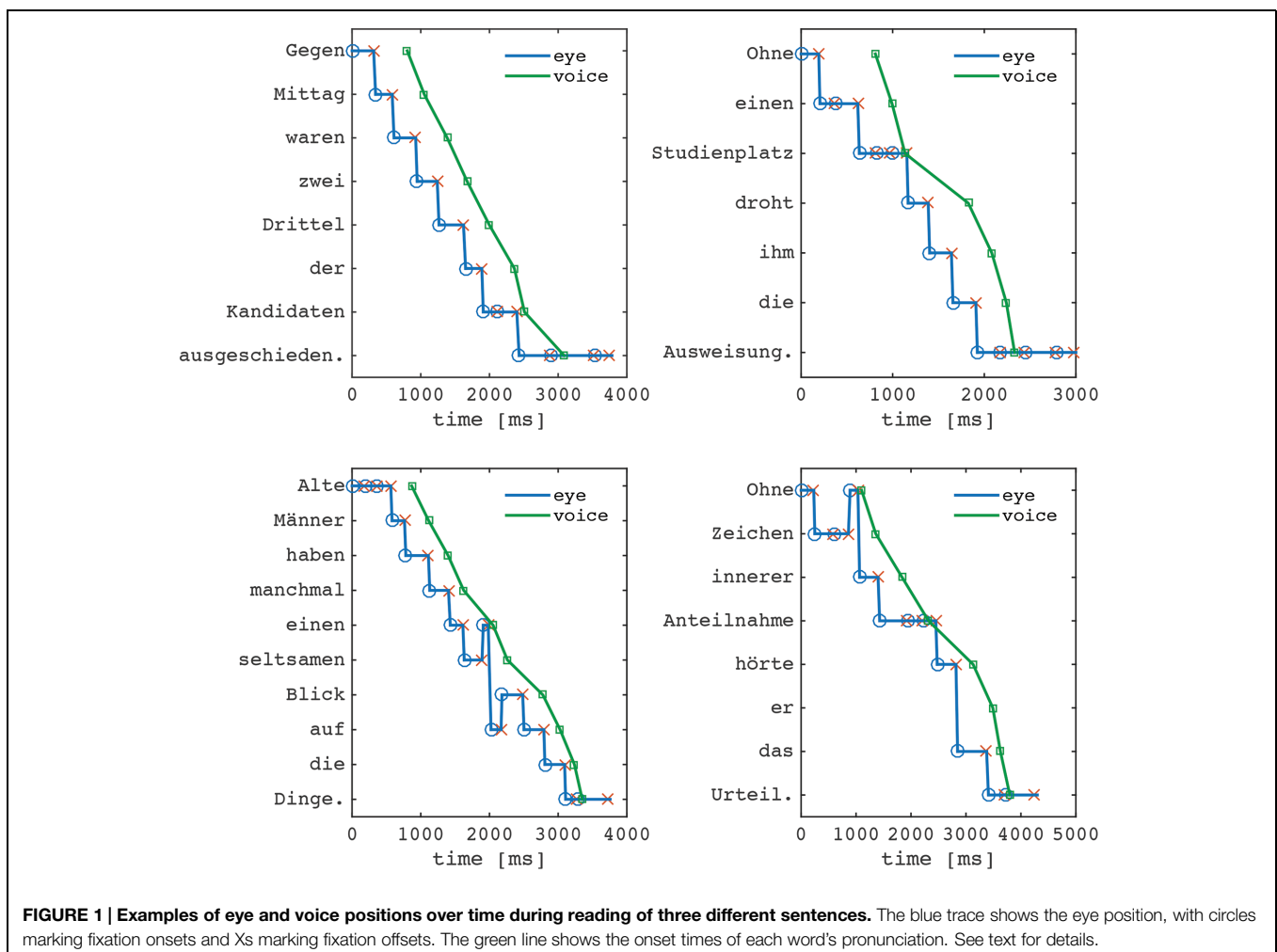
Eye-Voice Spans

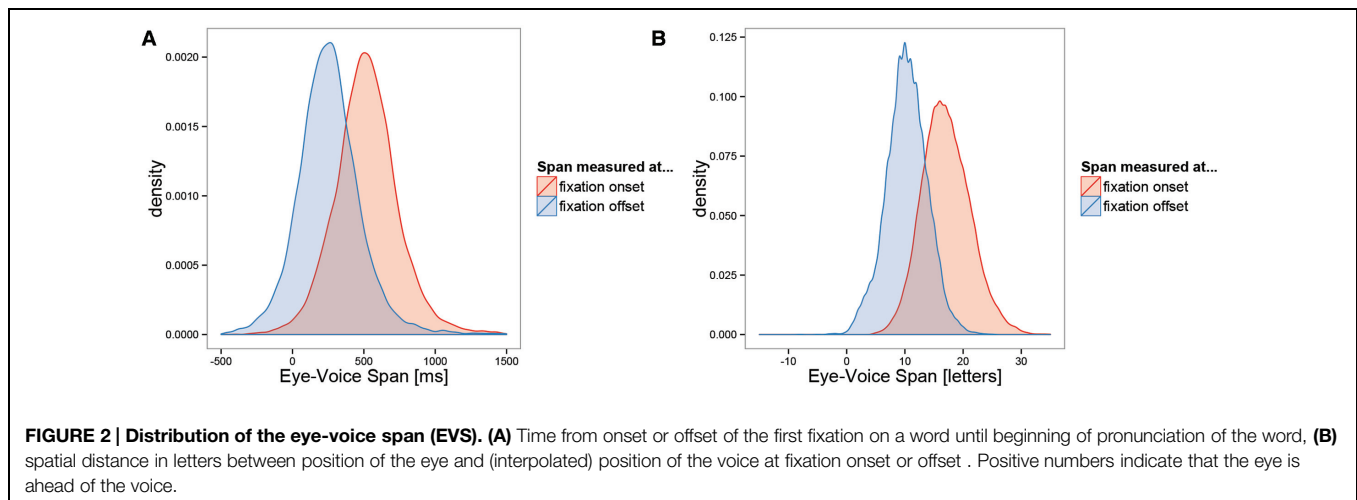
The signature marker of oral reading is the EVS, which can be measured with respect to the temporal or the spatial distance between eye and voice. We illustrate these concepts with three examples. Each panel in **Figure 1** shows the traces of the eye (blue line) and the voice (green line) over time during the reading of a sentence. In the top left panel the eye leads the voice by a fairly constant time or distance throughout the sentence. In the top right panel, the EVS all but vanishes during refixations of the word “Studienplatz.” In the bottom left panel, the eye regresses back twice to previous words to wait for the voice to catch up, followed by the eye jumping ahead of the voice again to ensure a distance similar to the one before the regression. Arguably, the latter two cases represent prototypes of how eye and voice take care of a local disturbance. Often this is due to a particularly difficult word, like in the refixations example where, in a way, the difficult word serves as a point of synchronization. The determiner “einen,” on the other hand, is unlikely to cause processing difficulties in normal reading, possibly the function of the regression is to reduce the distance between eye and voice. In the bottom right panel, finally, regressions and refixations are displayed, and a

particular pattern appears at the beginning of the sentence, where the eye initially scouts ahead, and makes a regression to the beginning word just before the voice starts pronouncing it. This sentence-initial pattern that looks like an initial re-synchronization to maintain a manageable buffer size was quite typical.

Temporal EVS

The temporal EVS distributions are displayed in the left panel of **Figure 2**. The distribution of the EVS in milliseconds from the beginning of the first fixation on a word to the onset of its pronunciation was nearly symmetric, with a mean of 561 ms and a standard deviation of 230 ms (**Figure 2**, right distribution in left panel). In contrast to most other measures during reading, the interindividual variability in temporal EVS ($SD = 73$ ms) was smaller than the intraindividual variability ($SD = 218$ ms). The mean EVS per subject ranged from 428 to 781 ms in our sample. Obviously, during oral-reading fixations the voice is able to catch up with the eyes. Consequently, the temporal EVS from the end of the last fixation on a word to the onset of its pronunciation was much shorter with a mean of 254 ms and a standard deviation of 216 ms (**Figure 2**, left distribution in left panel). The





standard deviations of the onset and offset distributions were not significantly different; Levene's test, $F = 2.66$, $p = 0.103$.

Spatial EVS

The spatial EVS distributions are displayed in the right panel of **Figure 2**. The distance in letters between the position of the eye and the position of the voice was estimated at each fixation onset after articulation of the words had started. Like the temporal EVS distribution, the spatial EVS distribution was nearly symmetric and showed considerable variability. The distribution at first fixation onset had a mean of 16.2 letters ($SD = 5.2$ letters). The interindividual variability ($SD = 1.5$ letters) was smaller than the intraindividual variability ($SD = 4.9$ letters). At last fixation offset, the eye was still 9.7 letters ahead of the voice ($SD = 3.6$ letters). Thus, during a fixation the spatial EVS was reduced on average by 6.5 letters (which is very close to the average saccade size); moreover, this reduction in spatial EVS went along with a significant reduction of its standard deviation; Levene's test, $F = 797$, $p < 0.001$. We interpret these results as evidence for active control of spatial rather than temporal EVS.

Eye-Voice Span as Predictor of Eye-Movement Control

A dominant goal of oral reading is to maintain a steady pace, modulated only for various prosodic effects. The observation that fixation durations are locally adjusted to keep the EVS at fixation offset at a fairly constant level of about 10 characters reflects this regulation. In this section we analyze by which means active control of spatial EVS is achieved. Specifically we show that at a given point in time the EVS is predictive of (1) regressions, (2) refixations, and (3) fixation durations that are followed by a forward saccade. Note that with this definition we analyze three non-overlapping sets of fixations and their associated EVSs from reading the same sentences.

Spatial EVS Predicts Regression and Refixation Probabilities

Moving beyond anecdotal evidence and descriptives, we demonstrate regulation with analyses of regression and refixation

probabilities as a function of EVS at the beginning and at the end of a fixation. Effects were tested with two GLMMs using the logit link function to predict binomial responses (either refixations or regressions) with EVS at onset and the difference between onset-EVS and offset-EVS as predictors, including both linear and quadratic trends.

The left panel of **Figure 3** shows the key results for regression and refixation probabilities as a function of the EVS at fixation onset. Both probabilities increased with an increase in EVS, suggesting that it is often determined already at the onset of a fixation whether a halt or a regressive eye movement will be programmed. **Table 2** shows that for both refixations and regressions, there were purely linear effects on the logit scale, indicating that the odds of making a regression or refixation increase with every character increase in the onset-EVS.

The right panel of **Figure 3** shows that the correlation between the offset-EVS and regression and refixations probabilities was considerably stronger at fixation offset than at fixation onset. This is captured by a significant coefficient for the Δ EVS-effects in **Table 2**. For both regressions and refixations, there was a strong increase in the linear effects. Additionally, there was a negative quadratic trend for refixations, meaning that when offset-EVS was very large, the likelihood of refixating increased no further; so that when offset-EVS was large, the probability of making a regression exceeded the refixations probability (the apparent positive quadratic trend for regressions was linear on the logit scale, indicating that with every character increase in the EVS, there is a proportional increase in the odds of making a regression). The fact that offset-EVS is more strongly related to regression behavior than onset-EVS suggests that the control of fixation durations is sometimes successful in decreasing the EVS.

In summary, the EVS is regulated by programming a refixation or a regression when the EVS gets too large. Whether a refixation or a regression is programmed is related to the size of the EVS at fixation offset: the likelihood of making a regression strongly increases with every additional character of EVS, whereas the likelihood of making a refixation initially increases, but then drops again for large EVS, for which regressions are the rule. The increase in regression or refixations probabilities with offset-EVS

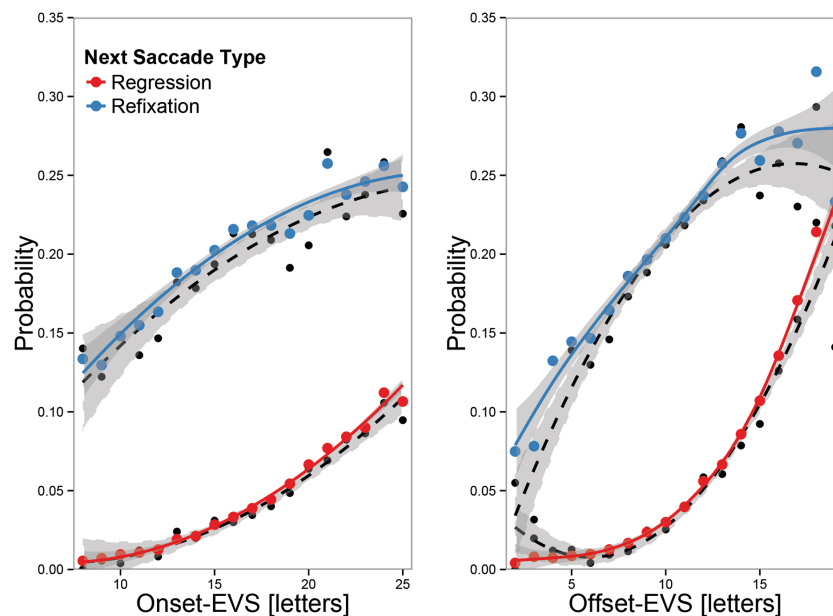


FIGURE 3 | Regression and refixation probabilities as function of EVS at fixation onset (left) and offset (right). Black dots represent overall means, and colored dots predicted means, adjusted for random effects. The lines represent second-order polynomial regression fits (black dotted) or GLMM fits (colored, solid). EVS at fixation onset is already predictive of an upcoming regression or refixations, but offset-EVS is more predictive. When EVS was large at offset, there was a high likelihood of making a regression.

was larger than with onset-EVS. Taken together, this suggests that regressions or refixations are programmed when the control of fixation duration is not sufficient in down-regulating the EVS.

Spatial Onset EVS Predicts Fixation Durations

Main effect of EVS

The analyses in the last section demonstrated that EVS at the end of a fixation (offset EVS) is strongly predictive of regressive and refixation saccades. In this section, we test whether fixation durations that are followed by a forward saccade are influenced by onset EVS. On the assumption that not only eye movements (i.e., regressions and refixations), but also fixation durations are in the service of maintaining fluent speech, the spatial EVS at fixation onset, should be predictive of the subsequent fixation duration. Specifically, the expectation is that if the EVS at fixation onset is large, long fixations should follow. There was clear evidence for this hypothesis in the data (see top left panel in **Figure 4**). The partial effect of onset-EVS on SFD (i.e., the regression line) represents a good fit of the observed mean SFDs at the various EVS levels (i.e., the dots). EVS at fixation onset was one of the strongest predictors of SFD, and had a substantial linear influence that was larger than well-established effects such as word frequency or word predictability.

The partial effect of EVS was estimated with statistical control of (a) the other covariates listed in **Table 3**, (b) differences between subject-related and sentence-related differences in mean fixation duration and effects, (c) subject-related and sentence-related differences in five effects each (i.e., variance components for N-length, N-frequency, N-predictability, N-1-length, and N-1-frequency effects, listed in **Table 4**), and (d) correlations

between subject-related (-0.43) and sentence-related ($+0.80$) effects of length and frequency (i.e., correlation parameters). Estimates, standard errors, and t -values are reported in **Table 3**. We describe effects as significant if t -values are larger than 2.0. This is a conservative criterion because, given our past research, all statistical inference is one-tailed.

The main EVS effect was moderated by (interacted with) length of the next word N+1 (i.e., N+1-length), N-frequency, N-predictability, and N-1-predictability. In addition, there were two three-covariate interactions: EVS \times N-frequency \times N-1-frequency and EVS \times N-1-length \times launch distance (see **Table 2**). These interactions are shown in the remaining panels of **Figure 4**.

EVS \times N+1-length

An effect of the length of the next word is obtained for short EVS. Presumably, with short EVS weight of processing can shift in the direction of reading, increasing the chances of observing a parafoveal-on-foveal effect of word length.

EVS \times N-1-predictability

If the last word was of low predictability the EVS slope was steeper than when the last word was highly predictable. High processing difficulty appears to be associated with stronger EVS effects.

EVS \times N predictability

An effect of the predictability of the fixated word is obtained for short onset-EVS, but not for long onset-EVS. This suggests that if the voice lags far behind the eye at fixation onset, prediction of the fixated word is limited. It can possibly be interpreted as a working memory effect; if the working memory buffer is too full, prediction of the upcoming word becomes very hard.

TABLE 2 | Estimates of GLMMs for regressions (upper part) and refixations (lower part) as a function of the Eye-Voice-Span.

Regressions					
Fixed effects					
	Estimate	SE	z value	p	Sig
(Intercept)	−3.82	0.16	−24.36	<0.001	***
Onset-EVS, linear	146.60	8.25	17.78	<0.001	***
Onset-EVS, quadratic	5.89	6.29	0.94	0.35	
ΔEVS, linear	97.68	7.06	13.83	<0.001	***
ΔEVS, quadratic	6.47	6.23	1.04	0.3	
Random effects					
Groups name	Variance	SD			
Sn	(Intercept)	0.31	0.56		
Id	(Intercept)	0.58	0.76		
Number of obs: 16451, groups: sn, 144; id, 32					
Refixations					
Fixed effects					
	Estimate	SE	z value	p	
(Intercept)	−1.59	0.10	−16.64	<0.001	***
Onset-EVS, linear	60.44	3.38	17.89	<0.001	***
Onset-EVS, quadratic	−2.39	2.83	−0.85	0.40	
ΔEVS, linear	49.86	3.32	15.02	<0.001	***
ΔEVS, quadratic	−10.92	3.04	−3.59	<0.001	***
Random effects					
Groups name	Variance	SD			
Sn	(Intercept)	0.28	0.53		
Id	(Intercept)	0.21	0.46		
Number of obs: 16451, groups: sn, 144; id, 32					

ΔEVS indicates the effect of the difference of offset-EVS minus onset-EVS.

EVS × N-frequency × N−1-frequency

The third row of **Figure 4** displays the interaction between current and last-word frequency for small and large EVS. This interaction also subsumes the EVS × N−1-frequency interaction. The most striking feature is the high-N-frequency hump after high frequency words N−1. This two-way interaction (also in its direction) was already reported in Kliegl et al. (2006; also Kliegl, 2007). The most plausible interpretation is that it reflects processing of word N+1 during a fixation on word N. We suggest that the attenuation of the high-frequency hump when word N−1 was of low frequency is evidence for less parafoveal processing during these fixations, presumably due to needs to deal with spillover from the last word. Qualitatively, this interaction was similar for short and large EVS. With a focus on differences, frequency effects were larger and more linear when the EVS was large. EVS moderated the frequency effect on fixation durations even more strongly when word N−1 was of low frequency; a strong and more or less linear N-frequency effect was observed in this case when EVS was large, whereas the N-frequency effect had little time to unfold when EVS was small. Thus when the onset-EVS is large, more cognitive resources seem to be allocated to processing of the current word rather than its neighbors.

EVS × N−1-length × launch site

The fourth row of **Figure 4** displays the interaction between launch site and length of word N−1 for small and large EVS. Fixation durations are especially long for the combination of large launch site and short words. Presumably the major source of this interaction is skipping which, on the one hand, is strongly linked to short words and, on the other hand, it is commonly accepted that fixations after skipped words are longer than average (e.g., Kliegl and Engbert, 2005, Table 1 for a review). Again, this interaction was qualitatively similar for short and large EVS. In this case, the effect of EVS for short last words was larger for long launch sites (i.e., high skipping probability).

Distributed Processing during Oral Reading

Fixation durations are not only predicted by the EVS, but also sensitive to numerous visual and lexical indicators of processing difficulty as well as to oculomotor demands. All the covariates listed in **Table 3** were used in previous research on silent reading and almost all of them showed consistent effects across nine samples of readers (e.g., Kliegl et al., 2006; Kliegl, 2007). In the previous section we used these variables as statistical control variables for assessing the effect

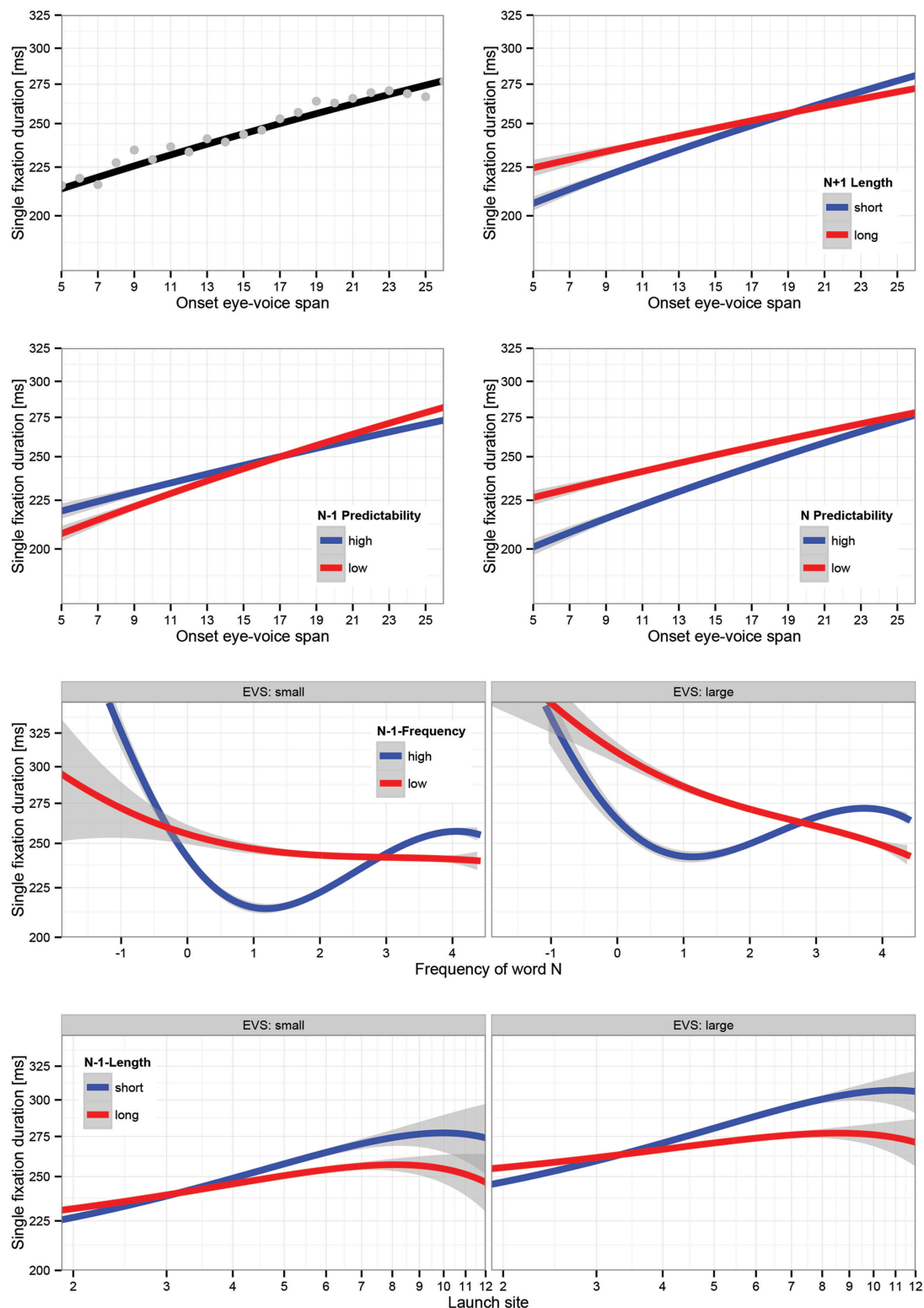


FIGURE 4 | Visualization of LMM estimates of main effect of onset EVS and three EVS-related interactions; LMM used three continuous covariates. Top left: main effect of EVS; dots are observed mean SFDs at levels of EVS; **top right:** EVS \times N+1 length interaction; **second row, left:** EVS \times N-1 predictability; **second row, right:** EVS \times N predictability; **third row:** EVS \times N-frequency \times N-1-frequency; **bottom row:** EVS \times launch site \times N-1 length. Factors in panels are based on median splits for visualization; LMM estimation used continuous covariates. Error bands represent 95% confidence intervals based on LMM residuals. Effects are plotted on a log-scale of fixation durations, thus they show the backtransformed effects as they were estimated in the LMM.

TABLE 3 | Fixed-effect estimates of LMM for single fixation durations (SFDs), including EVS as covariate.

		Estimate	SE	t-value			Estimate	SE	t-value
Grand Mean SFD		5.492	0.018	301.38	Main effect of EVS		0.012	0.001	14.68
N-1 length	Linear	−0.170	0.988	−0.17	EVS × N-1 length	−0.021	0.138	−0.15	
	Quadratic	0.771	0.478	−0.61		0.232	0.088	2.65	
N-1 frequency	Linear	−0.689	0.868	−0.79	EVS × N-1 frequency	0.115	0.131	0.88	
	Quadratic	−0.969	0.500	−1.94		0.050	0.078	0.64	
N-1 predictability	Linear	0.559	0.440	1.27	EVS × N-1 predictability	−0.246	0.083	−2.97	
	Quadratic	0.416	0.449	0.93		0.121	0.080	1.52	
N length	Linear	5.006	1.081	4.63	EVS × N length	−0.204	0.133	−1.54	
	Quadratic	0.342	0.441	0.78		0.070	0.078	0.90	
N frequency	Linear	−0.138	1.214	−0.11	EVS × N frequency	−0.656	0.138	−4.74	
	Quadratic	2.296	0.553	4.15		−0.159	0.096	−1.65	
	Cubic	−2.668	0.500	−5.34		−0.025	0.091	−0.28	
N predictability	Linear	−2.096	0.734	−2.86	EVS × N predictability	0.217	0.086	2.53	
	Quadratic	1.487	0.470	3.16		−0.142	0.078	−1.82	
N+1 length	Linear	0.166	0.708	0.24	EVS × N+1 length	−0.312	0.132	−2.36	
	Quadratic	−2.012	0.457	−4.40		0.064	0.081	0.80	
N+1 frequency	Linear	−2.350	0.738	−3.18	EVS × N+1 frequency	−0.167	0.138	−1.22	
	Quadratic	0.409	0.461	0.89		−0.045	0.085	−0.53	
N+1 predictability	Linear	1.055	0.466	2.27	EVS × N+1 predictability	−0.113	0.080	−1.42	
	Quadratic	0.843	0.441	1.91		−0.020	0.075	−0.27	
launch site distance	Linear	6.525	0.388	16.80	EVS × launch site distance	−0.152	0.083	−1.83	
	Quadratic	1.099	0.306	3.59		0.037	0.067	0.55	
landing site	Linear	6.137	0.370	16.60	EVS × landing site	0.015	0.083	0.18	
	Quadratic	−0.423	0.309	−1.37		−0.024	0.072	−0.34	
saccade size	Linear	7.542	0.363	20.78	EVS × saccade size	0.155	0.080	1.93	
	Quadratic	1.971	0.313	6.31		0.090	0.067	1.35	
N- freq × N-1 freq	Linear	3.258	0.325	10.02	EVS × N- frequency × N-1 freq	0.116	0.057	2.05	
	Quadratic	1.227	0.329	3.73		−0.048	0.055	−0.88	
	Cubic	−1.252	0.318	−3.94		−0.013	0.057	−0.23	
N-freq × N+1 freq	Linear	0.880	0.325	2.71	EVS × N-frequency × N+1 frequency	−0.094	0.061	−1.54	
	Quadratic	0.304	0.358	0.85		−0.105	0.064	−1.64	
	Cubic	0.055	0.380	0.14		0.035	0.068	0.52	
N-1 length x launch site distance		−0.076	0.012	−6.14	EVS × N-1 length x launch site distance		−0.007	0.003	−2.29

Eye-voice span was specified as a centered covariate. Therefore, the intercept estimates the Grand Mean SFD. Main effects of covariates (and associated test statistics) are presented in the left four columns; coefficients for their interactions with EVS in the right four columns; see text for details. Bold values indicate significant contrasts.

of onset EVS on fixation duration. In this section, we assess their effects on their own right, so to say, by comparing them directly with a group of readers who read the same sentences silently. With one exception, this second LMM was identical to the first LMM reported above. Instead of the within-subject covariate EVS, we included the between-subject variable oral vs. silent reading. Estimates, standard errors, and *t*-values for the second LMM are reported in **Table 5**; estimates of variance components are listed in **Table 4**. Again, we describe effects as significant if *t*-values are larger than 2.0. Please note that, as this is an article about the EVS, there is not enough space to discuss in detail effects that relate to other domains of research on eye-movement control during reading. Therefore, this section will be selective in highlighting results that are likely to be of interest beyond the EVS context of eye-movement control during reading.

Canonical Effects

Effects of word length, frequency, and predictability of the fixated word, corresponding effects of its left and right neighbor as well as effects of launch site, fixation position within word, and the amplitude of the outgoing saccade count among the best-studied covariates for single-fixation duration during silent reading. **Figure 5** is modeled on Figure 3 of Kliegl et al. (2006), but displays partial effects both for silent (red lines) and oral (blue lines) reading (i.e., the interaction of reading condition with each covariate). In addition, the gray lines and gray dots in each panel inform about the zero-order (i.e., simple) regression of SFD on the covariates and observed means categorized according to some covariate-dependent binning. Those panels in which the red and blue lines depart substantially from their gray-line neighbors were much affected by statistical control.

TABLE 4 | Variance components and correlation parameters for LMMs.

EVS LMM			Oral/Silent LMM		
Random factor	Variance component	SD	Random factor	Variance component	SD
Sentence (<i>N</i> = 144)	N-length	0.177	Sentence (<i>N</i> = 144)	N-length	0.185
	N-frequency	0.052		N-frequency	0.049
	N-predictability	0.171		N-predictability	0.192
	N-1 frequency	0.024		N-1 frequency	0.032
	N-1 length	0.092		N-1 length	0.105
	Mean SFD	0.062		Mean SFD	0.045
Subject (<i>N</i> = 32)	N-length	0.098	Subject (<i>N</i> = 63)	N-length	0.081
	N-frequency	0.029		N-frequency	0.023
	N-predictability	0.055		N-predictability	0.041
	N-1 frequency	0.012		N-1 frequency	0.010
	N-1 length	0.078		N-1 length	0.075
	Mean SFD	0.096		Mean SFD	0.114
Residual (<i>N</i> = 11709)		0.272	Residual (<i>N</i> = 31185)		

Correlation parameters for were 0.80 and -0.40 for sentence-related and subject-related N-length and N-frequency effects, respectively, in the EVS LMM; corresponding correlation parameters were 0.82 and -0.55 in the oral/silent LMM.

Obviously, aside from the generally longer fixation durations during oral than silent reading, there is much similarity with respect to the direction and profile of the canonical effects. In general, fixation durations increased when processing was difficult. The direction and shape of well-established effects of word length, frequency, and predictability were similar in oral and in silent reading. However, there were also some differences between reading modes, which we will discuss further below.

Controversial and Novel Effects

Aside from corroboration of well-established effects, the data also provided new information on controversial effects. An in-depth discussion of each topic is beyond the scope of this article. Moreover, the results attest to the reliability of effects, but do not really lead to resolution of the associated theoretical controversies. Therefore, the report of these results is to serve primarily as a pointer to the relevant literature. All effects are shown in panels of **Figure 5**.

N+1-frequency and N+1-predictability

There were two controversial effects that were replicated quite strongly in both oral and silent reading: negative N+1-frequency effect and positive N+1 predictability effect. The direction of the former effect is canonical (i.e., shorter fixation durations for high N+1 frequency words) whereas the direction the latter is non-canonical (i.e., longer fixation durations for high N+1 predictability words). The opposite direction of effects on fixation duration is remarkable, given that frequency and predictability of words are positively correlated. Both effects were reported in Kliegl et al. (2006), but are not well understood, and evidence has primarily been obtained from corpus studies (Kennedy and Pynte, 2005; Kliegl, 2007; Rayner et al., 2007; Angele et al., 2015). Their appearance during oral reading strongly supports their reliability and may provide new perspectives on their explanation. First note that there is

no statistical difference between oral and silent reading with respect to the negative N+1 frequency effect. Thus, this effect replicates across reading modes and with new sentence material. It likely indicates parafoveal preprocessing of the upcoming words. Second, the non-canonical positive N+1 predictability effect has been interpreted as an effect of memory retrieval (i.e., not as a parafoveal-on-foveal effect; Kliegl et al., 2006). Again the effect replicated across reading modes, although it also interacted with reading mode, as will be discussed below.

Fixation position

The signature effect of fixation position in word on SFD is the inverted u-shape of the function (Vitu et al., 2001). Again, several explanations have been advanced for this result (Nuthmann et al., 2005, for a review), including fast correction of mislocated fixations near the word boundaries. Our results reveal an important difference between the zero-order relation and the partial effects. The zero-order functions reveal a peak of SFDs in the word whereas for partial effects SFDs increase across the word. Note that all curves are of negative quadratic shape. The divergence between zero-order and partial effects suggests that the commonly observed decrease of SFDs toward the end of words is accounted for by covariates in the LMM. Most importantly, the result was obtained for the group of oral and the group of silent readers, despite minor differences, as will be discussed below.

N–1 frequency

The second example of a strong and quite unexpected difference between zero-order and partial effects concerns the effect of the frequency of the last word. The zero-order functions exhibit the negative effect known from past research (e.g., Kliegl et al., 2006) for both oral and silent reading. Usually this pattern is interpreted as evidence for spillover from processing the previous word. In this case, the partial effects for the reading condition \times N–1-frequency interactions are actually quite misleading and should

TABLE 5 | Fixed-effect estimates of LMM for SFDs, comparing silent and oral reading.

		Estimate	SE	t-value			Estimate SE t-value
Mean oral SFD		5.514	0.021	266.02	Δ (s – o) SFD)		–0.221 0.029 –7.62
N-1 length	Linear	–0.812	1.365	–0.60	Δ (s – o) N-1-length		5.565 1.622 3.43
	Quadratic	1.874	0.599	3.13			–0.613 0.648 –0.95
N-1 frequency	Linear	–0.340	1.177	–0.29	Δ (s – o)-N-1 frequency		3.613 1.169 3.09
	Quadratic	–2.683	0.627	–4.28			2.537 0.645 3.94
N-1 predictability	Linear	0.601	0.554	1.09	Δ (s – o)-N-1 predictor		–0.767 0.649 –1.18
	Quadratic	0.388	0.556	0.70			0.083 0.628 0.13
N length	Linear	7.583	1.564	4.85	Δ (s – o)-N length		–4.165 1.597 –2.61
	Quadratic	–0.190	0.613	–0.31			2.011 0.672 2.99
N frequency	Linear	–0.897	1.617	–0.55	Δ (s – o)-N frequency		2.284 1.648 1.39
	Quadratic	5.156	0.701	7.36			0.098 0.737 0.13
	Cubic	–2.667	0.663	–4.02			1.478 0.682 2.17
N predictability	Linear	–2.970	1.052	–2.82	Δ (s – o)-N predictor		0.071 0.807 0.09
	Quadratic	1.760	0.602	2.93			–0.041 0.617 –0.07
N+1 length	Linear	0.954	0.900	1.06	Δ (s – o)- N+1 length		–0.757 1.021 –0.74
	Quadratic	–2.717	0.571	–4.76			1.657 0.629 2.63
N+1 frequency	Linear	–3.244	0.944	–3.44	Δ (s – o)-N+1 frequency		–1.215 1.061 –1.15
	Quadratic	0.631	0.576	1.10			–0.926 0.649 –1.43
N+1 predictability	Linear	1.138	0.568	2.01	Δ (s – o)-N+1 predictor		1.587 0.627 2.53
	Quadratic	1.710	0.554	3.09			0.122 0.606 0.20
launch site distance	Linear	11.128	0.546	20.40	Δ (s – o)-launch site distance		3.475 1.075 3.23
	Quadratic	1.965	0.436	4.51			1.694 0.883 1.92
landing site	Linear	11.547	0.623	18.53	Δ (s – o)-landing site		–4.571 0.994 –4.60
	Quadratic	–1.803	0.532	–3.39			–2.259 0.858 –2.63
saccade size	Linear	12.684	0.552	22.97	Δ (s – o)-saccade size		–8.372 0.789 –10.61
	Quadratic	2.484	0.421	5.90			0.690 0.684 1.01
N-frequency \times N-1 freq	Linear	3.962	0.444	8.93	Δ (s – o)-N- frequency \times N-1 frequency		–1.289 0.464 –2.78
	Quadratic	1.650	0.440	3.75			–0.250 0.442 –0.57
	Cubic	–1.121	0.439	–2.55			0.444 0.449 0.99
N-frequency \times N+1 freq	Linear	1.461	0.431	3.39	Δ (s – o)- N-frequency \times N+1 frequency		–0.705 0.462 –1.53
	Quadratic	1.234	0.484	2.55			–0.068 0.508 –0.13
	Cubic	0.561	0.490	1.15			–0.543 0.514 –1.06
N-1 length \times launch site distance		–0.075	0.011	–6.67	Δ (s – o)- N-1 length \times launch site distance		–0.070 0.019 –3.75

Reading condition was specified as a treatment contrast with oral reading as reference. Therefore, main-effect coefficients in the left four columns represent mean and covariate effects (slopes) for oral reading; coefficients in the right four columns represent corresponding differences between oral and silent conditions (i.e., interactions between reading condition and covariate; differences in slopes between conditions). Thus, the sum of corresponding coefficients yields the effects for silent reading. Example: Δ (s – o) N-1-length: partial-effect estimate of difference between silent and oral condition for slopes associated with length of last word. Bold values indicate significant contrasts.

not be interpreted because this interaction is subordinated to the three-covariate interaction reading-condition \times N–1-frequency \times N-frequency, shown in **Figure 6** (top row) and discussed below.

Evidence for Differences between Oral and Silent Reading

The LMM provides test-statistics for the interaction between reading condition and each of the covariates. This interaction was significant for 9 of 12 covariates (see **Table 5**). Four of them were nested within a higher-order interaction and will be covered in this context (for sake of completeness all two-way interactions with reading mode are visualized in **Figure 5**). Others are due to a quantitative rather than qualitative change in the degree of non-linearity. For example, the negative cubic trend of word-N frequency was present in both reading modes,

but more pronounced in oral (–2.667) than in silent reading (–2.667 + 1.478 = –1.189). We had no specific expectation with respect to these differences; they were beyond the level of the current theoretical discourse. In the following we provide separate descriptions of these differences before an attempt at an integrative discussion.

Oral/silent main effect

As expected, silent reading was faster than oral reading. This at least partly reflects the need to wait for the slower voice, because otherwise working memory demands would become to great.

Oral/silent \times N length

There were positive linear and quadratic effects for silent reading, but only a (stronger) positive linear effect for oral reading, suggesting that the whole range of word lengths affects SFD in

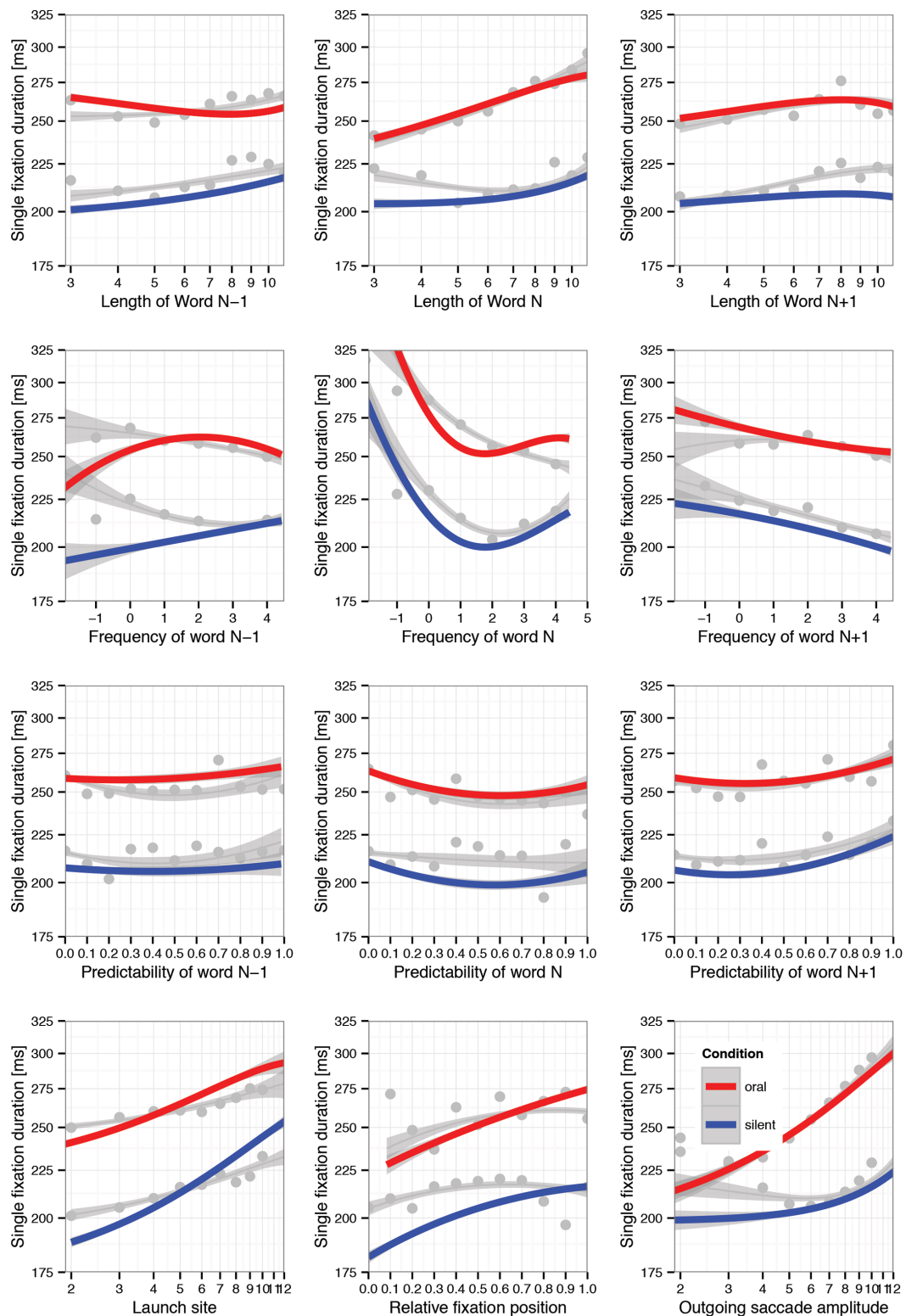


FIGURE 5 | Visualization of LMM estimates of interactions of reading condition (oral vs. silent) with 12 covariates. Colored lines represent partial effects; gray lines represent zero-order effects (i.e., simple regression of SFD on covariate); dots are observed mean SFDs suitably binned for the specific covariate; error bands represent 95% confidence intervals based on LMM residuals. The interactions of reading condition with N-1-frequency, N-frequency, N-1-length, and launch site distance should not be interpreted as such, because they are subordinated to higher-order interactions (see **Figure 6**). Note that effects are plotted on a log-scale of fixation durations.

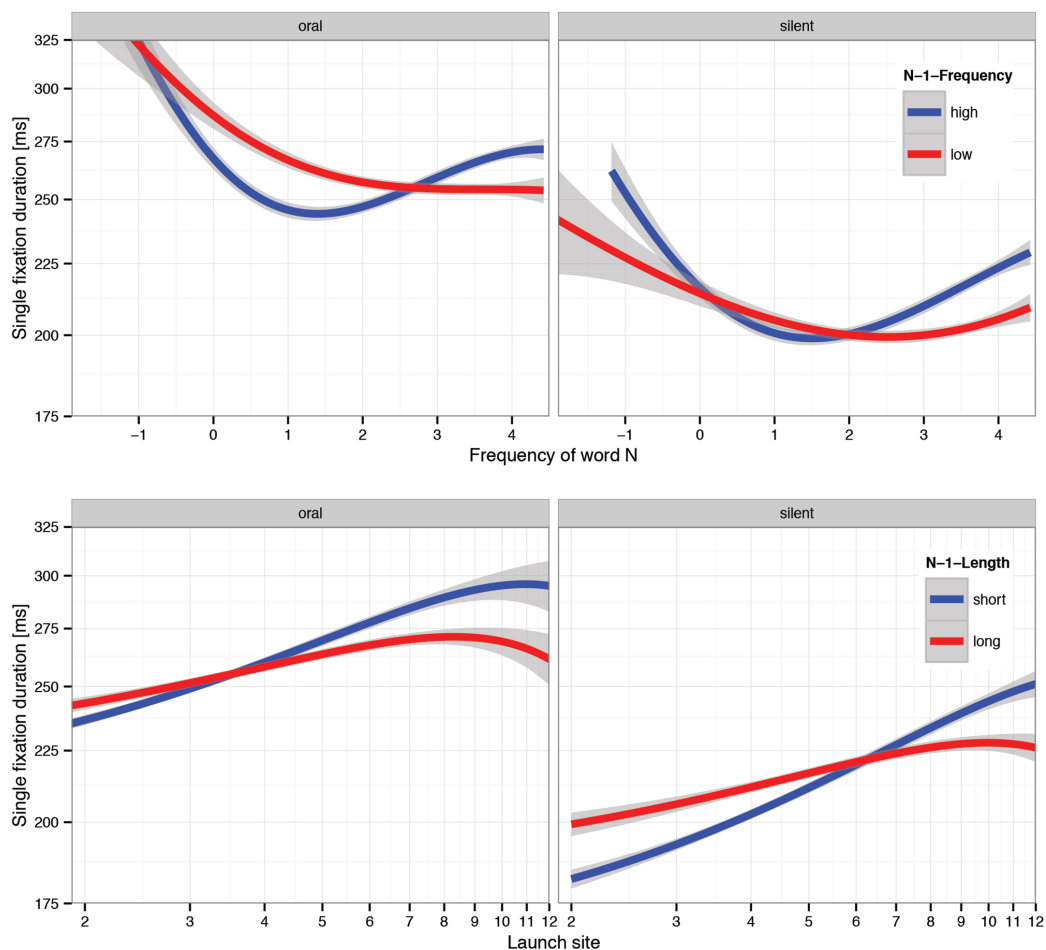


FIGURE 6 | Visualization of two LMM interactions involving three covariates; except reading condition, LMM used continuous covariates. Top row: oral vs. silent \times N-frequency \times N-1 frequency. **Bottom row:** oral vs. silent \times launch site \times N-1 length. Factors in panels are based on median splits; LMM estimation used continuous covariates. Error bands represent 95% confidence intervals based on LMM residuals. Note that effects are plotted on a log-scale of fixation durations.

oral reading, whereas the word length effect is restricted to longer words in silent reading.

Oral/silent \times N+1 length

There were negative quadratic effect for both reading modes, which were stronger for oral reading.

Oral/silent \times N+1 predictability

Positive linear and quadratic trends were observed in both reading modes; however, the linear component was stronger in silent reading. Since the effect of N+1 predictability has been linked to memory retrieval (Kliegl et al., 2006), this possibly indicates greater interference of ongoing articulatory planning with retrieval of expected words during oral reading.

Oral/silent \times landing site

Although there were strong positive linear and negative quadratic effects for both modes, the linear trend was stronger and the quadratic trend weaker in oral reading. We had no particular expectations about reading mode differences in landing position.

The IOVP-effect in silent reading has been linked to fast correction of mislocated fixations; it is possible that the oral reading constraint to maintain the EVS leads to a weaker influence of such lower-level oculomotor control mechanisms.

Oral/silent \times saccade amplitude

The most striking interaction with reading condition involved the outgoing saccade amplitude (see Figure 5, bottom right panel). There was a much stronger increase in SFD with the amplitude of the next saccade for oral than for silent reading. This interaction might be related to EVS: if a reader plans a long saccade, possibly involving skipping of the next word, and if at the same time aim the EVS must not become too large, one option (or even a necessity) is to wait a little longer.

Oral/silent \times N-1 frequency \times N frequency

Positive quadratic and negative cubic effects of word-N frequency were observed in both reading modes, but the latter was even stronger negative in oral reading. The quadratic trend, i.e., the upswing for the combination of high-frequency words N

and high-frequency words $N-1$, indicates preprocessing of the upcoming word; there is increased parafoveal preprocessing when foveal processing is easy (Henderson and Ferreira, 1990; Kliegl et al., 2006). Since the cubic effect mainly dampens the upswing caused by the quadratic effect, this is possibly related to the somewhat smaller perceptual span in oral reading. However, when word N received less preprocessing due to a difficult word $N-1$, frequency effects were monotonous across the whole range. This effect was even stronger during oral reading, when low-frequency words $N-1$ are also associated with articulatory difficulty. In support of this interpretation, this effect in oral reading appears to be linked to a large EVS (see **Figure 4**, third row).

Oral/silent \times launch site \times $N-1$ -length

There was also a very strong interaction between reading condition, launch site distance and length of the last word (see **Figure 6**, bottom row), analogous to the interaction between short vs. large EVS and the latter two covariates. The main source of this interaction is the steeper positive slope of launch site for short words $N-1$ during silent reading. This result is mainly due to a higher probability of skipping during silent reading (see **Table 1**) coupled with the well-known longer fixation durations following skipped words (Kliegl and Engbert, 2005). Again it suggests that parafoveal preprocessing of word N took place in both modes, but was more effective in silent reading.

In summary, although there were some differences due to reading mode, the overall pattern of effects looked rather similar for oral and silent reading. Most of the differences are probably related to the faster pace of silent reading. Some of them (i.e., the stronger linear outgoing saccade amplitude effect) appear to be linked to maintenance of the EVS; other effects (the stronger linear launch site distance effect, and the weaker negative cubic trend in current word frequency effect in silent reading) appear to indicate more parafoveal preprocessing in silent than in oral reading. The more restricted effects of both previous word length and previous word frequency suggest that lagged processing plays less of a role in oral than in silent reading. However, when word $N-1$ is of low frequency, word- N frequency effects are stronger in oral than in silent reading, suggesting a role of articulatory processing—note that during a fixation on word N , it is typically word $N-2$ that is pronounced, hence word $N-1$ is prepared for articulation. Finally, there was also a reading mode difference in the effect of $N+1$ predictability, which is stronger in silent than in oral reading, possibly suggesting phonological interference with lexical retrieval. Clearly, more experimental work is needed to support these interpretations.

Discussion

Oral reading is considerably slower than silent reading because of the demand to produce intelligible speech. In principle, longer fixation durations might offer a better chance to shift attention into the parafovea and thereby increase parafoveal aspects. The present results rather show that, despite some differences, eye movements during oral and silent reading are similar in many

respects². However, by analyzing the EVS, we have identified a previously unobserved, but very important regulatory influence on eye movements during reading. The present study is the first systematic investigation of how the spatial distance that the eye leads the voice regulates eye movement behavior. We have found the EVS to be predictive of regressions, refixations, and fixation durations. Indeed, effects of the EVS were among the strongest effects observed in the LMM analyses. Thus, the EVS during oral reading is a critical variable regulating eye movement behavior during reading. Given the documented effects of subvocalization on eye movements during silent reading, there is good reason to suspect that many of these influences are also at work during silent reading.

Before discussing the EVS in detail, we will focus on two methodological aspects that the present analyses brought forward. First, covariates of fixation durations typically exhibit substantial correlations (e.g., length and frequencies of word correlate around 0.70). Multivariate statistical tests of the significance of individual covariates take these correlations into account and yield partial effects. If covariates were uncorrelated, the direction and magnitude of the observed (zero-order) effect would be identical to the partial effect (i.e., there is no adjustment for uncorrelated covariates). With correlated covariates, in principle, there can be complete dissociation between zero-order and partial effects; Yan et al. (2014b) provide such dissociation for effects of word length and morphological complexity. In addition, in the presence of significant interactions between covariates, partial effects of the subordinate terms (i.e., the two main effects for a simple interaction) must not be interpreted independent of the interaction. The most striking example of this kind occurred for the $N-1$ -frequency effects in oral and silent reading, which were nested under higher-order interactions involving N -frequency and reading condition.

Second, the LMMs were based on continuous covariates (except, naturally, the oral vs. silent reading condition). For visualization of interactions we binned one or two such covariates. Therefore, when interpreting interaction plots one must keep in mind that the visualization may have missed a major source of the interaction, perhaps apparent with a different, usually more fine-grained binning. Notwithstanding this cautionary note, we are more impressed by the qualitative similarity of the interactions when comparing short or large EVS or when comparing oral and silent reading. In other words, as far as we can tell the significance of 3-covariate interactions are likely due to slight differences in the degree of non-linearity, not in the basic pattern. At this point such quantitative differences are

²One difference between oral and silent reading is that oral reading requires the participant to retrieve and articulate each word accurately while silent reading requires only that the reader have extracted a sufficient understanding of the sentence to answer an easy multiple choice question. Skimming strategies are likely to be used under these low comprehension requirements. This difference is potentially heightened by the removal of oral reading trials with articulation errors – i.e., the same cannot be done for silent reading trials. It's not necessarily clear how these selection effects would impact on the pattern of results, but it is unlikely to have had a major impact here, first, because of the overall similarity between the fixations duration patterns in oral and silent reading, and second, because of the relatively low number of sentences removed due to articulation errors.

clearly beyond the scope of theoretical proposals. Therefore, we primarily interpret the qualitatively similar interactions obtained across levels of EVS or across oral and silent reading as evidence for successful and non-trivial conceptual replications.

Returning to the EVS, the overall pattern of results suggests that the EVS is quite flexible, and is adjusted according to cognitive, oculomotor, and articulatory demands. Given that the voice proceeds fairly linearly through the text, most of the adjustment is actually performed by the oculomotor system. The eyes, and also the mind, could in principle proceed faster than the voice, since silent reading is faster than oral reading. However, the eyes need to wait for the voice because the size of the working memory buffer is limited. The major target value in the system controlling the eyes during oral reading is a constant EVS at fixation offset of about 10 letters, translating into an average temporal EVS of about 560 ms, in good agreement with Inhoff et al. (2011). The spatial EVS yielded a stronger signal for the dynamics than the temporal EVS, as suggested by the relatively narrow distribution of EVS at fixation offset compared to EVS at onset. This differentiation was much less pronounced for temporal EVS. There was also clear evidence that spatial offset-EVS is typically regulated within a fixation duration. Of course, sometimes this within-fixation adjustment fails and in these cases the probability of a refixation increases. If the EVS is too large for a refixation to effectively down-regulate the EVS, then a regression occurs with high probability.

It is worthwhile to put our results in a historical perspective. The absolute size of the onset-EVS is in surprisingly good agreement with Buswell's early recordings, using Charles Judd's sophisticated analog eye tracker with a tuning fork generating 50 Hz time stamps on a photo recording plate (Gray, 1917). In comparison, the EVS estimate from offline studies using the lights-off paradigm (Levin and Buckler-Addis, 1979) is widely off-track, and while it might measure something useful, the label "EVS" is somewhat of a misnomer. We suspect that our on-line EVS method measures how much is typically buffered, i.e., how much potential buffering capacity is actually used, whereas the offline method might measure its maximum under the most favorable circumstances. Why do the two estimates differ so widely? One reason could be the difference in tasks: whereas reading stops in the lights-off paradigm, it continues in the standard oral reading task, meaning that the working memory buffer needs to be continuously updated. Updating operations are costly and may be the reason for the much smaller estimate using the on-line measure.

Buswell furthermore reported that the EVS increased immediately prior to regressions, and was correlated with reading speed. Both of these results also hold in our data. Whereas Buswell had sophisticated recording equipment, he did not have any modern automated analysis tools or statistical models available. Thus, although he suspected that the EVS might be related to fixation duration, he was not able to find empirical evidence for this fact³, which was pronouncedly present in

our data. Failing to find evidence for a modulation of fixation duration by the EVS, Buswell examined other potential causes for long fixations, and found that difficult words like "hypnagogic" or "hallucinations" caused increased fixation durations. In modern terms, he discovered a word frequency effect on fixation duration.

Returning to our results, we went beyond Buswell by showing that the frequency effect, which is now well documented for fixation durations, also interacts with the EVS, such that the regulation of the EVS by fixation duration is much stronger for low frequency words. We also found this regulatory effect to be stronger for low-predictability words to the left of the fixated word. This pattern seems best explained by an oculomotor strategy that is influenced by cognitive processing and allows the eye to scout further ahead only when there is free capacity in the working memory buffer. Finally, the anecdotal observation that the eye often scouts ahead when a sentence is initially revealed, followed by a regression to synchronize with the voice and to maintain a manageable buffer size, is also consistent with the hypothetical oculomotor strategy. In summary, the oculomotor system has several means to regulate the EVS at offset, e.g., adjustment of fixation duration, of saccade direction, and of saccade amplitude, and all of them appear to be used.

Reading aloud involves working memory, specifically the phonological loop. Indeed, due to the serial output requirement, the working memory buffer during reading aloud is in some respect akin to a first-in, first-out queue. Phonological information is stored in the buffer in the serial order needed for output, since rearranging the phonological buffer is quite difficult. However, it is not clear whether the corresponding lexical units are also serially activated. In fact, one major difference between current computational models of eye movement control during reading is whether they assume serial or parallel lexical activation.

What then are the implications of our results for reading models? Although the temporal and spatial parameters are slightly different from silent reading, the general pattern of effects on fixation durations and probabilities speak for a similar control mechanism in both reading modes. Therefore, current models for silent reading can be used as a starting point for models of oral reading. Arguably, one necessary extension is an on-line working memory buffer that operates during reading. In particular, our results provide strong evidence that the oculomotor system is regulated by the cognitive system such that a relatively constant amount of information is buffered in working memory. Critically, this buffer is constantly updated during reading, requiring on-line control. The control process regulates both where- and when-decisions of eye movements: a large EVS goes along with increases in fixation durations as well as refixation and regression probabilities. Our data thus provided temporal constraints for eye movement models, since it can probably be assumed that a word that has been articulated is no longer a member of the set of potential saccade target locations. In the SWIFT model, for example, the lexical activation of a word should again be at zero by the time the word is articulated. Although oral reading is somewhat slower than silent reading due to the output demand to produce comprehensible speech, the size of the working memory buffer during silent reading is probably limited as well; it might be

³This is probably a consequence of the fact that Buswell (1920, pp. 80f) only examined the span differences between the 10 longest and the 10 shortest fixations, and not at all of the data points.

somewhat larger, but is surely on the same order of magnitude, given that fixation durations are not that dramatically different and given that sub-vocalization also takes place during silent reading. Indeed, it may well be that oral-reading models do a better job of predicting performance in silent reading than the original models.

Modeling oral reading would thus be a worthwhile effort, and has implications far beyond eye movement control. At least in the U.S. and the UK, oral reading fluency is a major arena of reading instruction and a benchmark of educational success. In most of the education-related reading literature this is treated as a monolithic construct that is examined in relation to other equally abstract latent variables like “decoding” and “comprehension.” Research on the EVS has the potential to crack this black box open and begins to understand oral reading fluency in a much more fundamental way.

We presented a first description of the EVS, mainly using the approach of statistical control in multivariate analyses. Of course, further experimental analyses looking at specific aspects

of the data will reveal new insights. In summary, we reported a detailed description of how during the EVS oral reading is regulated by cognitive processing difficulty. We discovered quite a few thought-provoking aspects of the cognitive regulation of the interplay between eye and voice during reading. The study provides an important first step at understanding how eye and voice are coordinated to achieve fast reading with a manageable working memory load.

Acknowledgments

This work was funded by ESF (05_ECRP-FP06) and DFG (KL 955/7-1) grants to RK. We acknowledge the support of the Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of University of Potsdam. We thank Petra Schienmann for help with data collection and Manon Jones, Alan Kennedy, Stephen Monsell, Ralph Radach, and Aaron Veldre for helpful comments on an earlier version

References

- Almargot, D., Dansac, C., Chesnet, D., and Fayol, M. (2007). “Parallel processing before and after pauses: a combined analysis of graphomotor and eye movements during procedural text production,” in *Writing and Cognition: Research and Applications*, eds M. Torrance, L. van Waas, and D. Galbraith (Amsterdam: Elsevier).
- Anderson, I. H., and Swanson, D. E. (1937). Common factors in eye movements in silent and oral reading. *Psychol. Monogr.* 48, 61–77. doi: 10.1037/h0093393
- Angele, B., Schotter, E. R., Slattery, T. J., Tenenbaum, T. L., Bicknell, K., and Rayner, K. (2015). Do successor effects in reading reflect lexical parafoveal processing? Evidence from corpus-based and experimental eye movement data. *J. Mem. Lang.* 7, 76–96.
- Ashby, J., Yang, J., Evans, K. H. C., and Rayner, K. (2012). Eye movements and the perceptual span in silent and oral reading. *Attent. Percept. Psychophys.* 74, 634–640. doi: 10.3758/s13414-012-0277-0
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends Cogn. Sci.* 4, 417–423. doi: 10.1016/S1364-6613(00)01538-2
- Baddeley, A. D., and Hitch, G. J. (1974). “Working memory,” in *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 8, ed. G. H. Bower (New York: Academic Press), 47–89.
- Bates, D., Kliegl, R., Vashishth, S., and Baayen, H. R. (2015a). Parsimonious mixed models. *arXiv*: 1506.04967
- Bates, D., Maechler, M., Bolker, B. M., and Walker, S. (2015b). *lme4: Linear Mixed-Effects Models Using Eigen and S4*. *ArXiv e-print; Submitted to Journal of Statistical Software*. Available at: <http://arxiv.org/abs/1406.5823>
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.* 5, 341–345.
- Boersma, P., and Weenink, D. (2010). *Praat: Doing Phonetics by Computer [Computer Program]*. Version 5.1. Available at: <http://www.praat.org/> [accessed 2, July 2010].
- Buswell, G. T. (1920). *An Experimental Study of the Eye-Voice Span in Reading*. Supplementary Educational Monographs No. 17. Chicago: Chicago University Press.
- Buswell, G. T. (1922). *Fundamental Reading Habits: A Study of Their Development*. Chicago: University of Chicago Press.
- Butsch, R. L. C. (1932). Eye movements and the eye-hand span in typewriting. *J. Educ. Psychol.* 23, 104–121. doi: 10.1037/h0073463
- De Luca, M., Pontillo, M., Primativo, S., Spinelli, D., and Zoccolotti, P. (2013). The eye-voice lead during oral reading in developmental dyslexia. *Front. Hum. Neurosci.* 7:696. doi: 10.3389/fnhum.2013.00696
- Eiter, B., and Inhoff, A. W. (2010). Visual word recognition during reading is followed by subvocal articulation. *J. Exp. Psychol.* 36, 457–470. doi: 10.1037/a0018278
- Engbert, R., Longtin, A., and Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Res.* 42, 621–636. doi: 10.1016/S0042-6989(01)00301-7
- Engbert, R., Nuthmann, A., Richter, E. M., and Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychol. Rev.* 112, 777–813. doi: 10.1037/0033-295X.112.4.777
- Ericsson, K. A., and Kintsch, W. (1995). Long-term working memory. *Psychol. Rev.* 102, 211–245. doi: 10.1037/0033-295X.102.2.211
- Fairbanks, G. (1937). Eye-movements and voice in oral reading. *Psychol. Monogr.* 48, 78–107. doi: 10.1037/h0093394
- Furneaux, S., and Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proc. R. Soc. Lond. B Sci.* 266, 2453–2440. doi: 10.1098/rspb.1999.0943
- Gray, C. T. (1917). *Types of Reading Ability as Exhibited Through Tests and Laboratory Experiments*. Supplementary Educational Monographs, Vol. 1, No. 5. Chicago: University of Chicago Press.
- Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., et al. (2011). dlexDB – eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychol. Rundschau* 62, 10–20. doi: 10.1026/0033-3042/a000029
- Heister, J., Würzner, K. M., and Kliegl, R. (2012). “Analysing large datasets of eye movements during reading,” in *Visual Word Recognition*, ed. J. S. Adelman (Hove: Psychology Press), 102–130.
- Henderson, J. M., and Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: implications for attention and eye movement control. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 417–429.
- Henderson, M., Dixon, P., Petersen, A., Twilley, L. C., and Ferreira, F. (1995). Evidence for the use of phonological representations during transsaccadic word recognition. *J. Exp. Psychol.* 21, 82–97.
- Hohenstein, S., and Kliegl, R. (2014). Semantic preview benefit during reading. *J. Exp. Psychol.* 40, 166–190.
- Hohenstein, S., and Kliegl, R. (2015). *remef: Remove Partial Effects*. R Package Version 1.0.6.9000. Available at: <https://github.com/hohenstein/remef/>
- Hohenstein, S., Laubrock, J., and Kliegl, R. (2010). Semantic preview benefit in eye movements during reading: a parafoveal fast-priming study. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 1150–1170. doi: 10.1037/a0020233
- Inhoff, A. W., Connine, C., Radach, R., and Heller, D. (2004). The phonological representation of words in working memory during sentence reading. *Psychon. Bull. Rev.* 11, 320–325. doi: 10.3758/BF03196577

- Inhoff, A. W., and Gordon, A. M. (1998). Eye movements and eye-hand coordination during typing. *Curr. Dir. Psychol. Sci.* 6, 153–157. doi: 10.1111/1467-8721.ep10772929
- Inhoff, A. W., Morris, R., and Calabrese, J. (1986). Eye movements in skilled transcription typing. *Bull. Psychon. Soc.* 24, 113–114. doi: 10.3758/BF03330519
- Inhoff, A. W., and Radach, R. (2014). Parafoveal preview benefits during oral and silent reading: testing the parafoveal extraction hypothesis. *Vis. Cogn.* 22, 354–376. doi: 10.1080/13506285.2013.879630
- Inhoff, A. W., Solomon, M., Radach, R., and Seymour, B. (2011). Temporal dynamics of the eye voice span and eye movement control during oral reading. *J. Cogn. Psychol.* 23, 543–558. doi: 10.1080/20445911.2011.546782
- Inhoff, A. W., and Wang, J. (1992). Encoding of text, manual movement planning, and eye-hand coordination during typing. *J. Exp. Psychol.* 18, 437–448.
- Jones, M. W., Ashby, J., and Branigan, H. P. (2013). Dyslexia and fluency: parafoveal and foveal influences on rapid automatized naming. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 554–567. doi: 10.1037/a0029710
- Jones, M. W., Snowling, M. J., and Moll, K. (2015). What automaticity deficit? Activation of lexical information by readers with dyslexia in a RAN Stroop-switch task. *J. Exp. Psychol. Learn. Mem. Cogn.* (in press).
- Kennedy, A., and Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Res.* 45, 153–168. doi: 10.1016/j.visres.2004.07.037
- Kliegl, R. (2007). Towards a perceptual-span theory of distributed processing in reading: a reply to Rayner. Pollatsek, Drieghe, & Reichle (2007). *J. Exp. Psychol.* 138, 530–537.
- Kliegl, R., and Engbert, R. (2005). Fixation durations before word skipping in reading. *Psychon. Bull. Rev.* 12, 132–138. doi: 10.3758/BF03196358
- Kliegl, R., Nuthmann, A., and Engbert, R. (2006). Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *J. Exp. Psychol.* 135, 12–35. doi: 10.1037/0096-3445.135.1.12
- Land, M. F., and Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philos. Trans. R. Soc. Lond. B* 352, 1231–1239. doi: 10.1098/rstb.1997.0105
- Land, M. F., and McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nat. Neurosci.* 3, 1340–1345. doi: 10.1038/81887
- Land, M. F., and Tatler, B. W. (2009). *Looking and Acting. Vision and Eye Movements in Natural Behaviour*. Oxford: Oxford University Press.
- Laubrock, J., and Hohenstein, S. (2012). Orthographic consistency and parafoveal preview benefit: a resource-sharing account of language differences in processing of phonological and semantic codes. *Behav. Brain Sci.* 35, 292–293. doi: 10.1017/S0140525X12000209
- Levin, H., and Buckler-Addis, A. (1979). *The Eye-Voice Span*. Cambridge, MA: MIT Press.
- Manguel, A. (1996). *A History of Reading*. New York: Viking.
- Nuthmann, A., Engbert, R., and Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Res.* 45, 2201–2217. doi: 10.1016/j.visres.2005.02.014
- Pan, J., Yan, M., Laubrock, J., Shu, H., and Kliegl, R. (2013). Eye-voice span during Rapid Automatized Naming of digits and dice in Chinese normal and dyslexic children. *Dev. Sci.* 16, 967–979. doi: 10.1111/desc.12075
- Pollatsek, A., Lesch, M., Morris, R. K., and Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *J. Exp. Psychol.* 18, 148–162.
- Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cogn. Psychol.* 7, 65–81. doi: 10.1016/0010-0285(75)90005-5
- Rayner, K. (1978). “Foveal and parafoveal cues in reading,” in *Attention and Performance VII*, ed. J. Requin (Hillsdale, NJ: Erlbaum), 149–162.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422. doi: 10.1037/0033-2909.124.3.372
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* 62, 1457–1506. doi: 10.1080/17470210902816461
- Rayner, K., and Duffy, S. A. (1986). Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. *Mem. Cogn.* 14, 191–201. doi: 10.3758/BF03197692
- Rayner, K., McConkie, G. W., and Ehrlich, S. F. (1978). Eye movements and integrating information across fixations. *J. Exp. Psychol.* 4, 529–544.
- Rayner, K., Pollatsek, A., Ashby, J., and Clifton, C. Jr. (2012). *The Psychology of Reading*, 2nd Edn. New York, NY: Psychology Press.
- Rayner, K., Pollatsek, A., Drieghe, D., Slattery, T. J., and Reichle, E. D. (2007). Tracking the mind during reading via eye movements: comments on Kliegl, Nuthmann, and Engbert (2006). *J. Exp. Psychol.* 136, 520–529.
- R Development Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: <http://www.R-project.org>.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., and Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychol. Rev.* 105, 125–157. doi: 10.1037/0033-295X.105.1.125
- Reichle, E. D., Rayner, K., and Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: comparisons to other models. *Behav. Brain Sci.* 26, 445–476. discussion 477–526. doi: 10.1017/S0140525X03000104
- Reilly, R. G., and Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cogn. Syst. Res.* 7, 34–55. doi: 10.1016/j.cogsys.2005.07.006
- Schotter, E. R. (2013). Synonyms provide semantic preview benefit in english. *J. Mem. Lang.* 69, 619–633. doi: 10.1016/j.jml.2013.09.002
- Sperlich, A., Schad, D. J., and Laubrock, J. (2015). When preview information starts to matter: development of the perceptual span in German beginning readers. *J. Cogn. Psychol.* 27, 511–530. doi: 10.1080/20445911.2014.993990
- Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.* 74, 1–29. doi: 10.1037/h0093759
- Tiffin, J. (1934). Simultaneous records of eye-movements and the voice in oral reading. *Science* 80, 430–431. doi: 10.1126/science.80.2080.430
- Truitt, F. E., Clifton, C., Pollatsek, A., and Rayner, K. (1997). The perceptual span and the eye-hand span in sight reading music. *Vis. Cogn.* 4, 143–161. doi: 10.1080/713756756
- Vitu, F., McConkie, G. W., Kerr, P., and O'Regan, J. K. (2001). Fixation location effects on fixation durations during reading: an inverted optimal viewing position effect. *Vis. Res.* 41, 3513–3533. doi: 10.1016/S0042-6989(01)00166-3
- Wotschack, C., and Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *Q. J. Exp. Psychol. (Hove)* 66, 548–562. doi: 10.1080/17470218.2011.625094
- Yan, M., Luo, Y., and Inhoff, A. W. (2014a). Syllable articulation influences foveal and parafoveal processing of words during the silent reading of Chinese sentences. *J. Mem. Lang.* 75, 93–103. doi: 10.1016/j.jml.2014.05.007
- Yan, M., Zhou, W., Shu, H., Yusupu, R., Krügel, A., and Kliegl, R. (2014b). Eye movements guided by morphological structure: evidence from Uighur language. *Cognition* 132, 181–215. doi: 10.1016/j.cognition.2014.03.008
- Yan, M., Richter, E., Shu, H., and Kliegl, R. (2009). Chinese readers extract semantic information from parafoveal words during reading. *Psychon. Bull. Rev.* 16, 561–566. doi: 10.3758/PBR.16.3.561

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

Copyright © 2015 Laubrock and Kliegl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Literacy transforms speech production

Meredith Saletta *

Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA, USA

Keywords: orthography, orthographic transparency, speech production, kinematics, implicit learning

Traditionally, literacy, and speech production have been investigated separately. Studies of development demonstrate that children are able to meet the challenge of language learning across modalities, and that adults may experience difficulties in one or both modalities. Yet, it is rare to find a conceptual connection between these two processes. I argue that speaking and reading actually share important mechanisms. Specifically, orthographic characteristics of written words influence spoken as well as written language, as indicated by measures of both explicit and implicit language processing. These effects can be quantified by examining speech movement variability. An important question regarding both limb and speech motor variability is whether it is interpreted as facilitating or inhibiting the process of learning. New lines of research may explore this question by quantifying the depth of learning when stimuli are produced with greater stability or greater variability. The developmental progressions of speaking and reading also contain important parallels, which are manifest differently in individuals with varying degrees of language and reading skills. This is an important and timely issue, as it can promote theoretical accounts of language processing and respond to the clinical reality that many individuals demonstrate both spoken and written language difficulties.

OPEN ACCESS

Edited by:

Simone Sulpizio,
University of Trento, Italy

Reviewed by:

Pierluigi Zoccolotti,
Sapienza University of Rome, Italy

***Correspondence:**

Meredith Saletta
meredith-saletta@uiowa.edu

Specialty section:

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

Received: 17 July 2015

Accepted: 11 September 2015

Published: 29 September 2015

Citation:

Saletta M (2015) Literacy transforms
speech production.
Front. Psychol. 6:1458.
doi: 10.3389/fpsyg.2015.01458

Orthographic Interference

As an individual acquires literacy skills, changes occur to his/her processing of spoken, as well as written, language (e.g., Ventura et al., 2004; Ziegler et al., 2004; Alario et al., 2007; Burgos et al., 2014). This phenomenon is known in the literature as *orthographic interference*; orthography has a facilitative or disruptive effect on the perception of the spoken word. Orthographic interference affects literate individuals when they learn a new word. Learners can integrate the new word's orthographic characteristics into its mental representation, thus changing their entire perception of the word. Orthographic interference is clearly present in experiences such as the Stroop Color and Word Test (Stroop, 1935; see the review by MacLeod, 1991), in which the reader is unable to deactivate the written word's orthography. These paradigms indicate that characteristics of the written word impact the processes of both speech and reading.

The influences of orthography on perception have been well documented using behavioral paradigms. Classic studies of orthographic interference revealed that individuals who are competent speakers but illiterate, or literate only in a non-alphabetic orthography, are unable to verbally blend or segment phonemes (Moraes et al., 1979; Read et al., 1986). Other early results indicate that orthography influences rhyme detection (Seidenberg and Tanenhaus, 1979), and that listeners report differences in the number of phonemes in homophones because of the presence of an additional grapheme (as in the pair "flour/flower," in which the second spelling was often thought to have an extra phoneme; Ehri and Wilce, 1980).

Recent works examining this phenomenon have focused on reaction time (Miller and Swick, 2003; Ziegler and Muneaux, 2007) or priming effects (Damian and Bowers, 2003). Rastle et al. (2011) manipulated spelling-sound consistency in novel words during picture naming and auditory

lexical decision tasks, and determined that orthographic factors influence speech production even when the speaker is not reading. Orthographic interference has also been examined via imaging studies, including measures of event-related potentials (ERP; Weber-Fox et al., 2003; Pattamadilok et al., 2009), positron emission tomography (PET; Castro-Caldas and Reis, 2003), and functional magnetic resonance imaging (fMRI; Shankweiler et al., 2008).

Influences of Orthography on Implicit Processing

Most of the above works focused on explicit learning. Participants were faced with a choice or required to give a response based on their conscious awareness of the stimuli. Fewer studies have investigated the influence of orthography on implicit learning. This type of processing can be quantified by measuring motor learning, which does not require conscious awareness; the participant need only produce the stimuli, not make decisions about them. Researchers may quantify individuals' articulatory stability as they speak and read aloud—a highly promising measure which provides a window into implicit learning.

Measures of articulatory stability have been used to assess implicit processing in relation to task load in children and adults with typical language development or speech and language disorders (Goffman et al., 2007; McMillan et al., 2009; Heisler et al., 2010). In these measures, kinematic parameters of movement are quantified, and the degree to which repeated movements, or productions of an utterance, converge on a single underlying template are then determined (Smith et al., 2000). These measures have been used to examine diverse phenomena in speech and language production, from the effects of altering a single phoneme (Goffman and Smith, 1999), to development and maturation (Wohlert and Smith, 1998), to stuttering and other motor speech disorders (Kleinow et al., 2001).

For the first time, this measure has been applied to individuals with differences in reading skills (Salletta et al., 2015). We indexed implicit learning by analyzing participants' segmental accuracy and articulatory stability as they learned non-words varying in modality of presentation (auditory or written) and orthographic transparency (transparent/consistent spelling vs. opaque/inconsistent spelling). Findings indicate that speech production is more accurate when non-word stimuli are read aloud than when they are simply heard and repeated. Crucially, this increase in accuracy is present even after the written text is removed. This indicates that the speakers integrated the orthographic characteristics of the non-words into their lexical representations, and supports conceptualizing reading as an interactive (rather than strictly top-down) process.

Movement Variability: Adaptive or Negative?

When examining these speech production findings, a crucial point is that the interpretation of the increased stability is unclear. Traditionally, movement stability has been viewed from

the perspective that greater stability is indicative of superior learning or production efficiency, and greater variability is a negative process. For instance, researchers exploring quiet stance on a forceplate considered increased sway to represent postural instability and decreased sway to indicate greater stability (Woollacott et al., 1986). Greater variability has been shown in elderly individuals who experience a slowing of online sensorimotor mechanisms, rendering them less able to modulate their sway (Fraizer and Mitra, 2008). Within the speech domain, children have been found to be more variable in their articulatory output than adults (Smith and Zelaznik, 2004), and clinical populations, such as individuals who stutter, are also generally more variable (MacPherson and Smith, 2013).

However, when investigating this effect more deeply, this interpretation is unclear. Greater movement variability may be an adaptive process which facilitates learning. In conditions of learning, such as when a child's system develops or an adult's system changes due to aging, motor variability can indicate flexibility. While perhaps counterintuitive, this has been shown in the motor control literature in several paradigms. Healthy adults and patients with Parkinson's disease may demonstrate increased sway as a strategy to enable the individual to overcome perturbations to his/her balance. In these cases, a decrease in postural sway could point toward stiffening and freezing of the degrees of freedom, reducing the individual's ability to recover from a perturbation (Chagdes et al., in revision). Studies of infants' reaching trajectories indicate that reaching is not restricted to the arm independently, but differs depending upon the body's posture, reaching from different positions and at different speeds, the freedom of the other arm, and other factors. Infants experience regression of trajectory control even after practicing reaching for several months, which indicates that early variability actually facilitates learning (Thelen and Spencer, 1998). Waddington and Adams (2003) discovered that wearing textured insoles to increase movement discrimination improved soccer players' abilities to discriminate ankle inversion, thus potentially diminishing their risk of lower limb injury. From this paradigm, Davids et al. (2003) argue that variability of motor output is essential for individuals to adapt to dynamic environments.

Viewing postural or limb motor variability as an adaptive process may be more intuitive than applying this concept to speech variability. However, it is important to note that increased variability in speech production is not always a function of a disordered system. Rather, it may actually aid developing speech and language learners in finding the optimal and dynamically changing (flexible) production patterns. We can apply this perspective to individuals' articulatory stability when speaking or reading aloud. Our previous work (Salletta et al., 2015) indicates that speech movement was more variable when reading words which were presented in the written modality with a relatively opaque spelling. Based on the motor control literature, we may conclude that participants' speech movements became more variable when they were exposed to orthography in the more challenging task because the participants were compelled to interact with the words at a deeper level. This facilitates their

reorganization of their representations of the non-word and the integration of this new information.

Influences of Orthography on Poor Readers

The interaction between speaking and reading aloud varies across individuals with differing degrees of reading skill. Children who are acquiring reading skills atypically may fail to integrate orthographic information into their process of developing phonological representations. Speech and reading development contain important parallels. In children with typical development, the processing of spoken language follows a continuum from holistic to segmental processing. According to Nitttrouer et al. (1989), children's earliest language is mediated by meaning. The earliest contrastive unit used by children is often one or a few syllables composing the word or formulaic phrase, rather than the phoneme or feature. By their second birthdays, children begin to reorganize their phonological processing from the whole word to a more segmental level (Dodd and McIntosh, 2009). Then, as toddlers mature into preschoolers, differentiation below the level of the syllable gradually emerges.

The onset of reading contributes to another reorganization, similar to that observed in spoken language. En route to achieving reading expertise, children pass through several stages. To achieve proficient reading, there is first a visual/logographic stage, during which children utilize salient graphic features to recognize the printed word (Masonheimer et al., 1984). This emergent literacy period gives way to the alphabetic stage, in which children are able to use the rules of grapheme-phoneme correspondence to decode new words (Kamhi and Catts, 2012). Proficient readers can achieve a more automatic identification of written words via visual sight word recognition (Ventura et al., 2007). Ultimately, children with reading difficulties fail to perform this reorganization effectively and efficiently. Although not every theorist supports this stage hypothesis of learning to read (e.g., Stuart and Coltheart, 1988)—indeed, specifically, there may not be a logographic stage in languages with regular orthographies (Wimmer and Hummer, 1990)—it is remarkable to consider how similarly the developmental courses of speaking and reading proceed, further supporting the interaction of these two phenomena.

This transformation is also apparent in the differences between typical and atypical adult readers (Castro-Caldas and Reis, 2003; Ziegler et al., 2003). Difficulty in acquiring literacy skills has cascading effects on neural organization. Numerous neuroimaging studies have revealed differences in visual skills

(Dehaene et al., 2015) and language processing in adults with poor reading skills (e.g., Shankweiler et al., 2008). Adults with reading disabilities may use a relatively global or coarse coding rather than the fine-grained grapheme-phoneme mappings used by typical readers. This means that they may rely to a greater extent on words' visual characteristics than their phonological characteristics (Lavidor et al., 2006), and thus, that poor readers are more influenced by orthographic irregularities. In contrast, according to Bolger et al. (2008), more proficient readers are influenced to a greater degree by phonological/orthographic inconsistency. Thus, individuals with higher reading skills should be more sensitive to changes to orthographic transparency. It remains to be seen which of these conclusions receives empirical support in future studies examining implicit learning and speech production. Furthermore, these differences may be more apparent in languages with varying degrees of orthographic consistency. Serrano and Defior (2008) state that languages with greater orthographic transparency may be associated with less severe reading difficulties. Furthermore, children with reading impairment may experience greater difficulties when reading languages which are more opaque (Kamhi and Catts, 2012).

Conclusions

Speaking and reading aloud are connected by shared mechanisms of processing and learning. Orthography influences not only reading, but speech production as well. Both reading and speaking are influenced by input, such as whether a new word is heard or read, in that reading and speaking (i.e., reading aloud) increases accuracy and stability over hearing and speaking (i.e., repeating). Speech production, from phonological encoding and articulatory planning to articulatory execution, is profoundly transformed by orthographic knowledge. Adding more auditory input does not change the production of the new word, but adding orthographic input may increase speech accuracy and cause shifts in articulatory variability. It is possible that, unlike previous interpretations of limb movement variability, speech movement variability might actually be an adaptive process which promotes depth of learning. Literate individuals can, and do, integrate orthography into a new word's representation—even without making a conscious decision to do so. All of these effects differ in speakers with varying degrees of reading proficiency. Ultimately, a word's written characteristics impact even the performance of tasks which do not involve written text. These concepts support the idea of reading and speaking as interactive processes which are mediated by differences in reading skill.

References

- Alario, F. X., Perre, L., Castel, C., and Ziegler, J. C. (2007). The role of orthography in speech production revisited. *Cognition* 102, 464–475. doi: 10.1016/j.cognition.2006.02.002
- Bolger, D. J., Hornickel, J., Cone, N. E., Burman, D. D., and Booth, J. R. (2008). Neural correlates of orthographic and phonological consistency effects in children. *Hum. Brain Mapp.* 29, 1416–1429. doi: 10.1002/hbm.20476
- Burgos, P., Cucchiari, 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
- Castro-Caldas, A., and Reis, A. (2003). The knowledge of orthography is a revolution in the brain. *Read. Writ.* 16, 81–97. doi: 10.1023/A:1021798106794
- Damian, M., and Bowers, J. (2003). Effects of orthography on speech production in a form-preparation paradigm. *J. Mem. Lang.* 49, 119–132. doi: 10.1016/S0749-596X(03)00008-1

- Davids, K., Shuttleworth, R., Button, C., Renshaw, I., and Glazier, P. (2003). "Essential noise"—Enhancing variability of informational constraints benefits movement control: a comment on Waddington and Adams (2003). *Br. J. Sports Med.* 38, 601–605. doi: 10.1136/bjsm.2003.007427
- Dehaene, S., Cohen, L., Morais, J., and Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nat. Rev. Neurosci.* 16, 234–244. doi: 10.1038/nrn3924
- Dodd, B., and McIntosh, B. (2009). Two-year-old phonology: impact of input, motor, and cognitive abilities on development. *J. Child Lang.* 37, 1027–1046. doi: 10.1017/S0305000909990171
- Ehri, L. C., and Wilce, L. S. (1980). The influence of orthography on readers' conceptualization of the phonemic structure of words. *Appl. Psycholinguist.* 1, 371–385. doi: 10.1017/S0142716400009802
- Fraizer, E. V., and Mitra, S. (2008). Methodological and interpretive issues in posture-cognition dual-tasking in upright stance. *Gait Posture* 27, 271–279. doi: 10.1016/j.gaitpost.2007.04.002
- Goffman, L., Gerken, L., and Lucchesi, J. (2007). Relations between segmental and motor variability in prosodically complex nonword sequences. *J. Speech Lang. Hear. Res.* 50, 444–458. doi: 10.1044/1092-4388(2007)031
- Goffman, L., and Smith, A. (1999). Development and phonetic differentiation of speech movement patterns. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 649–660. doi: 10.1037/0096-1523.25.3.649
- Heisler, L., Goffman, L., and Younger, B. (2010). Lexical and articulatory interactions in children's language production. *Dev. Sci.* 13, 722–730. doi: 10.1111/j.1467-7687.2009.00930.x
- Kamhi, A. G., and Catts, H. W. (2012). "Reading development," in *Language and Reading Disabilities, 3rd Edn.*, eds A. G. Kamhi and H. W. Catts (Boston, MA: Pearson), 24–44.
- Kleinow, J., Smith, A., and Ramig, L. O. (2001). Speech motor stability in IPD: effects of rate and loudness manipulation. *J. Speech Lang. Hear. Res.* 44, 1041–1051. doi: 10.1044/1092-4388(2001)082
- Lavidor, M., Johnston, R., and Snowling, M. J. (2006). When phonology fails: orthographic neighbourhood effects in dyslexia. *Brain Lang.* 96, 318–329. doi: 10.1016/j.bandl.2005.06.009
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychol. Bull.* 109, 163–203. doi: 10.1037/0033-2909.109.2.163
- MacPherson, M. K., and Smith, A. (2013). Influences of sentence length and syntactic complexity on the speech motor control of children who stutter. *J. Speech Lang. Hear. Res.* 56, 89–102. doi: 10.1044/1092-4388(2012)11-0184
- Masonheimer, P., Drum, P., and Ehri, L. (1984). Does environmental print identification lead children into word learning? *J. Read. Behav.* 16, 257–271.
- McMillan, C. T., Corley, M., and Lickley, R. J. (2009). Articulatory evidence for feedback and competition in speech production. *Lang. Cogn. Process.* 24, 44–66. doi: 10.1080/01690960801998236
- Miller, K. M., and Swick, D. (2003). Orthography influences the perception of speech in alexic patients. *J. Cogn. Neurosci.* 15, 981–990. doi: 10.1162/089892903770007371
- Morais, J., Cary, L., Alegria, J., and Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition* 7, 323–331. doi: 10.1016/0010-0277(79)90020-9
- Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (1989). The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *J. Speech Hear. Res.* 32, 120–132. doi: 10.1044/jshr.3201.120
- Pattamadilok, C., Perre, L., Dufau, S., and Ziegler, J. (2009). On-line orthographic influences on spoken language in a semantic task. *J. Cogn. Neurosci.* 21, 169–179. doi: 10.1162/jocn.2009.21014
- Rastle, K., McCormick, S. F., Bayliss, L., and Davis, C. J. (2011). Orthography influences the perception and production of speech. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1588–1594. doi: 10.1037/a0024833
- Read, C., Zhang, Y.-F., Nie, H.-Y., and Ding, B.-Q. (1986). The ability to manipulate speech sounds depends on knowing alphabetic spelling. *Cognition* 24, 31–44. doi: 10.1016/0010-0277(86)90003-X
- Saletta, M., Goffman, L., and Brentari, D. (2015). Reading skill and exposure to orthography influence speech production. *Appl. Psycholinguist.* 36, 1–24. doi: 10.1017/S0142716415000053
- Seidenberg, M. S., and Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *J. Exp. Psychol. Hum. Learn. Mem.* 5, 546–554. doi: 10.1037/0278-7393.5.6.546
- Serrano, F., and Defior, S. (2008). Dyslexia speed problems in a transparent orthography. *Ann. Dyslexia* 58, 81–95. doi: 10.1007/s11881-008-0013-6
- Shankweiler, D., Mencl, W. E., Braze, D., Tabor, W., Pugh, K. R., and Fulbright, R. K. (2008). Reading differences and brain: cortical integration of speech and print in sentence processing varies with reader skill. *Dev. Neuropsychol.* 33, 745–775. doi: 10.1080/87565640802418688
- Smith, A., Johnson, M., McGillem, C., and Goffman, L. (2000). On the assessment of stability and patterning of speech movements. *J. Speech Lang. Hear. Res.* 43, 277–286.
- Smith, A., and Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Dev. Psychobiol.* 45, 22–33. doi: 10.1002/dev.20009
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651
- Stuart, M., and Coltheart, M. (1988). Does reading develop in a sequence of stages? *Cognition* 30, 139–181. doi: 10.1016/0010-0277(88)90038-8
- Thelen, E., and Spencer, J. P. (1998). Postural control during reaching in young infants: a dynamic systems approach. *Neurosci. Biobehav. Rev.* 22, 507–514. doi: 10.1016/S0149-7634(97)00037-7
- Ventura, P., Morais, J., and Kolinsky, R. (2007). The development of the orthographic consistency effect in speech recognition: from sublexical to lexical involvement. *Cognition* 105, 547–576. doi: 10.1016/j.cognition.2006.12.005
- Ventura, P., Morais, J., Pattamadilok, C., and Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Lang. Cogn. Process.* 19, 57–95. doi: 10.1080/01690960344000134
- Waddington, G., and Adams, R. (2003). Football boot insoles and sensitivity to extent of ankle inversion movement. *Br. J. Sports Med.* 37, 170–175. doi: 10.1136/bjsm.37.2.170
- Weber-Fox, C., Spencer, R., Cuadrado, E., and Smith, A. (2003). Development of neural processes mediating rhyme judgments: phonological and orthographic interactions. *Dev. Psychobiol.* 43, 128–145. doi: 10.1002/dev.10128
- Wimmer, H., and Hummer, P. (1990). How German-speaking first graders read and spell: doubts on the importance of the logographic stage. *Appl. Psycholinguist.* 11, 349–368. doi: 10.1017/S0142716400009620
- Wohler, A. B., and Smith, A. (1998). Spatiotemporal stability of lip movements in older adult speakers. *J. Speech Lang. Hear. Res.* 41, 41–50. doi: 10.1044/jslhr.4101.41
- Woollacott, M., Shumway-Cook, A., and Nashner, L. (1986). Aging and posture control: changes in sensory organization and muscular organization. *Int. J. Aging Hum. Dev.* 23, 97–114. doi: 10.2190/VXN3-N3RT-54JB-X16X
- Ziegler, J. C., and Muneaux, M. (2007). Orthographic facilitation and phonological inhibition in spoken word recognition: a developmental study. *Psychon. Bull. Rev.* 14, 75–80. doi: 10.3758/BF03194031
- Ziegler, J. C., Ferrand, L., and Montant, M. (2004). Visual phonology: the effects of orthographic consistency on different auditory word recognition tasks. *Mem. Cognit.* 32, 732–741. doi: 10.3758/BF03195863
- Ziegler, J. C., Muneaux, M., and Grainger, J. (2003). Neighborhood effects in auditory word recognition: phonological competition and orthographic facilitation. *J. Mem. Lang.* 48, 779–793. doi: 10.1016/S0749-596X(03)00006-8

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

Copyright © 2015 Saletta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The segment as the minimal planning unit in speech production and reading aloud: evidence and implications

Alan H. Kawamoto^{1*}, Qiang Liu¹ and Christopher T. Kello²

¹ Department of Psychology, University of California, Santa Cruz, Santa Cruz, CA, USA, ² Department of Cognitive Science, University of California, Merced, Merced, CA, USA

OPEN ACCESS

Edited by:

Simone Sulpizio,
University of Trento, Italy

Reviewed by:

Markus F. Damian,
University of Bristol, UK
Cristina Burani,
Institute of Cognitive Sciences and
Technologies, CNR, Italy

*Correspondence:

Alan H. Kawamoto,
Department of Psychology, University
of California, Santa Cruz, 1159 High
St., Santa Cruz, CA 95064, USA
ahk@ucsc.edu

Specialty section:

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

Received: 21 July 2015

Accepted: 11 September 2015

Published: 29 September 2015

Citation:

Kawamoto AH, Liu Q and Kello CT
(2015) The segment as the minimal
planning unit in speech production
and reading aloud: evidence and
implications. *Front. Psychol.* 6:1457.
doi: 10.3389/fpsyg.2015.01457

Speech production and reading aloud studies have much in common, especially the last stages involved in producing a response. We focus on the minimal planning unit (MPU) in articulation. Although most researchers now assume that the MPU is the syllable, we argue that it is at least as small as the segment based on negative response latencies (i.e., response initiation before presentation of the complete target) and longer initial segment durations in a reading aloud task where the initial segment is primed. We also discuss why such evidence was not found in earlier studies. Next, we rebut arguments that the segment cannot be the MPU by appealing to flexible planning scope whereby planning units of different sizes can be used due to individual differences, as well as stimulus and experimental design differences. We also discuss why negative response latencies do not arise in some situations and why anticipatory coarticulation does not preclude the segment MPU. Finally, we argue that the segment MPU is also important because it provides an alternative explanation of results implicated in the serial vs. parallel processing debate.

Keywords: absolute latency, segment duration, serial vs. parallel encoding

In reading aloud and speech production experiments, participants produce a single word utterance as the response and thus the last processing stages—phonological encoding (assigning a segment to a position in a metrical frame), phonetic encoding (retrieving the motor plans required for articulation), and articulation (producing the gestures leading to an acoustic response)—are shared. Moreover, the 2 fields became closer as speech production researchers began to use chronometric measures (Meyer, 1992) and word reading researchers began to use error measures (Kello and Plaut, 2000). Also, models integrating both fields were being proposed (Roelofs, 2004).

One aspect of processing common to both fields is the degree to which processing is incremental. Incremental processing can be manifested in two non-mutually exclusive ways: (1) a segment (i.e., a consonant or vowel) as the minimal planning unit (MPU) (Kawamoto et al., 1998; Kawamoto, 1999), and (2) cascaded processing (Kello et al., 2000; Rapp and Goldrick, 2000). In this review, we consider the MPU.

Some researchers argue that articulation cannot begin before the currently queued word or syllable is fully planned, while others contend that articulation can start with just one segment planned. We begin by reviewing a variety of phonological units that have been proposed as the MPU, but focus on the segment. Next, we summarize more evidence for the segment and rebut

arguments against the segment. Finally, we discuss how the segment MPU provides alternative interpretations of results relevant to the serial vs. parallel processing debate in reading aloud.

Possible MPUs

Levelt (1989) initially assumed that the MPU was the phonological word—a stress group that may include multiple words. Under this assumption, the phonological word is completely phonologically encoded before it is sent to the phonetic encoding stage. After all syllables have been phonetically encoded, the motor plan for the entire word is sent to the articulator¹.

However, most researchers now assume that the syllable is the MPU (Schriefers and Teruel, 1999; Meyer et al., 2003). Under this assumption, the initial syllable is phonetically encoded after it has been phonologically encoded and then executed by the articulators after the entire word has been phonologically encoded. For Levelt's (1989; Levelt et al., 1999a) model, the syllable plays a unique role because the motor plan determined at the phonetic encoding stage is based on the syllable (either retrieved from a mental syllabary or assembled from the motor plans of individual segments). Levelt's speech production model has been implemented as the WEAVER model (Levelt et al., 1999a).

Models of reading aloud have also been implemented as computational models (Coltheart et al., 2001; Kello and Plaut, 2003; Perry et al., 2007). Unlike speech production models, however, models of reading aloud focus on the mapping from the spelling to a phonological representation and typically base response latency predictions on the time to generate the phonological representation.

Segment as the MPU

Various subsyllabic units have been considered as MPUs including the initial consonant(s) or the initial plosive consonant and following vowel (Frederiksen and Kroll, 1976) and the segment, the unit we focus on in this review (MacKay, 1987; Dell et al., 1993; Kawamoto et al., 1998)². If the segment is the MPU, then the motor plan for the initial segment can be retrieved and executed as soon as it has been phonologically encoded. Segment motor plans are part of Levelt's (1989) model, but are not intended to be executed individually.

It is theoretically straightforward to show that the segment is the MPU if phonological or other processes can be shown to affect absolute response latencies and initial segment durations. However, it is methodologically difficult to do so because many initial segments produce little or no acoustic energy.

¹Dell et al. (1993) use a simple recurrent network to produce segments sequentially without a stage that encodes segments to slots in a phonological frame.

²Some models (e.g., Dell et al., 1993) have been implemented as simple recurrent connectionist networks that generate each segment (or its features) after a fixed number of processing cycles without any buffering.

Problems Due to Acoustic Characteristics of the Initial Segment

The biggest problem is that the initial part of plosive and affricate segments are silent. In fact, for plosives, acoustic energy is not generated until the end of the segment when the second segment begins. Because there is no acoustic energy throughout the entire plosive segment, acoustic latency (response latency based on acoustic onset) conflates response latency and initial segment duration. Moreover, matching the initial segment across conditions isn't a solution because any factor that affects initial segment duration affects acoustic latency.

The conflation of response latency and initial segment duration extends to voiceless affricates and fricatives if voice-keys are used because voice-keys typically miss the low intensity acoustic energy of these segments (Pechmann et al., 1989; Sakuma et al., 1997). In fact, the first 2 segments might be missed if the target begins with /s/ followed by a plosive (Sakuma et al., 1997; Rastle and Davis, 2002).

The problems with using acoustic energy to assess processing difficulty arise because the onset of acoustic energy is arguably the last event occurring during articulation. Two alternatives are to index response latency to the initiation of muscular activity using electromyography (Riès et al., 2012) or movement of speech articulators (lips and jaw) using video (Kawamoto et al., 2008). The latter was used in the experiment described below.

Negative Response Latencies

To demonstrate that the segment is the MPU, the initial segment of a monosyllabic target word was primed in a reading aloud task (Kawamoto et al., 2014, Expt. 2). The initial letter was followed by underscores and presented for either 300 or 600 ms, at which time the underscores were replaced by the remaining letters of the target. The segment MPU predicts that a response can be initiated before the complete target is presented—resulting in a *negative* response latency—but the syllable MPU does not. The results below are from the 600 ms condition where there is sufficient time for a response to be initiated. Using acoustic onset, 2.5% of the trials had negative latencies measured from onset of the complete target, all beginning with non-plosives. However, using articulatory onset based on movement of the lips and jaw, 26.2% of the trials had negative latencies and these trials included plosives and non-plosives. These negative latencies provide unequivocal evidence for the segment MPU because the initial segment was provided early and measurements were able to detect its initiation early in the course of articulation.

Initial Segment Duration Differences

Additional evidence for the segment MPU is acoustic durations of responses. Duration effects arise because articulation of the current unit is prolonged while the speaker prepares the following unit to be articulated. Processing effects can be manifested as duration effects in different ways (see Kello, 2004), including duration effects measured across the entire word (Damian, 2003)³. However, a duration effect localized to the

³If the effect is localized to the initial segment, measuring the duration of the entire word would miss the effect for targets beginning with plosives.

initial segment is the strongest evidence for the segment as the MPU (Kawamoto et al., 2014).

Although the duration of the initial segment can be measured directly from the acoustic response only for non-plosives, the duration effect for plosives can be determined indirectly. In particular, the *difference* in the duration of initial plosive segments, ISD_P , in different priming conditions corresponds to the difference in acoustic latency for plosive and non-plosive segments, AL_P and AL_N , respectively:

$$ISD_P'' - ISD_P' = (AL_P'' - AL_N'') - (AL_P' - AL_N'), \text{ and (1a)}$$

$$ISD_P'' - ISD_P' = (AL_P'' - AL_P') - (AL_N'' - AL_N'), \quad (1b)$$

where double prime and prime denote the 600 ms and 300 ms prime durations, respectively (see also Kawamoto et al., 1998; Kawamoto, 1999).

Alternatively, the initial plosive duration can be determined by using articulatory onset to approximate the beginning of the segment and acoustic onset as the end of the segment. Using both of these approaches, the duration of plosives was also shown to be longer in the 600 ms than in the 300 ms condition due to early initiation of articulation (Kawamoto et al., 2014).

Rebutting Evidence Against Segment MPU

There are many studies demonstrating that a planning unit larger than the initial segment is used for different stimuli under various experimental conditions. These units include the syllable (Cholin et al., 2011), the initial fragment up to and including the first stressed syllable (Sulpizio et al., 2015), the word (Meyer et al., 2003), two phonological words (Damian and Dumay, 2007), and even the clause (Ferreira and Swets, 2002). These results demonstrate that planning units are variable, and can be as large as the clause.

We argue that the segment remains viable as the MPU because the planning unit varies by individuals, as well as stimuli and experimental design. Two different scenarios can arise with a variable planning scope. In one scenario, an effect can be found assuming a smaller unit than the putative MPU. For example, Damian (2003) found longer word durations when the initial segment was primed as predicted by the segment but not the syllable MPU, but only when a deadline was imposed. In the other scenario, a smaller unit might yield no effect as predicted. For example, monosyllabic words are named as quickly as bisyllabic words when presented in the same block as predicted by syllable and segment MPUs (Meyer et al., 2003; Damian et al., 2010), but more quickly when presented in different blocks as predicted by the word MPU (Meyer et al., 2003). Therefore, the planning unit was ostensibly larger in some studies without finding any effect for the smaller MPU. We further note that a smaller MPU does not always predict shorter latencies; longer latencies can arise if there is competition between different initial syllables (e.g., in assigning stress, Sulpizio et al., 2015) or segments (e.g., in mapping a letter or letters to a phoneme as discussed below).

Another argument is that anticipatory coarticulation precludes the segment MPU because knowledge of upcoming segments is required during articulation and because it is ubiquitous (Levelt et al., 1999b; Rastle et al., 2000). However,

Kawamoto and Liu (2007) found that anticipatory coarticulation is not ubiquitous. They had participants utter one member of a minimal pair (still-stool, spill-spool, still-spill, or stool-spool) and found that there was anticipatory coarticulation of the vowel on the initial segment when the vowels were identical, but not when the vowels were different. Moreover, the long interval between articulatory onset and acoustic onset when the initial segment alone is primed (Kawamoto et al., 2014) can be interpreted as coarticulatory effects of the initial segment on the preceding null phoneme.

Implications of the Segment as the MPU

Determining that the segment is the MPU is important in its own right, but it is also important because it provides an alternative account of results in other debates such as whether phonological encoding is purely parallel or has a sequential component. We examine a length effect and a position effect, effects that would be considered straightforward for sequential reading models to account for (e.g., Perry et al., 2007). However, we argue that current sequential models cannot account for the entire pattern of results, but that purely parallel models can if the segment is the MPU and if acoustic characteristics of the initial segment are considered.

Onset Complexity Effect

Researchers have examined whether words with a simple onset consisting of a single consonant have shorter or longer naming latencies than words with a complex onset consisting of two or more consonants. An early study by Frederiksen and Kroll (1976) found that when length was controlled, words with simple onsets had shorter naming latencies than words with complex onsets. However, interpreting these results is complicated by two acoustic characteristics of simple and complex onsets. First, complex onsets in English can only begin with plosives or voiceless fricatives, and many complex onsets beginning with /s/ are followed by a plosive. Second, segments have a shorter duration in a complex onset than in a simple onset (Klatt, 1974; Rastle and Davis, 2002).

Kawamoto and Kello (1999) reexamined the onset complexity effect for monosyllabic targets beginning with /s/. (Fillers beginning with plosives were also included.) In one experiment the second consonant of the complex onset was a plosive, and in another it was a non-plosive. Using measures of acoustic latency based on marking digitized responses, they found that targets with complex onsets had *shorter* acoustic latencies than targets with simple onsets despite being *longer* in length. They hypothesized that the inconsistency in their results and Frederiksen and Kroll's (1976) results was due to how acoustic latency was determined. This hypothesis was confirmed by Rastle and Davis (2002) who replicated Kawamoto and Kello's (1999, Expt. 2) results when acoustic latency was based on hand-marking digitized responses, but who found no effect when an integrator voice-key was used, and an opposite effect when a simple voice-key was used (see **Table 1**).

Although the difference in results due to the method of measuring acoustic latency reported by Rastle and Davis has

TABLE 1 | Acoustic latencies of words with simple vs. complex onsets beginning with /s/ followed by a plosive determined in different ways.

Kawamoto and Kello (Expt. 2)

	KK	RD	RD	RD	CDP+	DRC
	digitized	digitized	integrator-VK	simple-VK		
Simple	462.4	371	449	500	89.8	68
Complex	445.6	362	447	511	94.7	70
Difference	16.8	9	2	-11	-4.9	-2

Results from Kawamoto and Kello (1999, Expt. 2) (labeled "KK") based on marking a digitized acoustic waveform, from Rastle and Davis (2002) (labeled "RD") based on hand-marking a digitized acoustic waveform and 2 different voice-key (VK) methods, and simulation results from the CDP+ model (Perry et al., 2007) and the DRC model (Coltheart et al., 2001).

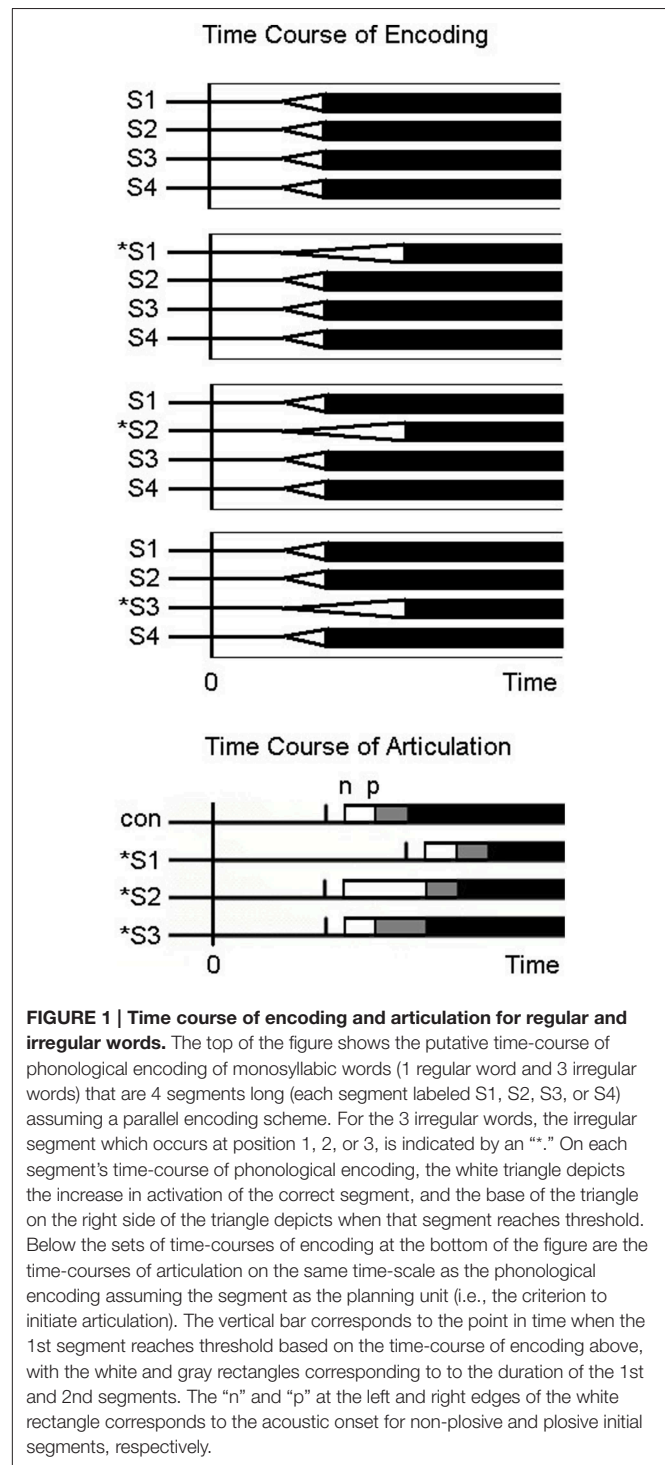
been widely recognized, the theoretical implication of the onset complexity effect has not. In particular, the partially sequential DRC (Coltheart et al., 2001) and CDP+ (Perry et al., 2007) models cannot account for the results (see Table 1) because the sequential rule route of these models processes the input from the beginning to the end of the word 1 letter or 1 grapheme at a time, respectively, independently of other letters and graphemes in the input. Thus, processing takes longer when the input has more letters and graphemes.

However, Kawamoto and Kello argued that the onset complexity result could be accounted for assuming parallel processing if the segment is the MPU. In particular, the initial consonant can be almost any consonant if the 2nd letter is a vowel as it is for simple onsets, but is almost always /s/ if the second segment of a complex onset is a plosive or a nasal consonant. If processing is parallel, the first segment is still being processed when the second segment is being processed. If information about the second segment can influence processing of the first segment, then the initial /s/ of a complex onset followed by a plosive or a nasal consonant should be encoded before the /s/ of a simple onset. Thus, for the segment MPU, articulation can be initiated earlier for targets with complex rather than simple onsets.

Regularity by Position of Regularity Interaction

Monosyllabic English words with irregular pronunciations have longer acoustic latencies than matched words with regular pronunciations. This regularity effect diminishes as the position of the irregular grapheme moves from left to right (Roberts et al., 2003). The authors argue that sequential models such as the DRC model could account for the data, but purely parallel models could not.

However, all the models considered by Roberts and colleagues assume that the MPU is the syllable (or word). Kawamoto et al. (1998) argued that purely parallel models could account for the regularity by position of regularity interaction if plosivity of the initial segment is taken into account and if the segment is the MPU. As illustrated in Figure 1, when the irregular grapheme is at position 1, targets beginning with plosives as well as non-plosives manifest the regularity effect because phonation cannot begin until the initial segment reaches threshold. When the



irregular grapheme is at position 2, the 2nd segment takes longer to reach threshold for irregular graphemes than for regular graphemes. However, acoustic latency is longer only for irregular targets beginning with plosives; no effect of regularity is predicted for targets beginning with non-plosives. This interaction of plosivity and regularity for targets with the irregular grapheme at position 2 has been found (Cortese, 1998; Kawamoto et al.,

1998). Finally, when the irregular grapheme is at position 3, only targets beginning with an /s/ followed by a plosive might manifest an effect, but only if a voice-key is used. On this account, the regularity effect diminishes from left to right because the proportion of stimuli that manifest an effect that can be detected acoustically diminishes from left to right. Roberts et al. (2003) rejected the account proposed by Kawamoto et al. (1998) because coarticulation was argued to be ubiquitous and thus the segment could not be the MPU. However, the coarticulation argument has been rebutted (see above). Moreover, Roberts and colleagues never provided an account of the regularity by plosivity interaction.

Cortese (1998) also reported simulations based on serial and parallel models showing that targets beginning with plosive as well as non-plosive targets predicted a regularity effect at position 2, but not the interaction. We argue that models fail to predict

the plosivity by regularity interaction at position 2 because the naming latency predictions assume that the MPU is the syllable (or word) and that the dependent measure is acoustic latency. If the MPU is the segment, sequential and parallel models should account for the plosivity by regularity interaction at position 2. Thus, the crucial distinction is not whether processing is serial or parallel, but whether the MPU is the segment or the syllable.

Final Remarks

The segment MPU suggests that written word processing can be highly incremental, with the degree of incrementality varying across individuals and with stimulus and task demands. More importantly, articulatory, and acoustic effects implied by the segment MPU also affect assumptions about earlier encoding stages.

References

- Cholin, J., Dell, G. S., and Levelt, W. J. M. (2011). Planning and articulation in incremental word production: syllable-frequency effects in English. *J. Exp. Psychol.* 37, 109–122. doi: 10.1037/a0021322
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual-route cascade model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256. doi: 10.1037/0033-295X.108.1.204
- Cortese, M. J. (1998). Revisiting serial position effects in reading. *J. Mem. Lang.* 39, 652–665. doi: 10.1006/jmla.1998.2603
- Damian, M. F. (2003). Articulatory duration in single-word speech production. *J. Exp. Psychol.* 29, 416–431. doi: 10.1037/0278-7393.29.3.416
- Damian, M. F., Bowers, J. S., Stadthagen-Gonzalez, H., and Spalek, K. (2010). Does word length affect speech onset latencies when producing single words? *J. Exp. Psychol.* 36, 892–205. doi: 10.1037/a0019446
- Damian, M. F., and Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *J. Mem. Lang.* 57, 195–209. doi: 10.1016/j.jml.2006.11.001
- Dell, G. S., Juliano, C., and Govindjee, A. (1993). A theory of frame constraints in phonological speech errors. *Cogn. Sci.* 17, 149–195. doi: 10.1207/s15516709cog1702_1
- Ferreira, F., and Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *J. Mem. Lang.* 46, 57–84. doi: 10.1006/jmla.2001.2797
- Frederiksen, J. R., and Kroll, J. F. (1976). Spelling and sound: approaches to the internal lexicon. *J. Exp. Psychol. Hum. Percept. Perform.* 2, 361–379. doi: 10.1037/0096-1523.2.3.361
- Kawamoto, A. H. (1999). Incremental encoding and incremental articulation in speech production: evidence based on response latency and initial segment duration. *Behav. Brain Sci.* 22, 48–49. doi: 10.1017/S0140525X99331770
- Kawamoto, A. H., and Kello, C. T. (1999). Effect of onset cluster complexity in speeded naming: a test of rule-based approaches. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 361–375. doi: 10.1037/0096-1523.25.2.361
- Kawamoto, A. H., Kello, C. T., Jones, R. M., and Bame, K. (1998). Initial phoneme versus whole word criterion to initiate pronunciation: evidence based on response latency and initial phoneme duration. *J. Exp. Psychol.* 24, 862–885. doi: 10.1037/0278-7393.24.4.862
- Kawamoto, A. H., and Liu, Q. (2007). “A unified account of duration and coarticulatory effects in speech production,” in *Proceedings of the 29th Annual Cognitive Science Society*, eds D. S. McNamara and G. Trafton (Austin, TX: Cognitive Science Society), 1151–1156.
- Kawamoto, A. H., Liu, Q., Lee, R. J., and Grebe, P. R. (2014). The segment as the minimal planning unit in speech production: evidence based on absolute response latencies. *Q. J. Exp. Psychol.* 67, 2340–2359. doi: 10.1080/17470218.2014.927892
- Kawamoto, A. H., Liu, Q., Mura, K., and Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *J. Mem. Lang.* 58, 347–365. doi: 10.1016/j.jml.2007.06.002
- Kello, C. T. (2004). Control over the time course of cognition in the tempo-naming task. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 942–955. doi: 10.1037/0096-1523.30.5.942
- Kello, C. T., and Plaut, D. C. (2000). Strategic control in word reading: evidence from speeded responding in the tempo-naming task. *J. Exp. Psychol.* 26, 719–750. doi: 10.1037/0278-7393.26.3.719
- Kello, C. T., and Plaut, D. C. (2003). Strategic control over rate of processing in word reading: a computational investigation. *J. Mem. Lang.* 48, 207–232. doi: 10.1016/S0749-596X(02)00512-0
- Kello, C. T., Plaut, D. C., and MacWhinney, B. (2000). The task-dependence of staged versus cascade processing: an empirical and computational study of Stroop interference in speech production. *J. Exp. Psychol.* 129, 340–360. doi: 10.1037/0096-3445.129.3.340
- Klatt, D. (1974). The duration of [s] in English words. *J. Speech Hear. Res.* 17, 51–63. doi: 10.1044/jshr.1701.51
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999a). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/S0140525X99001776
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999b). Multiple perspectives on word production. *Behav. Brain Sci.* 22, 61–75. doi: 10.1017/S0140525X99451775
- MacKay, D. G. (1987). *The Organization of Perception and Action: A Theory of Action and Other Cognitive Skills*. New York, NY: Springer-Verlag.
- Meyer, A. (1992). Investigation of phonological encoding through speech error analyses: achievements, limitations, and alternatives. *Cognition* 42, 181–211. doi: 10.1016/0010-0277(92)90043-H
- Meyer, A., Roelofs, A., and Levelt, W. J. M. (2003). Word length effects in object naming: the role of a response criterion. *J. Mem. Lang.* 48, 131–147. doi: 10.1016/S0749-596X(02)00509-0
- Pechmann, T., Reetz, H., and Zerbst, D. (1989). Kritik einer Meßmethode: Zur Ungenauigkeit von Voice-key Messungen [Critique of a method of measurement: on the unreliability of voice-key measurements.]. *Sprach. Kognit.* 8, 65–71.
- Perry, C., Ziegler, J. C., and Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychol. Rev.* 114, 273–315. doi: 10.1037/0033-295X.114.2.273
- Rapp, B., and Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychol. Rev.* 107, 460–499. doi: 10.1037/0033-295X.107.3.460
- Rastle, K., and Davis, M. H. (2002). On the complexities of measuring naming. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 307–314. doi: 10.1037/0096-1523.28.2.307

- Rastle, K., Harrington, J., Coltheart, M., and Palethorpe, S. (2000). Reading aloud begins when the computation of phonology is complete. *J. Exp. Psychol.* 26, 1178–1191. doi: 10.1037/0096-1523.26.3.1178
- Riès, S., Legoux, T., Burle, B., Alario, F., and Malfait, N. (2012). Why does picture naming take longer than word reading? The contribution of articulatory processes. *Psychon. Bull. Rev.* 19, 955–961. doi: 10.3758/s13423-012-0287-x
- Roberts, M. A., Rastle, K., Coltheart, M., and Besner, D. (2003). When parallel processing in visual word recognition is not enough: new evidence from naming. *Psychon. Bull. Rev.* 10, 405–414. doi: 10.3758/BF03196499
- Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Mem. Cogn.* 32, 212–222. doi: 10.3758/BF03196853
- Sakuma, N., Fushimi, T., and Tatsumi, I. (1997). Onseiha no shisatsu ni yoru kana no ondokusenji no sokutei: ondokusenji wa gotton no choonho ni ori okiku kotonaru [Naming latency measurements of kana based on inspection of voice waveforms: naming latency varies greatly depending on the manner of articulation of the initial phoneme.]. *Shinkeishinrigaku* 13, 126–136.
- Schriefers, H., and Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *Eur. J. Cogn. Psychol.* 11, 17–50. doi: 10.1080/713752301
- Sulpizio, S., Spinelli, G., and Burani, C. (2015). Stress affects articulatory planning in reading aloud. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 453–461. doi: 10.1037/a0038714
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Kawamoto, Liu and Kello. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Lexical frequency effects on articulation: a comparison of picture naming and reading aloud

Petroula Mousikou* and Kathleen Rastle

Department of Psychology, Royal Holloway, University of London, Egham, UK

OPEN ACCESS

Edited by:

Simone Sulpizio,
University of Trento, Italy

Reviewed by:

Niels O. Schiller,
Leiden University, Netherlands
Marc Brysbaert,
Ghent University, Belgium

*Correspondence:

Petroula Mousikou
betty.mousikou@rhul.ac.uk

Specialty section:

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

Received: 26 July 2015

Accepted: 28 September 2015

Published: 15 October 2015

Citation:

Mousikou P and Rastle K (2015)
Lexical frequency effects on
articulation: a comparison of picture
naming and reading aloud.
Front. Psychol. 6:1571.
doi: 10.3389/fpsyg.2015.01571

The present study investigated whether lexical frequency, a variable that is known to affect the time taken to utter a verbal response, may also influence articulation. Pairs of words that differed in terms of their relative frequency, but were matched on their onset, vowel, and number of phonemes (e.g., *map* vs. *mat*, where the former is more frequent than the latter) were used in a picture naming and a reading aloud task. Low-frequency items yielded slower response latencies than high-frequency items in both tasks, with the frequency effect being significantly larger in picture naming compared to reading aloud. Also, initial-phoneme durations were longer for low-frequency items than for high-frequency items. The frequency effect on initial-phoneme durations was slightly more prominent in picture naming than in reading aloud, yet its size was very small, thus preventing us from concluding that lexical frequency exerts an influence on articulation. Additionally, initial-phoneme and whole-word durations were significantly longer in reading aloud compared to picture naming. We discuss our findings in the context of current theories of reading aloud and speech production, and the approaches they adopt in relation to the nature of information flow (staged vs. cascaded) between cognitive and articulatory levels of processing.

Keywords: speech production, reading aloud, picture naming, articulation, acoustics, reaction times

INTRODUCTION

Speech production involves the combination of cognitive and articulatory processes. However, these processes have been traditionally investigated in separate domains of research, yielding a division between models of speech production that focus on psycholinguistic (e.g., Dell, 1986; Levelt et al., 1999) vs. motor control (Guenther et al., 2006) aspects of this process. This division is likely due to the widely held assumption that the transition from cognitive to articulatory levels of processing occurs in a *staged* manner, so that articulatory processes can only be initiated after cognitive processing is complete (Levelt et al., 1999). On this assumption, the articulation of an utterance should be unaffected by higher-level cognitive processes that are involved in selecting an abstract phonological code for speech production. However, several studies to date have shown that articulation is affected systematically by such higher-level processes (see Bell et al., 2009, and Gahl et al., 2012, for comprehensive reviews). The results of these studies suggest that articulation can be initiated before higher-level processes involved in the selection of a phonological code are finished. This finding offers support for the view that information from cognitive to articulatory levels of processing flows in a *cascaded* manner.

More specifically, these studies used different approaches to investigating whether high-level cognitive processes cascade down to articulation. Such approaches involved examining the nature of *speech errors* and showing that erroneous productions contain articulatory features of the non-produced target sound. This finding was thought to indicate that partial information from the target sound can cascade into articulation. Other approaches involved examining how certain *lexical variables* such as word frequency and phonological neighborhood density may influence articulatory detail, or how *syntactic predictability* and *semantic congruency* may affect articulation. Each of the four approaches is further elaborated below.

Speech Errors

Speech errors induced in the laboratory often reflect the simultaneous production of competing sounds. Goldrick and Blumstein (2006) designed a tongue twister task in which participants had to repeat a sequence of syllables (e.g., *keff geff keff*) at a rate faster than normal speech. They found that when participants erroneously produced /g/, there were phonetic traces of the target sound /k/. In these instances, /g/ had a longer Voice Onset Time (VOT) (i.e., it was more /k/-like) than correctly produced /g/ sounds. This finding shows that partial activation of both the target sound and the competing sound is reflected in the articulation of the spoken output. Hence, unselected phonemic representations can influence articulatory detail.

Similarly, in a study that used Electromagnetic Articulography (EMA) participants were asked to repeat as quickly as possible the phrase *top cop* (Goldstein et al., 2007). The results from this study showed that the articulatory gestures associated with the sounds /t/ and /k/ (i.e., raising of the tongue tip and the tongue body) were produced simultaneously. Similarly, Pouplier (2007) asked participants to read silently word pairs with a specific consonant order in their onsets (e.g., *gap dupe*, *gob dub*, *gum dam*) before they were unexpectedly asked to pronounce a word pair with opposite consonant order (*dome gimp*). Participants' productions revealed that the tongue tip was high during the initial /d/ in *dome*, but the tongue dorsum also displayed unexpected raising, which is characteristic of the articulatory gesture associated with /g/. Taken together, these results indicate that the partial activation of competing sounds cascades down to articulation.

Last, using a tongue-twister paradigm, McMillan and Corley (2010) asked participants to read groups of four ABBA syllables, where A and B differed in the onset by a voice feature (e.g., *kef gef gef kef*); a place feature (e.g., *kef tef tef kef*); both voice and place features (e.g., *kef def def kef*); or were identical (e.g., *kef kef kef kef*). Participants' responses were not categorized as "correct" or "wrong"; instead, the variability in the articulation of participants' *kef* productions during error invoking conditions (e.g., *kef gef gef kef*, *kef tef tef kef*, *kef def def kef*) were investigated relative to the baseline condition (e.g., *kef kef kef kef*). The results from this study showed significantly more articulatory variability in the VOT productions when the onsets of the A and B syllables differed in voice only (e.g., *kef gef gef kef*). There was also more articulatory variability in lingual contact with the palate, measured with Electropalatography (EPG), when the onsets of

the A and B syllables differed only in place of articulation (e.g., *kef tef tef kef*). Moreover, articulatory variability of both VOT and location of palate contact was significantly smaller when the onsets of the A and B syllables differed in both place and voice (e.g., *kef def def kef*). The results from this study provide further evidence in favor of the idea that properties of phonologically-similar competing utterances cascade into articulation.

Lexical Effects

High-frequency (HF) words are typically produced with shorter durations, reduced vowels, deleted codas, more tapping and palatalization, and reduced pitch range, compared to low-frequency (LF) words (e.g., Zipf, 1929; Fidelholz, 1975; Hooper, 1976; Rhodes, 1992, 1996; Fosler-Lussier and Morgan, 1999; Kawamoto et al., 1999; Bybee, 2000; Munson and Solomon, 2004; Pluymaekers et al., 2005a; Aylett and Turk, 2006; Gahl, 2008). For example, Pluymaekers et al. (2005a) used data from a corpus of spontaneous speech in Dutch to examine the production of the same affixes appearing in different words that varied in frequency. They observed that suffixes belonging to HF words were more reduced than those belonging to LF words. Using data from the Switchboard corpus of American English telephone conversations, Gahl (2008) also reported that HF English homophones (e.g., *time*) were produced with shorter durations than their LF counterparts (e.g., *thyme*). Accordingly, in a reading aloud task, Munson and Solomon (2004) observed that vowels in LF words were produced with longer durations and closer to the periphery of the vowel space (hence, with more extreme articulation) than vowels in HF words. Initial-phoneme durations were also found to be longer for LF words in reading aloud (Kawamoto et al., 1999), which led the authors to conclude that the criterion to initiate pronunciation is based on the initial phoneme and not on the whole word. This finding challenges the assumption that articulation is initiated only after phonological encoding is complete (Levelt et al., 1999). Taken together, these results suggest that *lexical frequency*, a variable that has been traditionally known to affect high-level cognitive processes, also affects low-level articulatory processes.

Words from dense neighborhoods (i.e., words which are phonologically similar with several other words) are hyperarticulated in reading aloud, compared to words from sparse neighborhoods (Wright, 1997, 2004; Munson and Solomon, 2004; Munson, 2007; but see Gahl et al., 2012, who observed that words from dense neighborhoods were phonetically reduced in spontaneous speech). In particular, in these studies, vowels in words from high-density neighborhoods were produced closer to the periphery of the vowel space (hence, with extreme articulation), whereas vowels in words from sparse neighborhoods were produced closer to the center of the vowel space. Accordingly, Baese-Berk and Goldrick (2009) observed that words with minimal pair onset neighbors (e.g., *cod-god*) were produced with more extreme VOTs (hence, were more hyperarticulated) than words with no minimal pair onset neighbors (e.g., *cop-gop*, where *gop* is a non-word). Last, Scarborough (2004) found that vowels in LF words from high-density neighborhoods were more coarticulated than vowels in HF words from low-density neighborhoods. Although

this finding seems to contradict previous findings, Scarborough (2004) took this result to indicate that speakers coarticulate the vowels more in words that are harder for listeners to recognize in order to facilitate lexical access (Luce and Pisoni, 1998). Taken together, these findings suggest that similarly to lexical frequency, *phonological neighborhood density* influences articulation.

Syntactic Predictability Effects

Words that are predictable in a sentence are produced with shorter durations and more reduced vowels (e.g., Lieberman, 1963; Liu et al., 1997; Griffin and Bock, 1998; Krug, 1998; Bybee and Scheibman, 1999; Gregory et al., 1999; Jurafsky et al., 2001; Aylett and Turk, 2004; Pluymaekers et al., 2005b). Further, repeated words (e.g., words that have occurred in a previous sentence) are more predictable, thus they tend to be shortened (e.g., Fowler and Housum, 1987; Fowler, 1988; Hawkins and Warren, 1994; Bard et al., 2000). Finally, words in less probable syntactic constructions are produced with longer durations (e.g., Gahl and Garnsey, 2004; Gahl et al., 2006; Tily et al., 2009). Taken together, these findings suggest that *syntactic predictability* influences articulatory detail.

Semantic Congruency Effects

Balota et al. (1989) observed that words that were cued by semantically congruent primes (e.g., *dog* preceded by *cat*) were produced with shorter durations compared to when these words were cued by semantically incongruent primes (e.g., *pen*). Using the Stroop paradigm, Kello et al. (2000) asked participants to name the color of rectangles with superimposed distractor words that were either semantically congruent, incongruent, or neutral (i.e., if the rectangle was colored in red the congruent condition consisted of the superimposed word *red*; the incongruent condition consisted of the superimposed word *blue*, and the neutral condition consisted of the superimposed letter string *iiii*). The results from this study showed a Stroop interference effect, so that the incongruent condition yielded significantly slower color-naming latencies compared to the neutral condition. In addition, when participants had a deadline within which they had to respond, color naming durations were significantly longer in the incongruent condition relative to the neutral condition. These findings support the idea that *semantic congruency*, another variable that is thought to affect high-level cognitive processes, also influences articulation.

However, the empirical evidence in this research domain is not entirely consistent. Meyer (1990), for example, observed that single words were produced faster when they occurred in a phonologically similar context, yet their durations were unaffected by the context in which they occurred. Similarly, Schriefers and Teruel (1999) found that naming latencies of adjective-noun utterances (e.g., *red house*) were affected by distractor words that were phonologically related to the adjective, yet the durations of either the adjectives or the nouns were unaffected by the same experimental manipulation. Moreover, using three different speech production paradigms, a picture-word interference task with semantic and phonological relatedness between pictures and distractors, a picture-naming task in which pictures were blocked either by semantic category

or by word-initial overlap, and a Stroop task, such as that used by Kello et al. (2000), Damian (2003) found no evidence for the idea that central cognitive processes influence articulation once a response has been initiated. As such, he argued that “articulation is not affected by prior processing stages—a finding that is easily accommodated by theoretical approaches that clearly separate articulation from preceding stages” (Damian, 2003, p. 429).

More recently, Riès et al. (2012, 2014) sought to determine the reason why naming pictures takes longer than reading aloud words. According to the literature in this domain, this is so because access to semantic information, which is required in picture naming but not necessarily in reading aloud, is time-consuming (Theios and Amrhein, 1989). In addition, it has been suggested that the stimulus-response association is equivocal in picture naming (i.e., some pictures may receive more than one name) but not in word reading aloud, thus yielding response uncertainty in the former task but not in the latter (Ferrand, 1999). These explanations imply that the response latency differences observed in the two tasks are due to differences in the processes that are involved in word-selection in the two tasks. However, verbal response latencies reflect not only the time that is required to select a word, but also the time to plan and initiate articulation. As such, the response latency differences observed in the two tasks could be due to a delay in planning and initiating articulation in picture naming compared to word reading aloud. If this hypothesis is true, strong evidence will be provided in favor of the idea that task-inherent cognitive processes (e.g., activation of semantic information, response uncertainty) cascade into articulation. Riès et al. (2012, 2014) tested this hypothesis using a reaction-time (RT) fractionation procedure in a reading aloud and a picture-naming task. RT was defined as the delay between stimulus presentation and the onset of the verbal response. Electromyographic (EMG) activity from several lip muscles was also recorded. The stimulus-response (SR) interval was divided into a premotor interval (from stimulus onset to EMG activity) and a motor interval (from EMG activity to verbal response). The results from the Riès et al. (2014) study showed that the difference between picture naming and reading aloud times was due to the premotor interval. This finding is consistent with Damian's (2003) results falsifying the theory that high-level cognitive processes affect articulatory processes.

In the present study, we re-examined this idea. In particular, we investigated whether lexical frequency affects initial-phoneme durations in picture naming and reading aloud. Lexical frequency is known to affect the time taken to select a phonological code for production. However, if it also influences durational aspects of the verbal response, we can conclude that cognitive processing is taking place after the verbal response is initiated. Such a finding will imply that information from cognitive to articulatory levels of processing flows in a cascaded manner. In contrast, if lexical frequency does not have an effect on durational aspects of the verbal response, we can conclude that processing at high cognitive levels is completed before the verbal response is initiated, and so the nature of information flow between cognitive and articulatory levels of processing must be staged. On the basis of previous results in the literature, we predicted that LF items would yield longer initial-phoneme durations than HF items.

In addition, we examined effects of lexical frequency on response times. Based on previous findings, we predicted that LF items would yield slower response times than HF items. Furthermore, we hypothesized that lexical frequency effects on verbal responses should be more prominent in picture naming than in reading aloud. This is because semantic activation of the target stimulus is required in picture naming; hence, its associated lexical frequency will have a robust effect on verbal responses. In contrast, reading aloud of a printed word can be performed, in principle, on the basis of sublexical information, and so lexical frequency effects on verbal responses are likely to be attenuated in this task. Accordingly, we predicted that both in the reaction time analyses and the analyses of initial-phoneme durations, the frequency effect would be bigger in size in picture naming than in reading aloud.

Last, we examined task effects on whole-word durations. Hennessey and Kirsner (1999) found that the same words were produced with longer durations in reading aloud compared to picture naming. They posited that reading aloud may be initiated on the basis of sublexical information (e.g., initial phoneme), and so processing of the rest of the word must be carried out during response execution, thus elongating response durations in this task compared to picture naming (see also Damian, 2003, and Kawamoto et al., 1998, for a similar account). Yet, this explanation is at odds with the idea that reading aloud begins when the computation of phonology is complete (Rastle et al., 2000). The present study further allows us to test these opposing views.

A common assumption in one of the most prominent psycholinguistic models of speech production (e.g., Levelt et al., 1999) is that the transition from cognitive to articulatory levels of processing during speech occurs in a staged (rather than a cascaded) manner, and so articulation can only be initiated after cognitive processing is complete. Similarly, the most prominent models of single word reading aloud (e.g., the DRC model of Coltheart et al., 2001; the CDP+ model of Perry et al., 2007; and the PDP model of Plaut et al., 1996) make the assumption that reading aloud cannot be initiated unless the orthography-to-phonology conversion of the printed letter string is complete. Thus, the results from the present study are critical for the evaluation of extant theories of speech production and reading aloud.

METHOD

Participants

Sixty undergraduate students from Royal Holloway, University of London, were paid £5 to participate in the study. Thirty of them participated in the picture naming task and the other 30 participated in the reading aloud task. Participants were monolingual native speakers of Southern British English and reported no visual, reading, or language difficulties.

Materials

In order to make the picture naming and reading aloud tasks as comparable as possible the same items were used in both

tasks. The selected items ($N = 72$) were between three and six letters long, had three or four phonemes, and had a CVC or CCVC structure. They were all regular words (i.e., with consistent spelling-to-sound mappings) that could be depicted as concrete objects.

The 72 items comprised 36 pairs of words that differed in their relative frequency, but were matched on number of phonemes and shared the same onset and vowel (e.g., *map* vs. *mat* and *brain* vs. *braid*, where *map* and *brain* are more frequent than *mat* and *braid*, respectively). Matching these pairs of words on their onset and vowel was important insofar as frequency effects on articulation were measured in terms of initial-phoneme durations, which are known to vary as a function of the identity of the following vowel or consonant (Klatt, 1975). Two lists were created using these word pairs, with one list containing items that were significantly higher in frequency than the items in the other list [$t_{(35)} = 8.27, p < 0.001$]¹. Age of acquisition (AoA) is known to have a robust effect on picture naming latencies that is independent of the frequency effect (see Bates et al., 2001; Meschyan and Hernandez, 2002). For this reason, we ensured that the items in the HF list had significantly lower AoA than the items in the LF list [$t_{(35)} = -4.42, p < 0.001$]. AoA values were obtained from Kuperman et al. (2012). The two lists were additionally matched on orthographic neighborhood, which was measured in terms of total orthographic neighbors [$t_{(35)} = -1.02, p > 0.05$] and substitution orthographic neighbors [$t_{(35)} = -1.57, p > 0.05$]; and phonological neighborhood, which was also measured in terms of total phonological neighbors [$t_{(35)} = 0.45, p > 0.05$] and substitution phonological neighbors [$t_{(35)} = -0.22, p > 0.05$]. The orthographic and phonological neighborhood information was extracted from the CLEARPOND database (Marian et al., 2012). The means of each of the linguistic variables for the HF and LF items are presented in Table 1. The paired words are provided as Supplementary Material.

TABLE 1 | Characteristics of the items used in the picture naming and reading aloud tasks.

	High frequency	Low frequency
SUBTLEX-UK log frequency (Zipf scale)	4.7	3.8
SUBTLEX-US frequency (per million)	60.6	10.0
Age of Acquisition	4.5	6.1
Number of letters	4.1	4.0
Number of phonemes	3.3	3.3
Orthographic neighbors (total)	14.5	15.6
Orthographic neighbors (substitution)	10.4	11.8
Phonological neighbors (total)	28.6	27.9
Phonological neighbors (substitution)	22.8	23.3

¹Frequency values were obtained from SUBTLEX-UK (Van Heuven et al., 2014). These values are expressed on a Zipf scale. Values 1–3 correspond to LF words and values 4–7 correspond to HF words. We also obtained frequency values from SUBTLEX-US (Brysbaert and New, 2009). These values are expressed per million. According to the frequency values obtained from SUBTLEX-US, the items in the HF list were also significantly higher in frequency than the items in the LF list [$t_{(35)} = 8.05, p < 0.001$].

The 72 pictures consisted of black-and-white line drawings of common objects. Most pictures were selected from the IPNP (International Picture Naming Project) database (Szekely et al., 2004) and the remaining were obtained from different sources, yet they were all comparable in style². The pictures varied slightly in width (226–400 pixels) and height (144–400 pixels) to avoid distorting the original shape of the depicted object; however, the longest side of each picture never exceeded 400 pixels and all pictures appeared in the center of the screen.

Design

In the picture naming task, each participant underwent a training phase and a test phase. The training phase consisted of two parts. During the first part, participants were told that the aim of this first training phase was to become familiar with the names of a set of pictures that they would be asked to name later. On each trial, participants saw a picture appearing on the computer screen and heard its corresponding name via headphones. The names of the pictures had been recorded by a female native speaker of Southern British English. Participants studied each picture for as long as they needed, and controlled the time at which the next picture was presented with a button press. The 72 pictures were presented to each participant in a different random order. During the second part of the training phase, we assessed whether participants remembered the picture names they had just learnt. Pictures were presented visually again in a random order and participants were asked to provide their names. Independently of whether participants produced the picture name correctly or incorrectly, on-screen feedback was provided subsequent to their response (i.e., the words “correct” or “incorrect” were displayed on the screen accordingly), and the correct picture name was presented aurally via headphones. Once the second part of the training phase was completed, participants proceeded to the test phase.

In the reading aloud task, there was no training phase. However, 16 words that had similar characteristics as the experimental words served as practice trials. A total of 72 experimental words were then presented to each participant in a different random order.

Apparatus and Procedure

Participants were tested individually in a quiet room, seated approximately 40 cm in front of a CRT monitor. Stimulus presentation and data recording were controlled by DMDX software (Forster and Forster, 2003). Verbal responses were recorded by a head-worn microphone. In the picture naming task, participants were told that they would see the same pictures that they had previously been familiarized with and that their task was to name each picture as quickly and as accurately as possible, without hesitation. The pictures appeared on a white background in the center of the screen and remained there for 2000 ms. The 72

pictures were presented to each participant in a different random order.

In the reading aloud task, participants were told that they would be shown a series of words and that their task was to read aloud each word as quickly and as accurately as possible, without hesitation. The words were presented in lowercase letters (14-point Courier New font) and appeared in black on a white background in the center of the screen for 2000 ms. Following 16 practice trials, the 72 words were presented to each participant in a different random order.

RESULTS

Participants' reaction times (RTs) in both the picture naming and reading aloud tasks were hand-marked using CheckVocal (Protopapas, 2007). Incorrect responses, mispronunciations, and hesitations (2.3% of the data in the picture naming task and 0.6% of the data in the reading aloud task) were treated as errors and discarded. Initial-phoneme durations and whole-word durations were measured using Praat (Boersma, 2001). Due to microphone clipping and mobile interference, 5.3% of the data in the picture naming task and 2% of the data in the reading aloud task could not be properly labeled and were therefore discarded. The hand-marking of participants' RTs and the acoustic labeling of initial-phoneme and whole-word durations were both performed by an independently trained rater who was naïve to the purposes of the experiment. The picture naming and reading aloud data were initially combined in a single analysis.

Reaction Times

To control for temporal dependencies between successive trials, the RT of the previous trial was taken into account in the analyses, so trials whose previous trial corresponded to an error and participants' first trial in each task (2.6% of all data) were excluded. The analyses were performed using linear mixed effects models (Baayen, 2008; Baayen et al., 2008) and the languageR (Baayen, 2008), lme4 1.0-5 (Bates et al., 2013), MASS (Venables and Ripley, 2002), and lmerTest (Kuznetsova et al., 2013) packages implemented in R (R Core Team, 2014, version 3.1.2).

The Box-Cox procedure indicated that inverse RT (1/RT) was the optimal transformation to meet the precondition of normality. We then multiplied 1/RT by -1000 ($-1000/\text{RT}$) to maintain the direction of effects, so that a larger inverse RT meant a slower response. In our model, inverse RT was the dependent variable. The fixed effects included the interaction between frequency type (HF vs. LF) and task type (picture naming vs. reading aloud), AoA, the RT of the previous trial, and trial order. The frequency type factor and the task type factor were both deviation-contrast coded (-0.5 , 0.5) to reflect the factorial design. Intercepts for subjects and items were included as random effects.

The results (obtained from 3990 observations) indicated a significant main effect of frequency: LF items were named slower than HF items ($t = 6.80$, $p < 0.001$). There was also a significant main effect of task: RTs were significantly faster in reading aloud compared to picture naming ($t = -16.40$, $p < 0.001$). The effect

²It is worth noting that due to an oversight, the American names of two of the objects were used in the study (*robe* instead of *gown* and *pants* instead of *trousers*). However, given that a training phase preceded the test phase, participants were already familiarized with the names of these two objects before carrying out the task.

of AoA was also significant ($t = 3.89, p < 0.001$), and so were the effects of the RT of the previous trial ($t = 3.49, p < 0.001$) and trial order ($t = 8.57, p < 0.001$). Importantly, frequency type interacted with task type ($t = -4.49, p < 0.001$), as the size of the frequency effect was significantly larger in picture naming compared to reading aloud (55 vs. 12 ms).

The picture naming and reading aloud tasks were then analyzed separately. In the analysis of the picture naming data, the fixed effects included frequency type (HF vs. LF), AoA, the RT of the previous trial, and trial order. Intercepts for subjects and items, and random slopes for the effect of frequency (for subjects) were included as random effects³. The results from this analysis (obtained from 1929 observations) showed a significant frequency effect, with LF items named slower than HF items ($t = 5.03, p < 0.001$), a significant effect of AoA ($t = 4.57, p < 0.001$), and a significant effect of trial order ($t = 5.86, p < 0.001$). The effect of the RT of the previous trial was not significant ($t = 1.40, p > 0.05$). In the analysis of the reading aloud data, frequency type (HF vs. LF), AoA, the RT of the previous trial, and trial order were included as fixed effects, and intercepts for subjects and items were included as random effects. The results (obtained from 2061 observations) showed a significant frequency effect with LF items read aloud slower than HF items ($t = 3.51, p < 0.001$), a significant effect of trial order ($t = 6.47, p < 0.001$), and a significant effect of the RT of the previous trial ($t = 7.72, p < 0.001$). The effect of AoA was not significant ($t = 1.02, p > 0.05$). The mean RTs for HF and LF items in the picture naming and reading aloud tasks are shown in **Table 2**.

Initial-phoneme Durations

The rater labeled the acoustic boundaries of the initial phoneme in each word via visual inspection of the waveform and spectrogram using the criteria established in the ANDOSL database (Croot et al., 1992). The analyses of the initial-phoneme durations were performed using the same version of R and the same R packages as those used in the analyses of the RT data. The Box-Cox procedure indicated that the logarithmic transformation was the best transformation for initial-phoneme durations to approach a normal distribution. Therefore, the logarithmic transformation of initial-phoneme duration was the dependent variable, while the fixed effects included the interaction between frequency type (HF vs. LF) and task type

(picture naming vs. reading aloud). The frequency type factor and the task type factor were both deviation-contrast coded ($-0.5, 0.5$) to reflect the factorial design. Intercepts for subjects and items were included as random effects.

The results (obtained from 4098 observations) showed a frequency effect, with LF items yielding longer initial-phoneme durations than HF items. However, this effect only approached significance ($t = 1.83, p = 0.07$). The main effect of task was significant: initial-phoneme durations were significantly longer in reading aloud compared to picture naming ($t = 2.1, p < 0.05$). Importantly, frequency type did not interact with task type ($t = -1.04, p > 0.05$). As in the RT analyses, the initial-phoneme durations in the picture naming and reading aloud tasks were subsequently analyzed separately. The analyses of the picture naming data (based on 1996 observations) showed a significant frequency effect ($t = 2.0, p < 0.05$), with LF items yielding significantly longer initial-phoneme durations than HF items. However, the analyses of the reading aloud task (based on 2102 observations) failed to show a significant frequency effect ($t = 0.57, p > 0.05$). The mean initial-phoneme durations for HF and LF items in the picture naming and reading aloud tasks are shown in **Table 2**.

Whole-word Durations

The same rater labeled the two acoustic boundaries that defined word duration. These were placed at the onset of acoustic energy, which was similarly denoted in all speech sounds by an increase in amplitude on the waveform, and at the offset of acoustic energy. When the last sound of the word was a stop, the second acoustic boundary that marked the end of the word was placed at the end point of the stop closure. Frequency effects on whole-word durations could not be examined given that the paired items in the HF and LF lists contained different codas. Therefore, in this analysis, we examined task effects (picture naming vs. reading aloud) on whole-word duration.

The analysis was performed using the same version of R and the same R packages as those used in the analyses of the RT and initial-phoneme duration data. The Box-Cox procedure indicated that the logarithmic transformation was the best transformation for the whole-word duration data. As such, the dependent variable in this analysis was the logarithmic transformation of whole-word duration, while task type (picture naming vs. reading aloud) was included as a fixed effect and intercepts for subjects and items were the random effects. The results (obtained from 4098 observations) showed a significant effect of task: whole-word durations were significantly longer in reading aloud compared to picture naming ($t = 3.42, p < 0.01$). The mean whole-word durations for all items in the picture naming and reading aloud tasks are shown in **Table 2**.

GENERAL DISCUSSION

Uttering a verbal response involves the combination of cognitive and articulatory processes; however, such processes have been traditionally investigated separately, perhaps due to the widely-held assumption that the relationship between cognitive and articulatory levels of processing is staged, so that articulation

TABLE 2 | Mean reaction times (RTs), initial-phoneme durations (IP durations), whole-word durations (WW durations), and Frequency effect (in milliseconds) in the picture naming and reading aloud tasks.

	RTs			IP durations			WW durations
	HF	LF	Freq effect	HF	LF	Freq effect	
Picture naming	664	719	55	58	60	2	358
Reading aloud	467	479	12	65	66	1	406

³The more complex model that included random slopes for the effect of frequency for subjects had a significantly better fit [$\chi^2(2) = 5.99, \text{Pr}(> \text{Chisq}) = 0.05$], hence this model was preferred over the simpler model.

can only begin once a phonological code has been generated (Levelt et al., 1999; Coltheart et al., 2001). A number of studies have provided evidence that challenges this assumption. Such evidence comes from speech errors, which contain articulatory characteristics of unselected sounds; and from effects of lexical frequency, phonological neighborhood density, syntactic predictability, and semantic congruency on the acoustic realization of verbal responses. Yet the evidence in this domain is not entirely consistent.

In the present study, we investigated effects of lexical frequency on articulation using the same stimuli in a picture naming and a reading aloud task. We reasoned that if lexical frequency affects durational aspects of verbal responses (e.g., initial-phoneme duration), we can conclude that cognitive processing continues to occur after the initiation of articulation. Such an observation would support the view that information from cognitive to articulatory levels of processing flows in a cascaded rather than a staged manner. In addition, we hypothesized that in a conceptually driven task such as picture naming, lexical frequency effects on articulation would be more prominent than in reading aloud. This is because semantic activation of the target stimulus is required in picture naming, and so its associated lexical variables (e.g., word frequency) are likely to cascade down to articulation (on the assumption that there is “leakage” of lexical activation from cognitive to articulatory levels of processing). However, reading aloud can be performed, in principle, on the basis of sublexical information, and so lexical variables associated with the printed word (e.g., its frequency) are less likely to trickle down to articulatory levels of processing.

Even though the analyses of RTs were not the focus of the present research, it is worth noting that the results were as expected. In particular, we observed a robust frequency effect, so that LF items were named slower than HF items. This was the case for both picture naming and reading aloud. Interestingly, the size of the frequency effect was significantly bigger in picture naming compared to reading aloud (55 vs. 12 ms). This result is consistent with the hypothesis that in conceptually driven tasks, where there is necessarily semantic activation of the target item, lexical variables associated with the target (e.g., word frequency) may have a robust effect on verbal responses (be that an effect on response latencies or durations)⁴. We also observed that response latencies were overall slower in picture naming than in reading aloud, a finding that was first observed over a century ago (Cattell, 1885).

⁴Taikh et al. (2015) recently published semantic decision times from a study in which participants saw a series of pictures (or a series of words) one at a time on the screen, and had to decide whether each represents something living or nonliving. Thirty-two of our stimuli overlapped with the items used in the Taikh et al. (2015) study. If picture naming involves semantic activation of the target stimuli, picture naming RTs for these 32 items in our study should correlate with semantic decision times for the same pictures in the Taikh et al. study. However, reading aloud RTs for these 32 items in our study may not correlate with semantic decision times for the same words in the Taikh et al. study. This was the case; the correlation between our picture naming RTs and their semantic decision times for the 32 pictures was significant ($r = 0.36, p < 0.05$), whereas the correlation between our reading aloud RTs and their semantic decision times for the 32 words was not ($r = 0.16, p > 0.05$). We thank Marc Brysbaert for pointing us to the Taikh et al. (2015) article and for suggesting this analysis.

The analyses of initial-phoneme durations, which were the focus of the present research, were overall consistent with the findings from previous studies that investigated effects of lexical frequency on acoustic durations (e.g., Pluymaekers et al., 2005a; Gahl, 2008; etc.). In particular, LF items yielded longer initial-phoneme durations than HF items, yet the size of this effect was very small and missed significance. Separate analyses of the picture naming and reading aloud data revealed a significant frequency effect on initial-phoneme durations for picture naming but not for reading aloud. Even though this finding is consistent with our hypothesis, namely that lexical frequency effects on articulation should be more prominent in picture naming than in reading aloud, the small size of this effect (2 ms) in combination with the absence of a significant interaction between frequency and task does not allow us to firmly conclude that lexical frequency trickles down to affect articulatory levels of processing in speech production.

Furthermore, we observed that both initial-phoneme and whole-word durations were significantly longer in reading aloud than in picture naming. This finding is consistent with the findings of Hennessey and Kirsner (1999) who reported that response durations of the same words were longer in reading aloud than in picture naming (for LF items only). To explain their findings, the authors postulated that reading aloud is initiated on the basis of partial information from the printed word. Because of this early start, the computation of phonology of the rest of the word needs to be carried out during response execution, thus resulting in longer response durations in this task compared to picture naming. This account could explain our data. If response execution in reading aloud is stretched out to compensate for an early start, we may observe that in our reading aloud data, faster RTs are associated with longer initial-phoneme and whole-word durations. As we expected, the nature of the relationship between RTs and initial-phoneme durations, and RTs and whole-word durations in the reading aloud task was negative, however the correlation was weak in both cases ($r = -0.27, p < 0.001$, and $r = -0.06, p < 0.01$, respectively).

To conclude, the present study investigated effects of lexical frequency on articulation using the same stimuli in a picture naming and a reading aloud task. In agreement with previous studies, we obtained longer initial-phoneme durations for LF items than for HF items. However, the observed frequency effect reached significance only in the picture naming task. Our data suggest that high levels of cognitive processing influence, to some extent, low levels of articulatory processing. Yet, given the small size of the effect, we are reluctant to draw firm conclusions about whether the nature of the relationship between cognitive and articulatory levels of processing in speech production is cascaded or staged.

AUTHOR NOTE

Ethics approval for this research was obtained from the department of Psychology at Royal Holloway, University of London (2012/086). The authors would like to thank Eva Liu for hand-marking participants' response latencies and labeling

the acoustic boundaries of participants' response durations. This research was supported by a British Academy Postdoctoral Fellowship to the first author. The second author was supported by a research grant from the Economic and Social Research Council (ES/L002264/1).

REFERENCES

- Aylett, M., and Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang. Speech* 47, 31–56. doi: 10.1177/00223830940470010201
- Aylett, M., and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *J. Acoust. Soc. Am.* 119, 3048–3059. doi: 10.1121/1.2188331
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. New York, NY: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Baese-Berk, M., and Goldrick, M. (2009). Mechanisms of interaction in speech production. *Lang. Cogn. Process.* 24, 527–554. doi: 10.1080/01690960802299378
- Balota, D. A., Boland, J. E., and Shields, L. W. (1989). Priming in pronunciation: beyond pattern recognition and onset latency. *J. Mem. Lang.* 28, 14–36. doi: 10.1016/0749-596X(89)90026-0
- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *J. Mem. Lang.* 42, 1–22. doi: 10.1006/jmla.1999.2667
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2013). *lme4: Linear Mixed-effects Models Using Eigen and S4*. R Package Version 1.0–5 [Computer Software]. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Bates, E., Burani, C., D'Amico, S., and Barca, L. (2001). Word reading and picture naming in Italian. *Mem. Cogn.* 29, 986–999. doi: 10.3758/BF03195761
- Bell, A., Brenier, J., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *J. Mem. Lang.* 60, 92–111. doi: 10.1016/j.jml.2008.06.003
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott Int.* 5, 341–345. Available online at: http://www.fon.hum.uva.nl/ptraat/manual/FAQ_How_to_cite_Praat.html
- Brysbaert, M., and New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990. doi: 10.3758/BRM.41.4.977
- Bybee, J., and Scheibman, J. (1999). The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics* 37, 575–596. doi: 10.1093/acprof:oso/9780195301571.003.0014
- Bybee, J. (2000). "The phonology of the lexicon: evidence from lexical diffusion," in *Usage-based Models of Language*, eds M. Barlow and S. Kemmer (Stanford, CA: CSLI), 65–85.
- Cattell, J. M. (1885). Über die zeit der erkennung und benennung von schriftzeichen, bildern und farben [The time it takes to recognize and name letters, pictures, and colors]. *Philis. Stud.* 2, 635–650.
- 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
- Croot, K., Fletcher, J., and Harrington, J. (1992). "Phonetic segmentation of the Australian National Database of Spoken Language," in *Proceedings of the 4th International Conference on Speech Science and Technology* (Brisbane), 86–90.
- Damian, M. F. (2003). Articulatory duration in single-word speech production. *J. Exp. Psychol.* 29, 416–431. doi: 10.1037/0278-7393.29.3.416
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* 93, 283–321. doi: 10.1037/0033-295X.93.3.283
- Ferrand, L. (1999). Why naming takes longer than reading? The special case of Arabic numbers. *Acta Psychol.* 100, 253–266. doi: 10.1016/S0001-6918(98)00021-3
- Fidelholz, J. (1975). Word frequency and vowel reduction in English. *Chic. Linguist. Soc.* 11, 200–213.
- Forster, K. I., and Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35, 116–124. doi: 10.3758/BF03195503
- Fosler-Lussier, E., and Morgan, N. (1999). Effects of speaking rate and word predictability on conversational pronunciations. *Speech Commun.* 29, 137–158. doi: 10.1016/S0167-6393(99)00035-7
- Fowler, C. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Lang. Speech* 31, 307–319. doi: 10.1177/002383098803100401
- Fowler, C. A., and Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *J. Mem. Lang.* 26, 489–504. doi: 10.1016/0749-596X(87)90136-7
- Gahl, S. (2008). "Time" and "thyme" are not homophones: word durations in spontaneous speech. *Language* 84, 474–496. doi: 10.1353/lan.0.0035
- Gahl, S., and Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: syntactic probabilities affect pronunciation variation. *Language* 80, 748–775. doi: 10.1353/lan.2004.0185
- Gahl, S., Garnsey, S. M., Fisher, C., and Matzen, L. (2006). "That sounds unlikely": syntactic probabilities affect pronunciation," in *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (Vancouver, BC), 1334–1339.
- Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *J. Mem. Lang.* 66, 789–806. doi: 10.1016/j.jml.2011.11.006
- Goldrick, M., and Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: evidence from tongue twisters. *Lang. Cogn. Process.* 21, 649–683. doi: 10.1080/01690960500181332
- Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., and Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition* 103, 386–412. doi: 10.1016/j.cognition.2006.05.010
- Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., and Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chic. Linguist. Soc.* 35, 151–166.
- Griffin, Z. M., and Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *J. Mem. Lang.* 38, 313–338. doi: 10.1006/jmla.1997.2547
- Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang.* 96, 280–301. doi: 10.1016/j.bandl.2005.06.001
- Hawkins, S., and Warren, P. (1994). Implications for lexical access of phonetic influences on the intelligibility of conversational speech. *J. Phon.* 22, 493–511.
- Hennessey, N. W., and Kirsner, K. (1999). The role of sub-lexical orthography in naming: a performance and acoustic analysis. *Acta Psychol.* 103, 125–148. doi: 10.1016/S0001-6918(99)00033-5
- Hooper, J. B. (1976). "Word frequency in lexical diffusion and the source of morphophonological change," in *Current Progress in Historical Linguistics*, ed W. Christie (Amsterdam: North Holland), 96–105.
- Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). "Probabilistic relations between words: evidence from reduction in lexical production," in *Frequency and the Emergence of Linguistic Structure*, eds J. Bybee and P. Hopper (Amsterdam: Benjamin), 229–254.
- Kawamoto, A. H., Kello, C. T., Higareda, I., and Vu, J. Q. (1999). Parallel processing and initial phoneme criterion in naming words: evidence from frequency effects on onset and rime duration. *J. Exp. Psychol.* 25, 362–381. doi: 10.1037/0278-7393.25.2.362
- Kawamoto, A. H., Kello, C. T., Jones, R., and Bame, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation: evidence based on response latency and initial phoneme duration. *J. Exp. Psychol.* 24, 862–885. doi: 10.1037/0278-7393.24.4.862

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01571>

- Kello, C. T., Plaut, D. C., and MacWhinney, B. (2000). The task-dependence of staged versus cascaded processing: an empirical and computational study of Stroop interference in speech production. *J. Exp. Psychol.* 129, 340–361. doi: 10.1037/0096-3445.129.3.340
- Klatt, D. H. (1975). Voice onset time, frication and aspiration in word-initial consonant clusters. *J. Speech Hear. Res.* 18, 686–706. doi: 10.1044/jshr.1804.686
- Krug, M. (1998). String frequency: a cognitive motivating factor in coalescence, language processing, and linguistic change. *J. Engl. Linguist.* 26, 286–320. doi: 10.1177/007542429802600402
- Kuperman, V., Stadthagen-Gonzales, H., and Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behav. Res. Methods* 44, 978–990. doi: 10.3758/s13428-012-0210-4
- Kuznetsov, A., Brockhoff, P. B., and Christensen, R. H. B. (2013). *lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models (Lmer Objects of lme4 Package). R Package Version 2.0–3 [Computer Software]*. Retrieved from <http://CRAN.R-project.org/package=lmerTest>
- Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/S0140525X99001776
- Lieberman, P. (1963). Some effects of the semantic and grammatical context on the production and perception of speech. *Lang. Speech* 6, 172–175. doi: 10.1177/002383096300600306
- Liu, H., Bates, E., Powell, T., and Wulfeck, B. (1997). Single-word shadowing and the study of lexical access. *Appl. Psycholinguist.* 18, 157–180. doi: 10.1017/S0142716400009954
- Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear.* 19, 1–36. doi: 10.1097/00003446-199802000-00001
- Marian, V., Bartolotti, J., Chabal, S., and Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE* 7:e43230. doi: 10.1371/journal.pone.0043230
- McMillan, C. T., and Corley, M. (2010). Cascading influences on the production of speech: evidence from articulation. *Cognition* 117, 243–260. doi: 10.1016/j.cognition.2010.08.019
- Meschyan, G., and Hernandez, H. (2002). Age of acquisition and word frequency: determinants of object-naming speed and accuracy. *Mem. Cogn.* 30, 262–269. doi: 10.3758/BF03195287
- Meyer, A. S. (1990). The time course of phonological encoding in language production: the encoding of successive syllables. *J. Mem. Lang.* 29, 524–545. doi: 10.1016/0749-596X(90)90050-A
- Munson, B. (2007). “Lexical access, lexical representation, and vowel production,” in *Laboratory Phonology 9*, eds J. S. Cole and J. I. Hualde (Berlin; New York, NY: Mouton), 201–228.
- Munson, B., and Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Speech Lang. Hear. Res.* 47, 1048–1058. doi: 10.1044/1092-4388(2004/078)
- Perry, C., Ziegler, J. C., and Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychol. Rev.* 114, 273–315. doi: 10.1037/0033-295X.114.2.273
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol. Rev.* 103, 56–115. doi: 10.1037/0033-295X.103.1.56
- Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005a). Lexical frequency and acoustic reduction in spoken Dutch. *J. Acoust. Soc. Am.* 118, 2561–2569. doi: 10.1121/1.2011150
- Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005b). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62, 146–159. doi: 10.1159/000090095
- Pouplier, M. (2007). Tongue kinematics during utterances elicited with the SLIP technique. *Lang. Speech* 50, 311–341. doi: 10.1177/00238309070500030201
- Protopapas, A. (2007). CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behav. Res. Methods* 39, 859–862. doi: 10.3758/B.F.03192979
- R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <http://www.R-project.org/>.
- Rastle, K., Harrington, J., Coltheart, M., and Palethorpe, S. (2000). Reading aloud begins when the computation of phonology is complete. *J. Exp. Psychol.* 26, 1178–1191. doi: 10.1037/0096-1523.26.3.1178
- Rhodes, R. A. (1992). “Flapping in American English,” in *Proceedings of the Seventh International Phonology Meeting* (Turin: Rosenberg and Sellier), 217–232.
- Rhodes, R. A. (1996). “English reduced vowels and the nature of natural processes,” in *Natural Phonology: The State of The Art*, eds B. Hurch and R. A. Rhodes (The Hague: Mouton), 239–259.
- Riès, S., Legou, T., Burle, B., Alario, F. X., and Malfait, N. (2012). Why does picture naming take longer than word reading? The contribution of articulatory processes. *Psychon. Bull. Rev.* 19, 955–961. doi: 10.3758/s13423-012-0287-x
- Riès, S., Legou, T., Burle, B., Alario, F.-X., and Malfait, N. (2014). Corrigendum to “Why does picture naming take longer than word naming? The contribution of articulatory processes.” *Psychon. Bull. Rev.* 22, 309–311. doi: 10.3758/s13423-014-0668-4
- Scarborough, R. A. (2004). *Coarticulation and the Structure of the Lexicon*. Unpublished doctoral dissertation. UCLA, Los Angeles, CA.
- Schriefers, H., and Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *Eur. J. Cogn. Psychol.* 11, 17–50. doi: 10.1080/713752301
- Szekely, A., Jacobsen, T., D’Amico, S., Devescovi, A., Andonova, E., Herron, D., et al. (2004). A new on-line resource for psycholinguistic studies. *J. Mem. Lang.* 51, 247–250. doi: 10.1016/j.jml.2004.03.002
- Taikh, A., Hargreaves, I. S., Yap, M. J., and Pexman, P. M. (2015). Semantic classification of pictures and words. *Q. J. Exp. Psychol.* 68, 1502–1518. doi: 10.1080/17470218.2014.975728
- Theios, J., and Amrhein, P. C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: reading, naming, and visual and conceptual comparisons. *Psychol. Rev.* 96, 5–24. doi: 10.1037/0033-295X.96.1.5
- Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., and Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Lang. Cogn.* 1, 147–165. doi: 10.1515/LANGCOG.2009.008
- Van Heuven, W. J., 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
- Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S, 4th Edn*. New York, NY: Springer.
- Wright, R. (1997). *Lexical Competition and Reduction in Speech: a Preliminary Report*. Indiana University research on spoken language processing progress report no. 21, 471–485.
- Wright, R. (2004). “Factors of lexical competition in vowel articulation,” in *Papers in Laboratory Phonology VI*, eds J. Local, R. Ogden, and R. Temple (Cambridge: Cambridge University Press), 26–50.
- Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harv. Stud. Classic. Philol.* 15, 1–95. doi: 10.2307/310585

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

Copyright © 2015 Mousikou and Rastle. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Incrementality in Planning of Speech During Speaking and Reading Aloud: Evidence from Eye-Tracking

Lesya Y. Ganushchak^{1,2,3*} and Yiya Chen^{1,3}

¹ Leiden University Centre for Linguistics, Leiden University, Leiden, Netherlands, ² Education and Child Studies, Faculty of Social and Behavioral Sciences, Leiden University, Leiden, Netherlands, ³ Leiden Institute for Brain and Cognition, Leiden University, Leiden, Netherlands

OPEN ACCESS

Edited by:

Simone Sulpizio,
University of Trento, Italy

Reviewed by:

Vitória Plai,
University of California Berkeley, USA
Linda Wheeldon,
University of Birmingham, UK

*Correspondence:

Lesya Y. Ganushchak
lganushchak@gmail.com

Specialty section:

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

Received: 27 October 2015

Accepted: 08 January 2016

Published: 26 January 2016

Citation:

Ganushchak LY and Chen Y (2016)
Incrementality in Planning of Speech
During Speaking and Reading Aloud:
Evidence from Eye-Tracking.
Front. Psychol. 7:33.
doi: 10.3389/fpsyg.2016.00033

Speaking is an incremental process where planning and articulation interleave. While incrementality has been studied in reading and online speech production separately, it has not been directly compared within one investigation. This study set out to compare the extent of planning incrementality in online sentence formulation versus reading aloud and how discourse context may constrain the planning scope of utterance preparation differently in these two modes of speech planning. Two eye-tracking experiments are reported: participants either described pictures of transitive events (Experiment 1) or read aloud the written descriptions of those events (Experiment 2). In both experiments, the information status of an object character was manipulated in the discourse preceding each picture or sentence. In the Literal condition, participants heard a story where object character was literally mentioned (e.g., *fly*). In the No Mention condition, stories did not literally mention nor prime the object character depicted on the picture or written in the sentence. The target response was expected to have the same structure and content in all conditions (*The frog catches the fly*). During naming, the results showed *shorter* speech onset latencies in the Literal condition than in the No Mention condition. However, no significant differences in gaze durations were found. In contrast, during reading, there were no significant differences in speech onset latencies but there were significantly *longer* gaze durations to the target picture/word in the Literal than in the No Mention condition. Our results show that planning is more incremental during reading than during naming and that discourse context can be helpful during speaker but may hinder during reading aloud. Taken together our results suggest that on-line planning of response is affected by both linguistic and non-linguistic factors.

Keywords: sentence planning, discourse context, reading aloud, naming, eye-tracking, incrementality

INTRODUCTION

To produce a sentence, speakers must prepare a preverbal message and then encode it linguistically (e.g., lexical selection and phonological encoding; Levelt et al., 1999). Current theories of speech planning agree that speaking is an incremental process: speakers plan what they want to say in small chunks rather than planning a whole sentence (for review see Wheeldon, 2013). Thus, during speaking, planning and articulation overlap in time. More recently, Konopka and Meyer (2014) have also argued that during planning, the size of the planning unit may vary in different situations,

resulting in a continuum of incrementality in planning (Konopka and Meyer, 2014). For instance, planning scope can be affected by the goal of the speaker (Ferreira and Swets, 2002), by language-specific linguistic features such as different phrasal word orders (Brown-Schmidt and Konopka, 2008), or even by the availability of cognitive resources (e.g., Wagner et al., 2010; Konopka, 2012). The goal of the present study is to investigate whether and how linguistic factors such as the information status of an event (i.e., given versus new) and non-linguistic factors such as the nature of the production task (i.e., picture naming versus reading aloud) affect the time course of on-line sentence formulation.

It is by now well-recognized that the process of planning for both picture naming and reading aloud is a highly dynamic one, a major reflection of which is the variability of the unit of planning within an utterance, ranging from an entire clause to a single phrase or a lexical item (for review see Konopka, 2012). Furthermore, zooming into the range of linguistic factors that affect planning, a consistent finding is that the accessibility or information status of the agent and patient of an event plays a significant role in the way utterances are formulated to describe the event. There is abundant evidence that speakers prefer to begin sentences with accessible characters (e.g., Prat-Sala and Branigan, 2000; Bock et al., 2004; Christianson and Ferreira, 2005; Branigan et al., 2008; Konopka and Meyer, 2014). So, easy-to-name characters become subjects more often than harder-to-name characters (e.g., Konopka and Meyer, 2014). This is in accordance with the so-called minimal load principle (Levelt, 1989) which states that completing easy processes before hard processes results in a lighter cognitive load on the production system, which in turn enables speakers to quickly begin and complete the encoding of individual increments (e.g., Ferreira and Henderson, 1998). For example, Konopka and Meyer (2014) showed that the planning of simple subject-verb-object (SVO) utterances was affected by the accessibility of a referent. Ganushchak et al. (2014) showed that information status (whether the information is new and therefor focused) also affects planning of utterances. In their experiments, Ganushchak et al. (2014) asked participants to describe pictures of two-character transitive events, while participants' eye-movements were recorded. Discourse focus was manipulated by presenting questions before each picture. Their results showed that speakers rapidly directed their gaze to the *new* character they needed to encode.

Planning has also been reported to be affected by non-linguistic factors, such as the nature of the production task: reading versus. naming. Word reading and picture naming have been extensively studied throughout the history of psycholinguistics. Previous studies that explored word reading and picture naming in sentence context showed shorter latencies for word reading compared to latencies for picture naming (e.g., Potter et al., 1986; Theios and Amrhein, 1989). Furthermore, during scene description, utterance formulation begins with an apprehension phase (0–400 ms after picture onset) during which speakers encode the “gist” of the event. The apprehension phase is then followed by linguistic encoding that lasts until

the end of articulation (Griffin and Bock, 2000). No such apprehension phase, however, is necessary during reading. Thus, picture naming requires conceptual preparation and selection of the correct name from other plausible alternatives, whereas reading could be achieved without access to the full semantic representation of the word (e.g., Potter et al., 1986; Levelt et al., 1999). There is evidence that during reading, the semantic system is recruited only when readers have difficulty to generate the pronunciation of a word by relying on orthography-to-phonology mapping alone (e.g., Cortese et al., 1997).

Thus far, the few studies that directly investigated the planning processes in reading versus naming have focused mainly upon the comparison between naming and reading of numerals (e.g., Ferrand, 1999; Meeuwissen et al., 2003; Korvorst et al., 2006). For instance, in an eye-tracking experiment, Korvorst et al. (2006) presented complex numerals in Arabic or an alphabetic format and asked participants to either name the numerals or read them aloud as house numbers or as clock times. They found that the degree of incrementality in planning was affected mainly by the nature of the utterance (house number versus clock times). Furthermore, utterance planning was influenced by different factors in the two production tasks but this was only evident in the production of clock times and not house numbers. Specifically, during the naming of clock times, gaze duration was affected by morpho-phonological (e.g., number of phonemes) as well as conceptual factors (e.g., factors related to telling time in Dutch; see Meeuwissen et al., 2003; Korvorst et al., 2006). However, during the reading of clock times, gaze durations reflected only morpho-phonological differences. This suggests that during reading aloud, conceptual preparation was no longer required (Meeuwissen et al., 2003; Korvorst et al., 2006). Thus, the presence and absence of conceptual preparation is responsible for the planning differences of an utterance between naming and reading tasks (Meeuwissen et al., 2003; Korvorst et al., 2006).

No study thus far, however, has investigated how linguistic (accessibility) and non-linguistic (production task) factors may interact to affect the planning of an utterance. To address this question, two comparable groups of participants were asked to describe a simple event (Experiment 1) or read aloud the written description of the same event (Experiment 2) while their eye-movements and onset speech latencies were recorded. Furthermore, we manipulated the accessibility of the object character of an event by providing two different discourse contexts prior to each picture or written sentence. In the *Literal condition*, the object character (e.g., *fly*) was literally mentioned in the preceding context. In the *No Mention condition*, stories did not literally mention nor prime any of the characters depicted on the picture. The target response was expected to have the same structure and content in all conditions (*The frog catches the fly*).

Differences in the planning of the target response in the naming and reading tasks were evaluated by the time needed for the preparation of speech, as reflected in speech onset time, but also by comparing speakers' eye-movements to the object characters in the picture and object words in the written sentence,

respectively. Gaze duration provides another good measure for estimating the total amount of speech planning that is required in order to produce an utterance (e.g., Meyer and Lethaus, 2004).

A distinction between early and late processing was also examined. Good index of early processing is (a) *first gaze durations*, which is the sum of all fixation durations on a target word/character prior to moving to another region. Measurement indexing the late processing is (b) *total gaze durations*, which is the sum of all fixations on a region. From previous literature, we know that first gaze duration is sensitive to earlier comprehension processes such as word recognition (see Clifton et al., 2007, for an overview), while total gaze duration reflects later processing such as re-analysis and discourse integration (e.g., Rayner, 1998; Frisson and Pickering, 1999; Sturt, 2007). Longer duration is usually taken as an indication of more effortful integration processes (e.g., Rayner and Sereno, 1994).

As mention above, in a picture description task, the formulation of a sentence begins with a short apprehension phase during which speakers encode the gist of the event (e.g., Griffin and Bock, 2000; Meyer and Lethaus, 2004; Konopka, 2014). Event apprehension is then followed by a longer phase of linguistic encoding. Typically, easy-to-name characters are fixated for less time than harder-to-name characters (e.g., Griffin and Bock, 2000; Meyer and Lethaus, 2004; Konopka, 2014). In a reading task, no apprehension phase is expected of the whole event described in the text. Readers typically do not need to read the whole sentence first, prior to start reading aloud. Another difference between reading and picture description is that during reading, a given word is likely to prompt the reader to try to integrate the word to the prior reference in the discourse context, whereas this integration process is likely to occur earlier (e.g., during the gist preparation stage) in the picture naming task.

Thus, we propose that during picture naming, the unit of planning is larger than during the reading aloud task. Speakers are unlikely to start speaking before they understood the gist of the event depicted in the picture. During reading, however, speakers start reading immediately after the onset of the written sentence. Therefore, we predict that first fixations to the object character will be earlier than speech onset in the naming task, but later than speech onset in the reading task. Consequently, the discourse context that we manipulated should also affect planning in naming and reading differently. Namely, the accessibility of an object character should ease the linguistic encoding phase in naming but not necessarily so in the reading task. We then predict that speakers will initiate their speech faster in the reading task than in the naming task. Furthermore, in the naming task, there should be faster onset latencies and shorter gaze durations in the Literal condition compared to the No Mention condition. In the reading task, however, no differences are expected in the onset latencies between the Literal and the No Mention condition. As for the eye gaze characteristics, we expect that the gaze duration to the object word should be less in the Literal condition than in the No Mention condition, as readers may recognize the target word from the preceding context, which in turn can facilitate the integration processes.

EXPERIMENT 1. PLANNING OF SPEECH DURING SPEAKING

Methods

Participants

Thirty-one native Dutch speakers (28 women) participated in the experiment (mean age: 20 years; $SD = 1.9$ years). All participants were students of Dutch universities. The study was approved by the ethical committee board at Leiden University. Participants gave written informed consent prior to participating in the study and received course credits for their participation. Due to technical problem, data of one participant was excluded from the analysis.

Materials

Seventy-eight colored pictures were used in the experiment (Konopka, 2014). All pictures displayed simple actions (**Figure 1**). There were 25 target pictures of transitive events, 50 fillers, and 3 practice pictures.

Accessibility was manipulated by means of short stories preceding each picture. All stories consisted of two sentences. The stories were only contextually related to the pictures, and were not intended to help participants understand the gist of the depicted event. Take the expected target sentence *De kikker vangt de vlieg* ('The frog catches the fly') as an example, the following illustrates the two conditions provided before the presentation of the target picture.

- (1) Literal condition: The object character was literally mentioned in the preceding story. Note that the target object character was always placed in the same grammatical role as in the intended target sentence and it was always placed in the second sentence of the story.

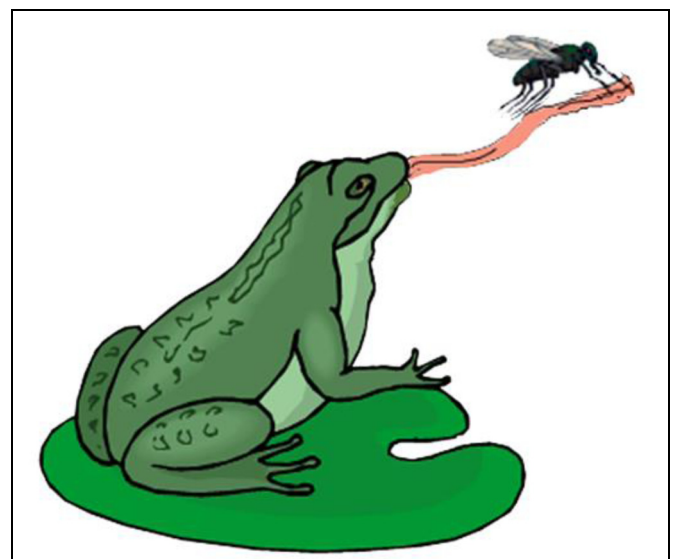


FIGURE 1 | Example of a target picture event.

David vist regelmatig en weet dus ook het een en ander over vissen. Hij gebruikt een kleine vlieg als aas. (David fishes regularly and knows a thing or two about fishing. He uses a small fly as bait.)

- (2) **No Mention condition:** The story did not include literal or associative mention of words that describe characters in the picture.

David gaat met zijn vader vissen. Ze gebruiken restjes van het avondeten als aas. (David is going fishing with his father. They use leftovers from dinner as bait.)

All stories were pre-recorded by a native Dutch female speaker and presented auditorily prior to picture onset. For 40% of the filler trials, after the story, a yes-or-no comprehension question was presented visually on the computer screen. The purpose of the questions was to make sure that participants listened attentively to the presented stories.

Design and Procedure

Lists of stimuli were created to counterbalance story types across target pictures. Each target picture occurred in each condition on different lists, so that each participant saw each picture only once. Each subject saw eight target pictures per condition. There were at least two filler pictures separating any two target trials in each list.

Participants were seated in a sound-proof room. Eye movements were recorded with an Eyelink 1000 eye-tracker (SR Research Ltd.; 500 Hz sampling rate). Screen resolution was set at 1024 × 768. A 9-point calibration procedure was used. Eye calibration was done at the beginning of the experiment. The task started with three practice trials. Each trial started with the blank screen of 500 ms. followed, by the auditory presentation of the story (presented through headphones). The duration of the story varied (mean = 5804 ms; *SD* = 1302 ms). Simultaneously with the story, a pictorial representation of 'Listen' was presented at the top-center of the screen (as shown in **Figure 2**). On 40% of the filler items, there was a yes/no comprehension question presented prior to a picture trial. Participants used computer mouse to give their response. For all target trials and 60% of filler trials, after the completion of the story, the experiment proceeded to the picture trial. The picture trials began with drift correction, which also served as a fixation point, presented at the top of the screen. Afterwards, a picture was presented on the screen. Participants were instructed to describe each picture with one sentence which should mention all the characters in the picture. The time interval between offset of the auditory story and picture onset slightly varied per trials, as it was dependent on how quickly the eye fixations were registered during the drift correction phase. Participants were not under time pressure to produce the response. When the participant finished speaking, the experimenter clicked with the mouse to proceed to the next trial. On average, the pictures were displayed on the screen for 5227 ms (*SD* = 1604 ms).

Scoring and Data Analysis

Only responses with active SVO structure were scored as correct. Trials with a different structure (e.g., passive), wrong description, or corrections during the description were excluded from further analysis (Literal: 6.3%; No Mention: 7.2%).

Interest areas were drawn around each character in the target pictures (allowing a 2–3 cm margin around each character). Note, that the fixations were concentrated around the characters themselves; a more tightly fit ROI would not affect the reported results. Trials in which the first fixation was within the subject or object character interest area, instead of the fixation point, were also removed from further analyses (2% of the data). This left 440 trials for analysis. Analyses were carried out on speech onsets of correct responses. For the eye-tracking data, we determined first and total *gaze durations* for the targets. Speech onsets of correct responses and gaze durations were first log-transformed to remove the intrinsic positive skew and non-normality of the distribution (Baayen et al., 2008). Mixed-effects model analyses were carried out with participants and items as random effects and Condition (i.e., Literal and No Mention) as fixed effect. All models included random by-participant and by-item random intercepts and slopes for the factor Condition.

Results

The time of first fixation on the subject character was on average 338 ms in the Literal condition and 336 ms in the No Mention Condition. Looks to the object character occurred

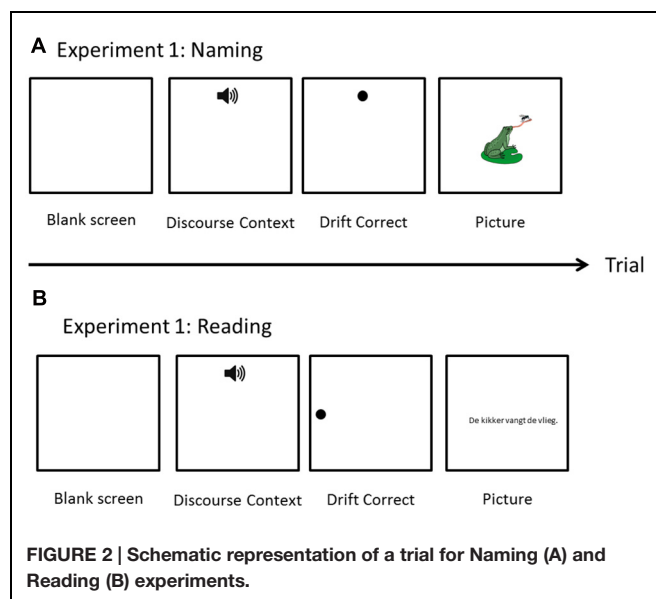


TABLE 1 | Mean response latencies in ms (and standard deviation) per condition in Naming (Experiment 1), in Reading (Experiment 2), and the mean difference across conditions (No Mention – Literal Mention).

	No Mention	Literal Mention	Mean Difference
Naming	1903 (496)	1841 (489)	62
Reading	834 (106)	827 (140)	7

at about 846 and 819 ms after the picture onset, in the Literal and No Mention condition, respectively. This is about 1000 ms *earlier* than the onset latencies (see **Table 1**). Note, the differences between times of the first fixation on the subject and object characters did not significantly differ across condition (all $t_s < 1.5$).

Speech Onsets

Participants started speaking significantly earlier in the Literal compared to the No Mention condition ($\beta = 0.05$, $SE = 0.02$, $t = 2.03$, $p = 0.04$; see **Table 1**).

Eye-Tracking Data on Object Character

No significant difference was found in both the *first and total gaze durations* on the object character between the Literal and No Mention conditions (all $t_s < 1$; see **Table 2** for means).

DISCUSSION

Overall, speech was initiated about 1000 ms after the first fixations to both the subject and object characters in the event pictures. This suggests that participants started to articulate the sentences after the apprehension phase and presumably after some of the linguistic encoding phase was completed. The onset of articulation was influenced by the discourse context manipulation. Speakers were significantly faster initiating production when the object character was given as compared to when the object character was contextually new. These results suggest that the activation of the object characters in the upcoming event facilitated planning. Note, that this does not necessarily mean that the speakers anticipated the upcoming events. Rather, we believe that it is likely due to the information about the object character given in the discourse context which made the encoding of the object character easier in the Literal condition than the No Mention condition. The lack of significant difference between the Literal and the No Mention condition in terms of both the *first and total gaze durations* on the object character suggest that neither object recognition nor integration into the discourse context were affected by our manipulation. We take this as evidence that the information status of the object character did not exert any effect on the planning of the initial chunk of speech in the naming task.

TABLE 2 | Mean first and total gaze durations on object character/word in ms (and standard deviation) per condition in Naming (Experiment 1), in Reading (Experiment 2) and the mean difference across conditions (No Mention – Literal Mention).

	Naming		Reading	
	First gaze Duration	Total gaze Duration	First gaze Duration	Total gaze Duration
No Mention	558 (225)	1857 (440)	563 (155)	680 (175)
Literal Mention	542 (205)	1796 (527)	580 (160)	755 (221)
Mean Difference (No Mention – Literal)	16	61	–17	–75

EXPERIMENT 2. PLANNING OF SPEECH DURING READING ALOUD

Methods

Participants

Thirty-one native Dutch speakers (28 women) participated in the experiment (mean age: 20 years; $SD = 1.9$ years). None of the participants took part in Experiment 1. All participants were students of Dutch universities. The study was approved by the ethical committee board at Leiden University. Participants gave written informed consent prior to participating in the study and received course credits for their participation. Due to technical problem, data of one participant was excluded from the analysis.

Materials

The description of events produced by participants from Experiment 1 were used as targets in this experiment. To account for variability in responses, responses of each participant from Experiment 1 were used in its own list in the present study. Thus, 30 unique lists were created. Trials with erroneous responses were replaced by the corresponding standard target sentence (e.g., *De kikker vangt de vlieg*, 'The frog catches the fly').

Design, Procedure, and Data Analysis

The design, procedure, and analyses were identical to Experiment 1. Interest areas were marked around target object words of each sentence as pre-defined by the analyzing software Data Viewer (SR Research Ltd.). Target trials with erroneous responses were removed from further analysis (Literal: 1.3%; No Mention: 3.0%). This left 472 trials for the analyses reported below.

Results

The time of first fixation on the subject words was on average 293 ms in the Literal condition and 334 ms in the No Mention Condition. First fixation to the object words occurred at about 1221 and 1336 ms after the sentence onset in the Literal and No Mention condition, respectively. This is about 1000 ms *later* than speech onset (see **Table 1**). The difference between first fixation time between Literal and No Mention condition was not significant for looks to the subject word ($t < 1$). However, participants fixated on the object word significantly earlier in the Literal Mention than No Mention condition ($\beta = 0.4$, $SE = 0.03$, $t = -2.40$, $p = 0.04$)¹.

Speech Onsets

No effects were found for speech onset latencies (all $t_s < 1$; see **Table 1** for means).

Eye-Tracking Data on Object Word

No significant effects were found for *first gaze durations* (all $t_s < 1.5$; see **Table 2** for means). However, there was a significant difference between conditions for *total gaze duration*. Namely,

¹Note, that the sentence length and number of words prior to the object word is comparable for the Literal (5.19 words; $SD = 0.38$ words) No Mention conditions (5.22 words; $SD = 0.25$ words).

participants looked longer at the target word in the Literal condition compared to the No Mention condition ($\beta = -0.2$, $SE = 0.06$, $t = -2.03$, $p = 0.04$).

Discussion

Contrary to Experiment 1, speakers initiated speech well before taking a look at the object word. This indicates that speakers started producing sentences before the comprehension of the whole event described in the written text. Onset latencies as well as first gaze durations were unaffected by the accessibility of the object character. However, total gaze durations were affected by the accessibility of the object character. Namely, in the Literal condition, participants looked at object character longer than in the No Mention condition.

GENERAL DISCUSSION

We reported two eye-tracking experiments that investigated the extent to which speakers' simultaneous planning and articulation of an utterance is influenced by linguistic (accessibility) and non-linguistic (production task) factors. Our results show clearly that planning processes differ during naming and reading aloud. This is in accordance with previous findings (e.g., Meeuwissen et al., 2003; Korvorst et al., 2006). The crucial factor that influences planning in these two production tasks is conceptual preparation (Meeuwissen et al., 2003; Korvorst et al., 2006). In the naming task, participants had to describe events depicted on the pictures, which required conceptual preparation and selection of appropriate names for characters from the competing alternatives. In the reading task, however, no conceptual preparation and no word selection were necessary.

Another way to account for the differences between the naming and reading tasks is that in the picture naming task, the unit of planning was larger than during reading. In the picture description task, speakers initiated their speech around 1872 ms, much later than when they gazed upon the subject (337 ms after picture onset) and object characters (833 ms after picture onset). In the reading task, however, the initiation of speech was much earlier (830 ms), after they have looked at the subject word (312 ms after sentence onset), but much earlier than when they paid attention to the object word (1279 ms after sentence onset). This suggests that in naming, speakers had to encode the object character before they started speaking; while in the reading task, the object word is encoded only after the participants have already started articulating the first part of the sentence.

The differences in naming and reading aloud were also reflected in how discourse context affected the planning processes. In both speech production tasks, we observed effects of accessibility, which though manifested in two different ways. For the naming task, literal mention of the object character resulted in *facilitated* speech onset latency while no such effect was found in the reading task. This may be taken to indicate that the accessibility of the object did help to speed up the planning process during naming, probably all the way from the conceptualization of the message down to the retrieval and phonological encoding of the lexical item for the object.

Our results do not allow to disentangle with certainty at which stage of planning (e.g., conceptual versus phonological) did the facilitation effect arise. In future studies, one might manipulate different levels of information that is provided by the discourse context (e.g., only conceptual information versus only phonological information).

In contrast, literal mention of the object character resulted in *inhibition* (as suggested by the longer gaze durations) during the reading task. Specifically, readers looked at the object word significantly longer in the Literal condition (755 ms) than in the No Mention condition (680 ms). Interestingly, the readers fixated on the object word significantly earlier in the Literal condition (1796 ms) than in the No Mention condition (1857 ms). These effects were not found for the naming task. Thus, it appears that readers look at the object word more quickly but also look at it for longer in the Literal condition than in the No Mention condition. The initial facilitation in the processing of the object word may come from the preview benefits from the parafoveal viewing. The preview benefits have been often demonstrated for the words that are orthographically or phonologically related to the target (for review see Schotter et al., 2012). There is also some evidence of processing of semantic information during the preview (e.g., Yan et al., 2009; Hohenstein et al., 2010). It is likely that the processing of object words was initially sped up by the orthographical and phonological (and possibly semantical) information that was activated by the discourse context. Note that the duration of the first fixation on a target word was slightly shorter in the Literal condition (370 ms; $SD = 139$ ms) than in the No Mention condition (378 ms; $SD = 120$ ms), supporting the argument that word recognition processes might have benefited by the available information about the object word. The question that arises here is the later inhibition effects in the Literal Mention condition compared to the No Mention condition.

One reason could be that the inhibition effect resulted from competition of the phonology (and maybe orthography) of the previously activated word in the preceding discourse during the recognition of that same word in the reading of the post-discourse target sentence. Similar effects have been reported in Frisson et al. (2014), which though found an inhibition effect only when the two words overlapped in both phonology and orthography and were close to each other within one sentence. In our experiment, the effect, if verified to result from the same mechanism, was present even when they were as far apart as across different sentences.

Alternatively, this effect may be resulted from the fact that readers were trying to integrate the word to the prior reference in the preceding story. Two possible scenarios could have led to the observed gaze pattern. Possibility one is that such integration process might have been skipped or was shallow in the No Mention condition, compared to the Literal condition, since there was no obvious reference between the preceding context and the target sentence. Another possibility is that such an integration process turned out to be more costly when the given information of the object (provided in the discourse in the Literal condition) was coded with its full name as if it was new information. Further research is needed to find evidence for or against these speculations.

Taken together, our results show that planning is more incremental during reading, where planning and speaking are closely interleaved, than during naming. Reading tasks are often used to investigate language production processes. Our results show that nature of processes may differ across the tasks and that the time course of these processes may not be comparable for reading and naming tasks. Furthermore, our results showed that discourse context can be helpful during speaking but may hinder during reading aloud. Overall, our results suggest that planning is a dynamic process which is affected by both linguistic and non-linguistic factors.

AUTHOR CONTRIBUTIONS

LG: Substantial contributions to the conception or design of the work; the acquisition, analysis, and interpretation of data for the work; drafting the work and revising it critically; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the

accuracy or integrity of any part of the work are appropriately investigated and resolved. YC: Substantial contributions to the conception or design of the work; interpretation of data for the work; critically revising the manuscript; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

ACKNOWLEDGMENTS

We thank Gouming Martens and Maxime Tulling for help with data collection for Experiment 1 and 2, respectively. Special thank you to Agnieszka Konopka who provided the pictures for the Experiment 1. We thank Linda Wheeldon for her insightful comments on data analysis and its interpretation. This research was supported from the European Research Council to YC (ERC Starting grant 206198).

REFERENCES

- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Bock, K., Irwin, D. E., and Davidson, D. (2004). "Putting first things first," in *The Integration of Language, Vision, and Action: Eye Movements and the Visual World*, eds F. Ferreira and M. Henderson (New York, NY: Psychology Press), 249–278.
- Branigan, H. P., Pickering, M. J., and Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua* 118, 172–189. doi: 10.1016/j.lingua.2007.02.003
- Brown-Schmidt, S., and Konopka, A. (2008). Little houses and casas pequeñas: message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition* 109, 274–280. doi: 10.1016/j.cognition.2008.07.011
- Christianson, K., and Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition* 98, 105–135. doi: 10.1016/j.cognition.2004.10.006
- Clifton, C. Jr., Staub, A., and Rayner, K. (2007). "Eye movements in reading words and sentences," in *Eye Movements: a Window on Mind and Brain*, eds R. P. G. van Gompel, M. H. Fischer, W. S. Murray, and R. L. Hill (Oxford: Elsevier), 341–373.
- Cortese, M. J., Simpson, G. B., and Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychon. Bull. Rev.* 4, 226–231. doi: 10.3758/BF03209397
- Ferrand, L. (1999). Why naming takes longer than reading? The special case of Arabic numbers. *Acta Psychol.* 100, 253–266. doi: 10.1080/0361073X.2011.568805
- Ferreira, F., and Henderson, J. M. (1998). Linearization strategies during language production. *Mem. Cogn.* 26, 88–96. doi: 10.3758/BF03211372
- Ferreira, F., and Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *J. Mem. Lang.* 46, 57–84. doi: 10.1006/jmla.2001.2797
- Frisson, S., Koole, H., Hughes, L., Olson, A., and Wheeldon, L. (2014). Competition between orthographically and phonologically similar words during sentence reading: evidence from eye movements. *J. Mem. Lang.* 73, 148–173. doi: 10.1016/j.jml.2014.03.004
- Frisson, S., and Pickering, M. J. (1999). The processing of metonymy: evidence from eye movements. *J. Exp. Psychol.* 25, 1366–1383.
- Ganushchak, L. Y., Konopka, A. E., and Chen, Y. (2014). Focus planning during sentence production: an eye-tracking study. *Poster Presented on International Seminar on Speech Production (ISSP)*, Cologne.
- Griffin, Z. M., and Bock, J. K. (2000). What the eyes say about speaking. *Psychol. Sci.* 11, 274–279. doi: 10.1111/1467-9280.00255
- Hohenstein, S., Laubrock, J., and Kliegl, R. (2010). Semantic preview benefit in eye movements during reading: a parafoveal fast-priming study. *J. Exp. Psychol. Learn. Mem. Cogn.* 36:1150. doi: 10.1037/a0020233
- Konopka, A. E. (2012). Planning ahead: how recent experience with structures and words changes the scope of linguistic planning. *J. Mem. Lang.* 66, 143–162. doi: 10.1016/j.jml.2011.08.003
- Konopka, A. E. (2014). Speaking in context: discourse influences formulation of simple sentences. *Poster presented at the 27th CUNY Human Sentence Processing Conference*, Columbus, OH.
- Konopka, A. E., and Meyer, A. S. (2014). Priming sentence planning. *Cogn. Psychol.* 73, 1–40. doi: 10.1016/j.cogpsych.2014.04.001
- Korvorst, M., Roelofs, A., and Levelt, W. J. M. (2006). Incrementality in naming and reading complex numerals: evidence from eyetracking. *Q. J. Exp. Psychol.* 59, 296–311. doi: 10.1080/17470210500151691
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/S0140525X99001776
- Meeuwissen, M., Roelofs, A., and Levelt, W. J. M. (2003). Planning levels in naming and reading complex numerals. *Mem. Cogn.* 31, 1238–1248. doi: 10.3758/BF03195807
- Meyer, A. S., and Lethaus, F. (2004). "The use of eye tracking in studies of sentence generation," in *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, eds J. M. Henderson and F. Ferreira (New York, NY: Psychology Press), 191–212.
- Potter, M., Kroll, J., Yachzel, B., Carpenter, E., and Sherman, J. (1986). Pictures in sentences: understanding without words. *J. Exp. Psychol.* 115, 281–294. doi: 10.1037/0096-3445.115.3.281
- Prat-Sala, M., and Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: a cross-linguistic study in English and Spanish. *J. Mem. Lang.* 42, 168–182. doi: 10.1006/jmla.1999.2668
- Rayner, K. (1998). Eye movement in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422. doi: 10.1037/0033-2909.124.3.372
- Rayner, K., and Sereno, S. C. (1994). "Eye movements in reading: psycholinguistic studies" in *Handbook of Psycholinguistics*, ed. M. A. Gernsbacher (San Diego, CA: Academic Press), 57–81.

- Schotter, E. R., Angele, B., and Rayner, K. (2012). Parafoveal processing in reading. *Attent. Percept. Psychophys.* 74, 5–35. doi: 10.3758/s13414-011-0219-2
- Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition* 105, 477–488. doi: 10.1016/j.cognition.2006.10.009
- Theios, J., and Amrhein, P. C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: reading, naming, and visual and conceptual comparisons. *Psychol. Rev.* 96, 5–24. doi: 10.1037/0033-295X.96.1.5
- Wagner, V., Jescheniak, J. D., and Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: effects of cognitive load on multiple lexical access. *J. Exp. Psychol.* 36, 423–440. doi: 10.1037/a0018619
- Wheeldon, L. (2013). “Producing spoken sentences: the scope of incremental planning,” in *Cognitive and Physical Models of Speech Production, Speech Perception, and Production-Perception Integration*, eds P. Perrier and P. Lang (Berlin: Springer-Verlag GmbH).
- Yan, M., Richter, E. M., Shu, H., and Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychon. Bull. Rev.* 16, 561–566. doi: 10.3758/PBR.16.3.561

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

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



The segment-to-frame association in word reading: early effects of the interaction between segmental and suprasegmental information

Simone Sulpizio^{1,2*} and Remo Job¹

¹ Department of Psychology and Cognitive Science, University of Trento, Trento, Italy, ² Fondazione Marica De Vincenzi ONLUS, Trento, Italy

OPEN ACCESS

Edited by:

Jonathan Grainger,
Laboratoire de Psychologie Cognitive,
CNRS, France

Reviewed by:

Ludovic Ferrand,
CNRS and University Blaise Pascal,
France
Fabienne Chetail,
Université Libre de Bruxelles, Belgium

*Correspondence:

Simone Sulpizio
simone.sulpizio@unitn.it

Specialty section:

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

Received: 25 May 2015

Accepted: 06 October 2015

Published: 20 October 2015

Citation:

Sulpizio S and Job R (2015)
The segment-to-frame association
in word reading: early effects
of the interaction between segmental
and suprasegmental information.
Front. Psychol. 6:1612.
doi: 10.3389/fpsyg.2015.01612

In four reading aloud experiments we investigated the operations occurring at the level of the phonological buffer by manipulating stress and phoneme information. In all experiments we adopted a masked priming paradigm with three-syllable Italian word targets. Experiments 1 and 2 tested the effect of pure segmental (e.g., fe%%%% – FEcola) and pure suprasegmental (Clnema – FEcola) overlap, respectively. Experiments 3 and 4 tested the joint manipulation of segmental and suprasegmental information, by using prime-target pairs that shared the first syllable and did or did not share their stress pattern (e.g., FEgato – FEcola vs. feNlce – FEcola). The results showed that both segmental and suprasegmental primes affect reading at an abstract phonological level. Moreover, the joint manipulation of stress and phonemes showed an asymmetric pattern for different stress patterns, suggesting that the phonemic and the stress systems address the articulation planning through a process that starts as soon as the relevant information about the to-be-planned unit is active.

Keywords: stress assignment, phonological encoding, masked priming, reading aloud, articulation

INTRODUCTION

Reading aloud involves computing the sound of a word from its visually presented form. In order to carry out such process the execution of multiple operations is required, e.g., perceiving the written stimulus, computing the phonological code, and converting it into a speech signal. Giving its specific nature, reading aloud thus has similarities and differences with both the process of (silent) reading and the process of speech production, the former being about getting from print to meaning and the latter being about getting from concepts to sounds. Since reading aloud may be construed as a print-to-sound mapping process, a key issue for such a process is the understanding of how a phonological code is translated into a sequence of articulatory gestures that correspond to the word's sounds. Despite their importance, the operations involved in the planning and execution of articulation in reading aloud have not been investigated with the same fervor that word recognition or lexical access received. As a consequence, little empirical evidence is available on how readers perform the two steps assumed to follow, i.e., the lexical retrieval and/or the orthography-to-phonology mapping, the phonological encoding – that is the building of a sequence of well-formed phonological syllables – and the phonetic encoding – that is the computation of the phonetic-articulatory gestures of the to be uttered stimulus (Levelt et al., 1999). In most computational models of reading aloud phonological and phonetic encoding are

implemented as an oversimplified set of operations (see, e.g., Rastle and Coltheart, 2000; Coltheart et al., 2001; Arciuli et al., 2010; Perry et al., 2010).

Recent empirical work has shown evidence for a double process at the level of phonological encoding in reading. Similarly to what happens in word production, reading polysyllabic words implies retrieving both segmental (i.e., word sounds) and suprasegmental information (i.e., stress) and these two types of information may be computed separately (Colombo and Zevin, 2009; Sulpizio et al., 2012a,b; Sulpizio and Job, 2013). Ascribing the computation of stress and the computation of phonemes to two separate mechanisms has important consequences on the structure of phonological and phonetic encoding since the assembling of the phonological unit will require the reader to carry out at least three operations: (a) activating the word's segments, (b) activating the stress pattern, and (c) assembling segmental and suprasegmental information. Data on (c) are lacking, but some evidence is available for both (a) and (b).

An insight into the phonological encoding in reading has been provided by the masked onset priming effect (henceforth MOPE; Forster and Davis, 1991; see Grainger and Ferrand, 1996): target words (e.g., sink) are named faster when preceded by a masked prime with the same initial phoneme (e.g., save), than by a prime with a different initial phoneme (e.g., ball). The main account of the MOPE – the speech planning account (Kinoshita, 2000) – assumes that the effect has a serial nature and affects the segment-to-frame association (Kinoshita, 2000; Kinoshita and Woollams, 2002; Malouf and Kinoshita, 2007; Dimitropoulou et al., 2010; but see Mousikou et al., 2010). Such process allows for the active phonological segments to be assigned to an abstract frame – i.e., the word metrical frame – specifying the number of syllables and the stress pattern of the word (e.g., for the word FEcola 'starch,' the metrical word is 'σ σ σ). The MOPE was also found by Schiller (2004) with a slightly different masked priming paradigm, in which participants had to read aloud Dutch words (e.g., banaan, 'banana') under two conditions: when preceded by a prime consisting of an onset-related word embedded in a sequence of symbols (e.g., %%balans%, 'balance') and when preceded by an onset-related sequence prime that consisted of one or two letters embedded in a sequence of symbols (e.g., %%ba%%%%%). Responses to targets were faster in both onset-related conditions than in the control, all symbols condition (%%%%%%%%%) and Schiller suggested that the pre-activation of congruent phonological segments by the prime facilitates the phonological encoding of the target (see, e.g., Schiller and Kinoshita, 2007). Taken together, these findings offer support for a stage of phonological encoding in the reading system; during this stage, after having retrieved/computed word's phonemes and stress, the reader assembles the phonological word through a rightward serial process that associates the phonological segments to a metrical frame. The resulting unit is then used to address the articulatory system (see Levelt et al., 1999 for a detailed description of the phonological encoding in speech production).

With regard to stress, some studies have investigated stress assignment to polysyllabic words addressing the question whether the computation of stress may be independent of the computation of segmental information. The results have been mixed. In a series of implicit form-priming experiments – participants first learn pairs of words (e.g., meer-water 'lake-water'), and then had to produce the second word (e.g., water) of the pair in response to the presentation of the first (e.g., meer) – Roelofs and Meyer (1998) manipulated the stress pattern of the to be produced words (all having either the same or different stress) and did not find any stress priming effect. However, adopting different priming methodologies (all involving visible primes), some reading aloud studies have shown that the metrical structure of a word may be primed independently from its segmental content, and this is possible both when stress is assigned to pseudowords and when it is lexically retrieved (Colombo and Zevin, 2009; Sulpizio et al., 2012a,b). Possible explanations for the divergent results are offered in the General Discussion, but for the time being we assume that computation of stress and segmental information are to some extent independent. To illustrate this issue, we may refer to the Sulpizio et al.'s (2012b) study: readers were presented with prime-target word pairs that did or did not share the stress pattern (e.g., TESsera – BUfala, 'card' – 'hoax' vs. cuGIno – BUfala¹, 'cousin' – 'hoax') and were found to be faster in reading the targets when preceded by a congruent stress prime, than when preceded by an incongruent-stress prime. The finding invites the conclusion that readers have an abstract representation of stress, quite independent from the segmental material and that the representation of stress is involved in the segment-to-frame association and in the articulatory planning of the stimulus, thus affecting target processing.

While phonemic computation and stress assignment are to some extent handled by autonomous systems, they need to interact during processing. Specifically, articulation requires a segment-to-frame association, in which the system associates the computed phonological segments to a metrical frame, and such a well-formed phonological unit will allow articulation (Dell, 1986, 1988; Levelt et al., 1999).

The speech production literature may help to shed light on the functioning of the segment-to-frame association in word reading. Since both reading aloud and speech production require the construction of a phonological unit and its conversion into articulatory programs, they share (at least in part) the stages of processing finalized to encode the phonological word and to use such a phonological word to produce the phonetic realization of the stimulus (Roelofs, 2004).

To investigate the processing of segment-to-frame association and phonological-to-phonetic mapping in word reading we run four experiments in Italian capitalizing on the fact that in such language stress is not graphically marked neither solely determined by orthographic structure² and that, therefore, any

¹Capital letters indicate the stressed syllable.

²The unique Italian rule for stress assignment requires to assign penultimate stress to those words that have a heavy penultimate syllable (e.g., biSONte 'bison'). The rule shows also some exceptions (e.g., MANdorla 'almond', LEpanto).

particular word's stress pattern can only be reliably established through lexically stored information. Our results will be then generalizable to the other polysyllabic languages such as English, with a similar stress system.

Although distributional cues allow Italian readers to assign stress to pseudowords to some extent (Colombo and Zevin, 2009; Sulpizio et al., 2013), such cues play no role in word reading (Paizi et al., 2011; Sulpizio and Colombo, 2013). The fact that in Italian word stress is lexically based may be helpful to investigate phonological encoding: since there is no algorithmic procedure to assign a stress to a stimulus, the metrical structure has to be lexically retrieved and then combined with the segmental material to shape the phonological word, which will be then used by the system to address articulatory programs.

Experiments 1 and 2 investigated the MOPE and the stress priming effects by means of a masked priming paradigm, with a set of tightly controlled stimuli, trying to establish whether the two effects are facilitatory or inhibitory. Moreover, with regard to the MOPE, the use of a pure segmental prime (e.g., *fe*%%%) – FEcola, 'starch') allowed us to test whether the activation of the first phonological segments of the word automatically activates suprasegmental information as the masked segment (e.g., <*fe*>) might activate either a syllabic unit – which may be phonetically specified for stress (i.e., as stressed or unstressed) – or only its segmental constituents (i.e., /f/ and /e/).

We adopted the masked priming paradigm also in Experiments 3 and 4 but the aim here was to test the effect of the joint manipulation of segmental and suprasegmental information. Thus, for each prime-target pair, the prime either shared both the initial phonemes and the stress pattern with the target (e.g., FEgato – FEcola, 'liver' – 'starch'), or shared the initial phonemes with the target but had a different stress pattern (e.g., feNIce – FEcola, 'phoenix' – 'starch'). In the control condition, the prime-target pair shared neither segmental nor suprasegmental information, the prime being composed of a string of symbols (%%%%%%%%). The manipulation is particularly interesting for the fact that Italian three-syllable words have two main stress patterns (Thornton et al., 1997): antepenultimate stress (i.e., the first syllable bears stress, e.g., TAvolo 'table'), and penultimate stress (i.e., the second syllable bears stress, e.g., coLOre 'color'). Although their distribution differs – 80% of three-syllable words bear penultimate stress and 18% bear antepenultimate stress³ – reading of words bearing the dominant penultimate stress pattern is not faster, and the two patterns are assumed to be stored in the phonological lexicon (Burani and Arduino, 2004; Paizi et al., 2011). Thus, a further question we may ask is whether the prime-target manipulation affects similarly penultimate- and antepenultimate-stress targets. For the manipulation we proposed – prime-target pairs sharing both initial phonemes and stress vs. prime-target pairs sharing initial phonemes but not stress – we may sketch the following predictions: congruent primes should facilitate, and incongruent primes should inhibit, target articulation. The facilitation would

be brought about by the prime pre-activating either segments and/or stress (cf. Roelofs and Meyer, 1998) congruent with the target, while in the incongruent condition the stress mismatch would be enough to delay the articulation. In fact, if we assume – according to current computational models of polysyllable word reading (Perry et al., 2010) – that readers do not start articulation until stress has been fully activated – since only determining which syllable is stressed guarantees correct performance –, we may expect that the incongruency at the suprasegmental level may be sufficient to delay the articulation, irrespective of any overlap at the segmental level. Moreover, since previous stress priming studies have shown that stress priming effects seem not to be modulated by the word stress position (Sulpizio et al., 2012b), no difference is expected between penultimate- and antepenultimate-stress targets.

EXPERIMENT 1

In Experiment 1 we tested the MOPE in a reading aloud experiment with Italian penultimate- and antepenultimate-stress words as targets. We adopted the paradigm proposed by Schiller (2004; see also Schiller and Kinoshita, 2007), in which the target word (e.g., FEcola, 'starch') is preceded by an onset-related or -unrelated sequence (e.g., *fe*%%%; *mi*%%%). In this way, we are able to exclude any effect of suprasegmental material that, in case of a whole word prime (as, e.g., FEgato 'liver'), might be elicited by the activation of stress information. In addition, in order to establish the direction of the effect we also included a control condition that did not involve orthographic information.

The aim of Experiment 1, however, was not only to replicate previous studies showing that onset-related primes facilitate the computation of target phonology during reading aloud, but also to test whether a pure segmental prime may also activate suprasegmental information. In the onset-related condition, prime and target shared the first syllable as they were segmentally identical; however, the target syllable was either stressed (e.g., FEcola 'starch') or unstressed (e.g., feNIce 'phoenix'), and thus the prime syllable could or could not be congruent with the target first syllable for stress pattern. This allows us to propose two alternative predictions: first, if the prime affects an abstract phonological level of computation, such as the segment-to-frame association, then readers should be faster reading a target word in the onset-related condition than in either the onset-unrelated condition or the control condition (Schiller, 2004), and this should be true for both antepenultimate and penultimate stress words. Alternatively, if the prime affects the phonetic level of target computation – by activating a phonetic syllabic unit containing also information about stress – then we should expect different results for penultimate- and antepenultimate-stress targets. The reason for this is that penultimate-stress targets start with an unstressed syllable whereas antepenultimate-stress targets start with a stressed syllable. Thus, if the prime activates a stressed syllable, it might facilitate antepenultimate-, but not penultimate-stress targets; differently, if the prime activates an unstressed syllable,

³The remaining 2% of three-syllable words bear stress on the final syllable, and in this case stress is graphically marked (e.g., *colibri*).

it might facilitate penultimate-, but not antepenultimate-stress targets.

Method

Participants

Twenty-four students (six males, mean age: 23.33; *SD*: 4.73) from the University of Trento took part in the experiment. They received course credit for their participation. All participants were Italian native speakers with normal or corrected-to-normal vision. This and all the following experiments were carried out in accordance with the recommendations of the University of Trento ethics committee.

Materials

Targets were two sets of 24 three-syllable words each. One set comprised penultimate-stress words and the other antepenultimate-stress words. Words were selected from the CoLFIS database (Bertinetto et al., 2005) and were matched on: frequency, orthographic neighborhood size, orthographic neighbors' summed frequency, and bigram frequency (Table 1). Words in the two sets were also matched on their first syllable, i.e., for each word in a set there was a word in the other set starting with the same syllable as, e.g., FEcola 'starch' and feRIta 'wound.' All words were six letters long and had the same CVCVCV syllabic structure. All stimuli are listed in the Appendix.

Each target (e.g., FEcola 'starch') was preceded by three different primes: (i) a control condition, in which the prime consisted of a string of symbols (%%%%%); (ii) an onset-related condition, in which prime and target share the first syllable (e.g., fe%%%%); (iii) an onset-unrelated condition, in which prime and target differ in the first syllable (e.g., mi%%%%). Three different lists were created, and each target appeared once in each list in a different prime condition. Within each list the three prime conditions appeared the same number of times.

Procedure

Participants were tested individually. They were instructed to read the targets aloud as quickly and accurately as possible. No information was given about the presence of the primes, which was revealed only after the experiment.

The experiment was run using E-Prime software (Psychology Software Tools, Pittsburgh, PA, USA). Each target started with a

fixation cross, in the center of the screen, for 400 ms. The fixation cross was followed by a forward mask of hash marks (#), which was displayed for 500 ms in the center of the screen. The prime was then presented for 50 ms in lower-case letter, in the same location, followed by the target word, displayed in upper-case letters in the same position as the prime. The target remained on the screen until the participant began to read or for a maximum of 1,500 ms. A voice key connected to the computer measured reaction times (RTs) in ms from the onset of pronunciation.

The inter-stimulus interval was 1,500 ms. A short practice session preceded the experiment.

Each participant received all three lists, each list in a separate block separated by a short interval. Each block contained only one token of target and an equal number of the three prime-target pairs; the order of blocks was counterbalanced across participants and the order of prime-target pairs was randomized within each block. The experimenter noted the naming errors or apparatus failures on the fly.

Results

Responses shorter than 200 ms and invalid trials due to technical failures accounted for 1.3% of all data points and were discarded from the analyses; outliers (0.9% of all data points) were identified and removed following the Van Selst and Jolicoeur's (1994) procedure. Three items (PATina 'patina,' coLEra 'cholera,' Mitilo 'mussel,' all above 30% of errors) were also excluded from analyses due to the very high percentage of errors participants made. Naming errors were few (2.4% of all data points) and were not analyzed. Naming times were analyzed using mixed-effects models (Baayen et al., 2008). The models were fitted using the lmer function in R software. The models included prime type (related, unrelated, and control) and stress of the target (penultimate and antepenultimate) as fixed factors⁴. For the random factors, a maximal random structure approach was used (by participants and by items random intercepts and slopes; see Barr et al., 2013). The analysis started with a full factorial model including the main effects and the two-way interaction. The model was progressively simplified by removing the variables that did not significantly contribute to the goodness of fit of the model. Variables were evaluated one by one on the basis of likelihood ratio tests: those whose exclusion did not decrease significantly the model goodness of fit were removed from the analysis. Statistics of the best model are reported. Statistical significance of the fixed parameters was evaluated using the MCMC procedure, sampling 10,000 times (Baayen et al., 2008). Results are reported in Figure 1.

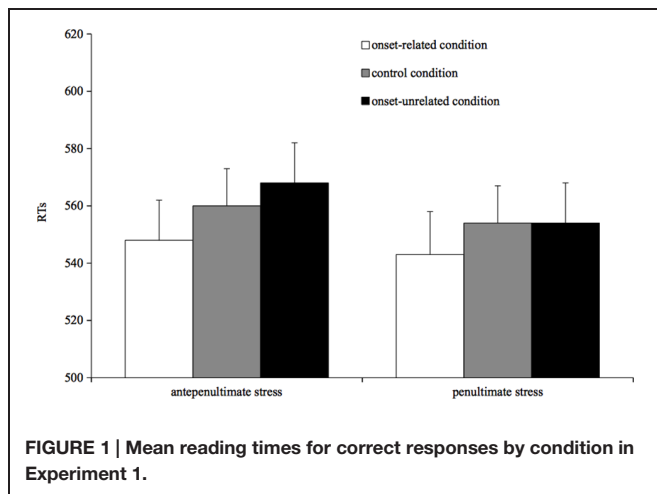
The full factorial model revealed that the prime type by stress of the target interaction was not significant, and it was dropped from the analysis as it did not significantly increase the model goodness of fit ($\chi^2 = 2.47, p > 0.2$). The reduced model revealed that prime type significantly affected reading of target words, with slower reading times for targets preceded either by unrelated primes ($\beta = 14.46, SE = 3.37, t = 3.61, p < 0.001$) or by control

TABLE 1 | Summary statistics: mean (and standard deviation) for target words used in Experiments 1–3.

Variables	Stress type	
	Antepenultimate	Penultimate
Word frequency	10.58 (17.42)	10.7 (17.63)
N of orthographic neighbors	3.66 (2.38)	3.29 (2.27)
Neighbors' frequency	25.38 (60.46)	17.49 (24.98)
Bigram frequency	11.33 (0.48)	11.31 (0.34)

Word frequency measures are calculated out of one million occurrences (Bertinetto et al., 2005); bigram frequency is log transformed on the basis of the natural logarithm.

⁴The analyses were also run with the block order as fixed factor. As the pattern of results was the same, we decided not to report such analyses. The analyses were also run with the block order as fixed factor. As the pattern of results was the same, we decided not to report such analyses. The same was done for all experiments.



primes ($\beta = 10.01$, $SE = 3.61$, $t = 2.76$, $p = 0.005$) than for targets preceded by related primes. The unrelated and the control conditions did not differ ($t = 1.22$, $p > 0.2$). The effect of target stress was not significant ($t = -1.25$, $p > 0.2$).

Discussion

The results of Experiment 1 show a clear effect of the segmental overlap on reading times: readers were facilitated in reading a target word in the onset-related condition in comparison to both the onset-unrelated and the control condition. The pattern goes in the same direction for penultimate- and antepenultimate-stress targets, suggesting similar processing in the computation of segmental information for both types of words.

The pattern we obtained is entirely compatible with Schiller's (2004; Schiller and Kinoshita, 2007) explanation: in the onset-related condition the prime pre-activates the initial phonological segments of the target at the level of phonological encoding. According to such a view, the active units are phonological segments and not phonetically specified syllabic units.

The analogous pattern obtained for penultimate- and antepenultimate-stress target supports this claim. In our experiment, the congruent prime always coincided with the first syllable of the target and, thus, the prime might have activated a syllabic unit rather than two phonological segments. However, in Italian a syllabic unit is realized in one of two different phonetic versions, i.e., as stressed or unstressed. Thus, the prime could have affected the target at a phonetic level, by activating a phonetically specified syllabic unit, which would also activate information about stress. This being the case, a different pattern for penultimate- and antepenultimate-stress targets would be expected since pre-activation of stressed syllables would facilitate reading antepenultimate-stress targets (which start with a stressed syllable) but not penultimate-stress targets (which start with an unstressed unit) and pre-activation of unstressed syllables would lead to the opposite pattern. The results of our experiment showing a parallel pattern for both penultimate- and antepenultimate-stress words suggest that the prime exerts its effect at an abstract phonological level, with a benefit for onset-overlapping

targets during the word phonological encoding (Schiller, 2004)⁵.

In Experiment 2 we investigated the effect of suprasegmental priming on the phonological encoding of the word using the same set of target words of Experiment 1.

EXPERIMENT 2

The aim of the present experiment was to establish whether the masked stress priming is effective in generating a stress priming effect, and whether such an effect is facilitatory or inhibitory in nature. The stress priming effect reported by previous studies has never been tested against a control condition (Colombo and Zevin, 2009; Sulpizio et al., 2012a,b), with the consequence that it is still unclear whether priming the metrical structure of a word facilitates or inhibits reading it aloud. Moreover since all aforementioned studies adopted a visible priming technique – in which readers explicitly processed the prime – it cannot be excluded that the effect of stress priming they reported may have a strategic component. To rule out this hypothesis, we used the masked priming paradigm with prime-target pairs that differed at the segmental level but did or did not share the metrical structure. In this way, we would be able to assess whether primes sharing or not sharing stress with the targets (i.e., the congruent vs. incongruent condition) affect target reading, with respect to a non-linguistic control condition, over and above any effect due to the prime and target mismatch at the segmental level.

Method

Participants

Twenty-four student (four males, mean age: 20.26; SD : 1.99) from the University of Trento took part in the experiment. They received course credit for their participation. All participants were Italian native speakers with normal or corrected-to-normal vision.

Materials

The same target words of Experiment 1 were used. Prime words had the same syllabic length and structure of the targets. Penultimate- and antepenultimate-stress prime words were matched on frequency, orthographic neighborhood size, orthographic neighbors' summed frequency, and bigram frequency (Table 2). All stimuli are listed in the Appendix. Primes and targets were paired in such a way as to obtain three prime conditions for each targets: a stress congruent condition, with prime and target sharing the same stress pattern (e.g., *CI*nema – *FE*cola, 'cinema' – 'starch'); a stress incongruent condition, with prime and target bearing a different stress (e.g., *ca*NAle – *FE*cola, 'channel' – 'starch'); and a control condition, in which the target word was preceded by a string of symbols (e.g., %%%%%%%%% – *FE*cola, 'starch'). Primes and targets were not semantically related and never shared the initial syllable.

⁵The activation of phonetic rather than phonological units would be compatible with our pattern only by assuming that stressed and unstressed syllables are each activated and roughly at the same time, with the additional assumption that the

TABLE 2 | Summary statistics: mean (and standard deviation) for prime words used in Experiments 2 and 3.

Variables	Stress type	
	Antepenultimate	Penultimate
Word frequency	48.41 (105.33)	44.08 (59.39)
N of orthographic neighbors	3.37 (2.08)	3.41 (2.16)
Neighbors' frequency	15.26 (21.98)	29.82 (38.11)
Bigram frequency	11.3 (0.33)	11.42 (0.39)

Word frequency measures are calculated out of one million occurrences (Bertinetto et al., 2005); bigram frequency is log transformed on the basis of the natural logarithm.

Procedure

The same procedure as in Experiment 1 was adopted.

Results

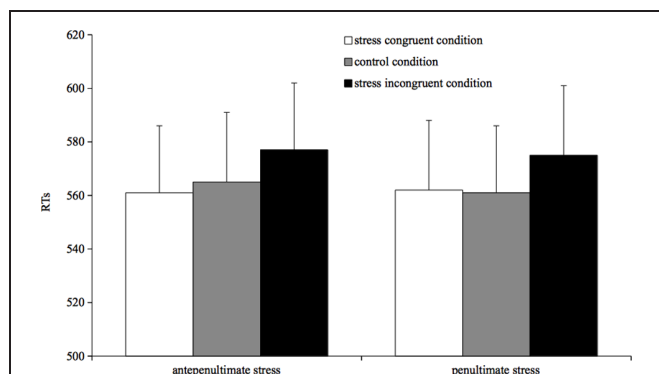
Responses shorter than 200 ms or longer than 1500 ms as well as invalid trials due to technical failures accounted for the 2.6% of all data points and were discarded from the analyses; outliers (1% of all data points) were identified and removed using the Van Selst and Jolicoeur's (1994) procedure. Due to the high number of errors, two items (PATina 'patina,' Mitilo 'mussel,' above 30% of errors) were excluded from analyses. Naming errors were few (2.5%) and were not analyzed.

Naming times were analyzed using mixed-effects models (Baayen et al., 2008). Results are reported in **Figure 2**.

The model was run with RTs as dependent variable and prime type (congruent stress, incongruent stress, and control) and stress target (penultimate and antepenultimate) as fixed factors.

The full factorial model revealed that the prime type by stress target interaction was not significant, and as it did not significantly increase the model goodness of fit ($\chi^2 < 1$) it was dropped from the analysis. The simplified model showed that prime type significantly affected target reading times: participants were slower when reading targets preceded by incongruent stress

stress-inconsistent syllable does not interfere. This alternative cannot be totally ruled out, but it seems unlikely.

**FIGURE 2 | Mean reading times for correct responses by condition in Experiment 2.**

primes than preceded by both congruent stress primes ($\beta = 10.49$, $SE = 4.47$, $t = 2.34$, $p = 0.01$) and control primes ($\beta = 9.19$, $SE = 4.51$, $t = 1.91$, $p = 0.05$). No difference was found between targets preceded by congruent stress primes and by control primes ($t < 1$). No main effect of stress target was found ($t < 1$).

Discussion

The pattern shown by the analyses of reading times is clear: readers are slower when reading a target word preceded by a prime bearing a different stress pattern than a target preceded by a prime bearing the same stress pattern or by a control prime. Moreover, the stress prime effect is not affected by the type of word stress pattern as revealed by the absence of a prime type by stress type interaction.

The results of Experiment 2 replicate findings on stress priming reported previously (Sulpizio et al., 2012b), but they add new insights about the computation of stress in reading. In particular, the finding of a stress priming effect when the prime is masked not only corroborates the view that the metrical structure of a word may be primed independently from its segmental content, but also suggests that the word stress pattern is automatically activated by lexical computation as well as by segmental phonological information.

With regard to the nature of the priming effect, our results show that target words preceded by stress-incongruent primes were read more slowly than those preceded by stress-congruent primes, and this was true for both penultimate- and antepenultimate-stress targets. Thus, the findings extend Sulpizio et al.'s (2012b) results by showing that the stress priming effect on naming times is automatic, i.e., it is not driven by strategic mechanisms, since it emerges also when readers are not aware of the presence of primes. As previous works suggest (Colombo and Zevin, 2009; Sulpizio et al., 2012b), the locus for the stress priming effect is the stage of phonological encoding.

Note that the prime-target pairs of Experiment 2 always differed at the segmental level. It might be argued that such segmental mismatch might have contributed to the pattern we found, as the phonological segments activated by the prime could have interfered with the selection of the segments of the target. However, if that were the case, the effect of segmental mismatch would have been visible also in the congruent-stress condition, with slower reading times than the control condition, which is clearly not the case. The absence of segmental inhibition is also in line with the results of Experiment 1, where there was no difference between targets in the control and in the segmentally incongruent condition, and reinforces the idea that, under our experimental conditions, segmental information may facilitate but does not hinder word processing (e.g., Schiller and Kinoshita, 2007).

Taken together, the results of Experiments 1 and 2 show an interesting asymmetry: a segmental prime without stress information speeds up the reading of a segmentally consistent target; a stress prime, keeping segmental information (incongruently) constant, slows down the reading of stress inconsistent targets. This is *prima facie* evidence that segmental and suprasegmental information affect word reading independently and with an opposite pattern. Moreover, in both

Experiments 1 and 2 antepenultimate- and penultimate-stress targets were similarly affected by segmental and suprasegmental priming; following previous research, both the facilitation for prime-target segmental overlapping pairs and the inhibition for prime-target incongruent stress pairs may be located at the level of phonological output buffer when the segment-to-frame association takes place (Kinoshita, 2000; Sulpizio et al., 2012b).

In Experiments 3 and 4 we further tested how readers encode the phonological word by jointly manipulating the overlap between stress and phonemes in prime/targets pairs. To our knowledge, this issue has never been investigated in the reading literature, in spite of being crucial for any model of polysyllabic word reading.

EXPERIMENT 3

In this experiment we investigated the processes of segment-to-frame association by directly testing how readers assemble the phonological segments with the stress metrical structure of the word they have to produce. We used the same target words of the two previous experiments, and varied the degree of overlap of segmental and suprasegmental information between primes and targets. To illustrate, each target (e.g., FEcola, 'starch') could be preceded by: (a) a congruent prime, in which prime and target shared both the first syllable and the stress pattern (e.g., FEgato, 'liver'); (b) an incongruent prime, in which prime and target shared the first syllable but not the stress pattern (e.g., feNIce, 'phoenix'); (c) a control prime, i.e., a sequence of symbols (%%%%%%%%). The incongruent prime condition is the critical one. In fact, although both congruent and incongruent primes would cause pre-activation of the segmental level, in the latter case the pre-activated phonemes might not be associated to the correct metrical frame until the stress pattern has been identified (cf. Perry et al., 2010). This would interfere with the segment-to-frame association and with the processes occurring further down stream by delaying the selection of the correct metrical frame and its association with the phonological segments and the planning of articulation. No such delay would occur in the congruent prime condition, where the pre-activation of both the initial phonemes and the correct stress pattern would speed up articulation.

Method

Participants

Twenty-four students (11 males, mean age: 28, *SD*: 5.57) took part in the experiment. None participated to both experiments. Participants were all from the University of Trento and received course credit for their participation. All participants were Italian native speakers with normal or corrected-to-normal vision.

Materials

The same target and prime words of Experiment 2 were used. However, the pairing of primes and targets was modified in order to obtain three conditions: 16 prime-target pairs sharing both the initial syllable and the stress pattern (e.g., FEgato–FEcola, 'liver' – 'starch'); 16 pairs sharing the same initial syllable but having a

different stress pattern (e.g., feNIce – FEcola, 'phoenix' – 'starch'), and 16 pairs not sharing either segmental or stress information (control condition; e.g., %%%%%%%%% – FEcola, 'starch'). Primes and targets were never semantically related. Three different lists were created, so that each target appeared only once in each list in a different prime condition. Within each list the three prime conditions appeared the same number of times.

Procedure

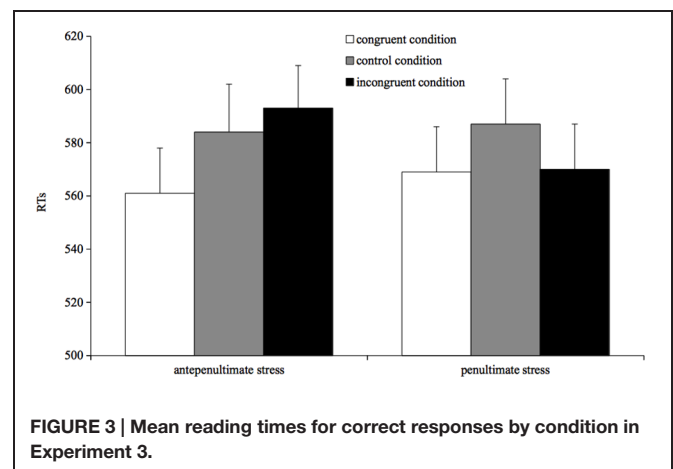
The same procedure as in Experiment 1.

Results

Responses shorter than 200 ms or longer than 1500 ms as well as invalid trials due to technical failures accounted for the 2.2% of all data points and were discarded from the analyses; outliers (2.5% of all data points) were also removed using the Van Selst and Jolicoeur's (1994) procedure. Due to its high number of error (above 30%), one item (PATina 'patina') was removed and not further considered in the analyses. Naming errors were few (2.9%) and were not analyzed.

Naming times were analyzed using mixed-effects models (Baayen et al., 2008). Results are reported in **Figure 3**.

The model was run with RTs as dependent variable and prime type (congruent, incongruent, and control) and stress target (penultimate and antepenultimate) as fixed factors. The prime type by stress target interaction was significant ($\beta = -14.03$, $SE = 6.78$, $t = -2.10$, $p = 0.03$), showing that the three primes affected antepenultimate- and penultimate-stress targets differently. Direct comparisons between conditions were assessed through separate analyses on the two types of targets. For antepenultimate-stress targets, the model showed that participants were faster in reading targets when preceded by congruent primes than when preceded by either incongruent primes ($\beta = 17.58$, $SE = 4.88$, $t = 3.59$, $p < 0.001$) or control primes ($\beta = 11.94$, $SE = 4.92$, $t = 2.42$, $p = 0.01$). Incongruent and control conditions did not differ from each other ($t = 1.13$, $p > 0.2$). A different pattern was found for penultimate-stress words: participants were faster in reading a penultimate-stress target when preceded by a congruent prime than when preceded by a control prime ($\beta = 14.92$, $SE = 4.68$, $t = 3.18$, $p = 0.001$);



surprisingly, participants were also faster in reading a target when preceded by an incongruent prime than when preceded by a control prime ($\beta = 15.35$, $SE = 4.69$, $t = 3.26$, $p = 0.001$). No difference was found between congruent and incongruent prime condition ($t < 1$).

To sum up, the effect of the incongruent prime condition on naming speed appears to be asymmetric: it does not differ from the control condition for antepenultimate-stress targets, but it is facilitatory for penultimate-stress targets.

Discussion

The results of Experiment 3 show that penultimate- and antepenultimate-stress targets are processed more rapidly in the congruent than in the control condition. However, the two types of targets differ in the incongruent prime condition: reading times to incongruent antepenultimate-stress targets and to control targets do not differ, and both are read more slowly than congruent antepenultimate-stress targets; however, incongruent penultimate-stress targets are read as quickly as congruent penultimate-stress targets and both are processed more rapidly than control targets. Thus, the incongruent prime condition hinders responses to antepenultimate-stress targets but does not affect penultimate-stress targets.

The results for penultimate stress targets show that the overlap of segmental information between prime and target is sufficient to drive a process for incongruent targets that is quantitatively analogous to that driven by the congruent condition, in which primes and targets overlap for suprasegmental as well as segmental information. This result for penultimate-stress words is not only sharply different from the pattern obtained for antepenultimate-stress words in the same condition, but it is also quite in contrast with the pattern obtained in Experiment 2, where slower reading times were obtained for incongruent pairs when prime-target pairs had different stress but were also entirely different at the segmental level.

For antepenultimate-stress targets, however, the data pattern differently: targets in the incongruent prime condition are read more slowly than targets in the congruent prime condition. To understand such pattern we may look at the first two experiments, which show that the prime-target segmental congruency speeds up responses (Experiment 1), whereas the prime-target suprasegmental incongruency slows down the reading times (Experiment 2). Thus, in Experiment 3, in which the two factors were jointly manipulated, the actual pattern for the incongruent condition could be the outcome of the combination of the segmental overlap and the suprasegmental mismatch. Specifically, segmental match speeds up frame to segment association, but the concurrent presence of incongruent stress information slows down such process, with the result that the two effects cancel out. Thus, the crucial aspect becomes how the system incorporates congruent and incongruent segmental and suprasegmental information in time.

The asymmetry between antepenultimate- and penultimate-stress targets we obtained for the incongruent condition in Experiment 3 lends itself to several possible interpretations. The processing account we provide below seems to us to be both

empirically consistent and theoretically valid, but further data will be necessary to rule out alternative accounts.

We ascribe the asymmetric pattern to the operations that take place at the level of phonological output buffer, where lexical and sub-lexical routes converge and the system pools together the information coming from the two routes to drive the stimulus pronunciation; in our view, the buffer comprises a system for phonemic activation and one for stress assignment (for a similar proposal, see Perry et al., 2010). Within the phonological output buffer, we assume that the segment-to-frame association – i.e., the association of phonemes to a metrical frame – and the phonological-to-phonetic mapping – i.e., the mapping of abstract linguistic information into motor commands – take place rightward incrementally (cf. Kinoshita, 2000). Thus, for the first syllable of three-syllable antepenultimate-stress words there is activation of both its phonemes and the stress pattern while for the first syllable of penultimate-stress words there is activation of its phonemes while it is the second syllable that requires the activation of both its phonemes and the stress pattern. Accordingly, we assume that, during the segment-to-frame association, the stress system specifies the tonic syllable among the available segmental material: specifically, at the level of phonological output buffer, once enough evidence (coming from lexical and sub-lexical route) for a stress pattern is available, the stress system specifies which syllable should be articulatory implemented as stressed. Note that for reading there may not be the need to specify information about the number of syllables, as in word reading, the number of syllables and their internal organization may be inferred by orthography, with the system able to arrange the identified letters (or group of letters) into a graphosyllabic representation (see, e.g., Caramazza and Miceli, 1990; Perry et al., 2007, 2010; see also Chetail et al., 2014 for evidence that the structure of a letter string can be determined simply on the basis of consonant and vowel identification). Furthermore, we assume that the reading system starts the planning of articulation as soon as the relevant information for the to-be-planned unit is active. We may call this the use-information-as-soon-as-possible (UIASAP) approach. That is to say, within the phonological output buffer, as soon as usable information becomes available it is incorporated in the open frame the system builds for the phonetic encoding of the stimulus, which then addresses the motor programs to execute articulation. This would yield different patterns for antepenultimate- and penultimate-stress targets in the incongruent condition: for the former, articulation may start as soon as the first syllable is encoded, since both segmental and suprasegmental information is already available; for the latter, articulation needs to wait up to the second syllable since information about stress becomes available at that point. The inconsistent stress prime would differently affect the two types of words as a function of the unit to be processed. For antepenultimate-stress words the interference would be stronger as it would impact on the to-be-articulated syllable. For penultimate-stress words, however, there would be time to mitigate the impact of the incongruent stress prime since articulation cannot start until the information about stress becomes available on the second syllable; this being the case, the

system might capitalize on the available segmental information, which is not affected by the suprasegmental-stress mismatch, performing in a similar manner for targets with congruent- and incongruent primes.

An alternative explanation of our results may rest on the distributional asymmetry of the two stress patterns. In Italian, 80% of three-syllable words bears penultimate stress while 18% bears antepenultimate (Thornton et al., 1997), and it might be argued that the penultimate stress pattern would work as a default pattern (Colombo, 1992). Thus, penultimate stress would reach the activation level quite easily, with low chance to be interfered with by any other pre-activated, less frequent stress pattern. The antepenultimate-stress pattern would show the opposite picture, as it is less represented in the lexicon and it would need a high activation level to be selected; as a consequence, it would have a high probability to be interfered with by the partial activation of the penultimate-stress (default) pattern. According to this distributional view, the asymmetry we found for penultimate- and antepenultimate-stress targets in Experiment 3 would be fully accountable for by the different weight the two stress patterns have in the reading system, the former being the default, more available pattern.

The Italian lexicon offers a good test to adjudicate between the UIASAP and the distributional pattern hypotheses, i.e., three-syllable words with final stress, which is the least frequent stress patterns (around 2% of three-syllable words). Thus, in Experiment 4 we performed a critical test of the two alternative accounts by using final-stress words as targets, i.e., words bearing stress on the last syllable, which is orthographically marked (e.g., *colibri*, *hummingbird*). Note that for these words the suprasegmental information may be computed sub-lexically, as the accent mark may directly activate the corresponding stress pattern. However, as in other domains of orthographic processing, we think that the system always engages in lexical as well as non-lexical processing (see, e.g., Peressotti et al., 2003). Thus, we think that final-stress words are a very good test for the UIASAP hypothesis.

The distributional pattern and the UIASAP hypothesis make opposite predictions about the pattern the final-stress words should elicit. If the different pattern of results found for antepenultimate- and penultimate-stress words is due to their distributional properties, then we expect final-stress words to behave as the antepenultimate-stress words, since both are rare patterns in the language. This being the case, for antepenultimate-stress and final-stress words we expect both incongruent and control condition to be slower than the congruent condition, and not to differ from each other. On the other hand, if the difference between antepenultimate- and penultimate-stress pattern is due to left-to-right processing, as assumed by the UIASAP proposal, then we may expect the final-stress words to pattern with the penultimate-stress words. In particular, antepenultimate-stress words should show slower reading times in the incongruent than in the congruent condition, while both penultimate-stress and final-stress words should show similar reading times in the congruent and incongruent condition due to the earlier availability of the segmental information and to the absence of mismatching stress information on the

first syllable that allows for the articulation of the word to begin.

In Experiment 4 we used a new set of stimuli comprising final-stress words as well as new penultimate- and antepenultimate-stress words. The aim of the experiment is twofold: first, to test for the replicability and robustness of the effect we found in Experiment 3; second, to adjudicate between the UIASAP and the distributional pattern hypotheses.

EXPERIMENT 4

Method

Participants

Twenty student (five males, mean age: 23.7; *SD*: 4.9) from the University of Trento took part in the experiment. They received course credit for their participation. All participants were Italian native speakers with normal or corrected-to-normal vision.

Materials

Three sets of three-syllable words were selected as targets. One set included final-stress words, one set penultimate-stress words, and one set antepenultimate-stress words. No stimulus was a target in any of the previous experiments. Of the 56 targets, half were final-stress words (mean frequency: 79.85 occurrence per million) and the other half was equally divided between penultimate- and antepenultimate-stress words (mean frequency: 100.71 and 43.50 occurrence per million, respectively). Words were selected from the CoLFIS database (Bertinetto et al., 2005). The final stress targets included 19 words of six letters and nine words of seven letters. The penultimate- and antepenultimate-stress words were all six letters in length and had a CVCVCV syllabic structure.

As in Experiment 3, for each target (e.g., *REsina*, ‘resin’) there were three primes: (i) a word sharing the initial syllable and the stress pattern with the target (e.g., *REgola*, ‘rule’); (ii) a word sharing the initial syllable but not the stress pattern with the target (e.g., *reGIme*, ‘regime’); (iii) a string of symbols (% % % % %). In condition, the incongruent-stress condition, primes for final-stress targets were either penultimate-stress words (14/28) or antepenultimate-stress words (14/28), for penultimate-stress targets primes were antepenultimate-stress words, and for antepenultimate-stress targets primes were penultimate-stress words.

The sets of prime words were matched on: frequency, orthographic neighborhood size, orthographic neighbors’ summed frequency, bigram frequency (Table 3). Primes and targets were not semantically related. All stimuli are listed in the Appendix. Three different lists were created, with each target appearing only in one list in a different prime condition.

Procedure

The same procedure as in Experiment 1 was adopted.

Results

Responses shorter than 200 ms or longer than 1500 ms as well as invalid trials due to technical failures accounted for the 2.9% of all data points and were discarded from the analyses; outliers

TABLE 3 | Summary statistics: mean (and standard deviation) for prime words used in Experiment 4.

Variables	Antepenultimate stress targets		Penultimate stress targets		Final stress targets	
	Congruent stress prime	Incongruent stress prime	Congruent stress prime	Incongruent stress prime	Congruent stress prime	Incongruent stress prime
Word frequency	73.57 (121.41)	58 (82.31)	96.42 (129.4)	91.07 (159.72)	2.35 (3.58)	3.64 (7.01)
Letters length	6 (0)	6 (0)	6 (0)	6 (0)	6.32 (0.61)	6.46 (0.69)
N of orthographic neighbors	3.14 (1.91)	4 (1.88)	3.92 (2.52)	3.14 (1.4)	2.42 (1.23)	2.96 (1.62)
Neighbors' frequency	43.41 (67.31)	15.35 (14.29)	30.69 (40.3)	32.91 (29.47)	14.95 (27.53)	31.34 (61.87)
Bigram frequency	11.74 (0.15)	11.79 (0.32)	11.7 (0.24)	11.71 (0.3)	11.65 (0.28)	11.62 (0.31)

Word frequency measures are calculated out of one million occurrences (Bertinetto et al., 2005); bigram frequency is log transformed on the basis of the natural logarithm.

(0.3% of all data points) were also removed using the Van Selst and Jolicoeur's (1994) procedure.

Participants did few naming errors (2.7% all data points) and were not analyzed.

Naming times were analyzed using mixed-effects models (Baayen et al., 2008). Results are reported in **Figure 4**.

The full factorial model was run with RTs as dependent variable and prime type (congruent, incongruent, and control) and stress target (antepenultimate, penultimate, and final stress) as fixed factors. The model showed that prime type and stress target interacted, and that the effect of prime type on antepenultimate-stress targets differed from that the prime type had on both penultimate- ($\beta = -18.44$, $SE = 9.39$, $t = -1.96$, $p = 0.04$) and final-stress targets ($\beta = -16.11$, $SE = 8.18$, $t = -1.96$, $p = 0.04$). No effect of stress target was reported for the control condition (antepenultimate vs. penultimate: $t < 1$; antepenultimate vs. final: $t < 1$; penultimate vs. final: $t < 1$). To further explore the interaction, we run separate analyses on the three types of targets.

Antepenultimate Stress Targets

The pattern parallels that of Experiment 3. Specifically, participants were faster in reading a target preceded by congruent primes than preceded by a control primes ($\beta = 19.18$, $SE = 6.73$, $t = 2.84$, $p = 0.004$), and were faster in reading a target preceded by a congruent prime than preceded by an incongruent prime

($\beta = 17.68$, $SE = 6.86$, $t = 2.57$, $p = 0.01$). The incongruent prime and the control prime condition did not differ from each other ($t < 1$).

Penultimate Stress Targets

Again, the pattern parallels that of Experiment 3: participants were faster in reading a target preceded by a congruent prime than preceded by a control prime ($\beta = 14.36$, $SE = 6.33$, $t = 2.26$, $p = 0.02$), and were faster in reading a target preceded by an incongruent prime than preceded by a control prime ($\beta = 14.97$, $SE = 6.40$, $t = 2.33$, $p = 0.01$). No difference was found between the congruent and incongruent prime condition ($t < 1$).

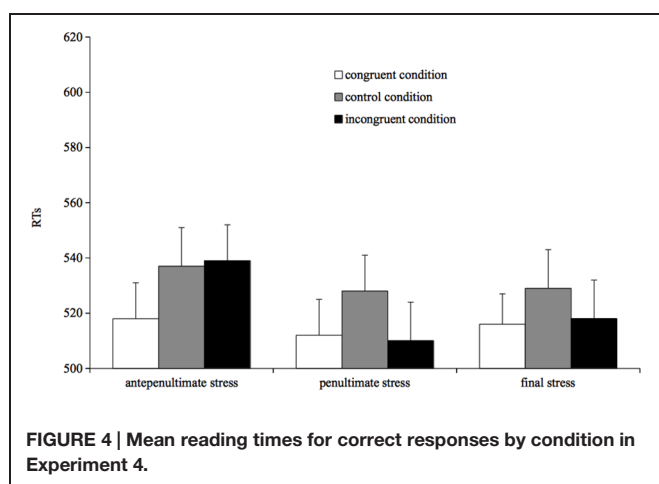
Final Stress Targets

Participants were faster in reading a target preceded by a congruent prime than preceded by a control prime ($\beta = 10.38$, $SE = 4.65$, $t = 2.23$, $p = 0.02$), and were faster in reading a target preceded by an incongruent prime than preceded by a control prime ($\beta = 8.54$, $SE = 4.66$, $t = 1.83$, $p = 0.06$). No difference was found between congruent and incongruent prime condition ($t < 1$).

Discussion

The results of Experiment 4 are straightforward: for penultimate- and antepenultimate-stress targets we replicated the pattern found in Experiment 3, with a facilitation for both congruent and incongruent targets for penultimate-stress words, and a dissociation between congruent and incongruent targets for antepenultimate-stress words, the former being faster than the control and the latter being as slow as the control. This pattern strengthens the results of Experiments 3 generalizing them to a new set of stimuli.

The novel result is that final-stress targets show a pattern identical to penultimate-stress words, that is to say they show similar reading times for both stress congruent and stress incongruent targets (both faster than the control condition). The analogous pattern found for penultimate- and final-stress targets suggests that the difference between antepenultimate-stress words, on the one hand, and penultimate- and final-stress words, on the other hand, is not a consequence of distributional differences among the stress patterns, the latter (final stress) being much less frequent than the former (with 2% of words bearing final stress vs. 18% of words bearing antepenultimate stress). Instead, the asymmetry in the incongruent condition is more



consistent with the UIASAP proposal that ascribes the difference to the way in which the operations within the phonological output buffer take place during reading aloud.

GENERAL DISCUSSION

In four reading aloud experiments, using a masked priming paradigm, we investigated the timing of the operations that occur in the phonological output buffer in order for readers to assemble segmental and suprasegmental information for articulating the word phonological form. Across experiments, we manipulated the degree of segmental and suprasegmental overlap between prime-target pairs of three-syllable Italian words varying in stress position. The results shed new light on several issues relevant for the understanding of how the stage of segment-to-frame association and phonological-to-phonetic mapping takes place in polysyllabic word reading.

The effect of segmental overlap we found in Experiment 1 is a robust, often replicated effect (see, e.g., Kinoshita, 2000; Schiller, 2004; Malouf and Kinoshita, 2007; Schiller and Kinoshita, 2007; Dimitropoulou et al., 2010). Since the effect emerges for orthographically dissimilar but phonologically similar prime-target pairs, and not for phonologically dissimilar but orthographically similar prime-target pairs, it has been ascribed to the stage of phonological encoding: an onset-congruent prime speeds up target reading by facilitating the segment-to-frame association process, which proceeds rightward incrementally and may thus benefit from a segmental phonological pre-activation occurring at the beginning of the word (Schiller, 2004). The results of Experiment 1 are consistent with the previous studies and provide further evidence that the prime onset overlap affects target reading at an abstract phonological level, before the articulatory programs are addressed. The claim follows from the fact that the same segmental prime (e.g., fe%%) affected equally penultimate-stress targets (e.g., feRIta, 'wound') – whose first syllable is unstressed – and antepenultimate-stress targets (e.g., FEcola, 'starch') – whose first syllable is stressed. As syllables may be phonetically implemented as either stressed or unstressed, the implication is that the segmental prime activates the graphemes up to their phonological representation without specifying any phonetic detail. In fact, by using primes consisting of two graphemes (letters), wholly overlapping or wholly not overlapping with the first syllable of the target, followed by a sequence of %, we were effective in withholding information about lexical stress from the prime (i.e., whether a word bears penultimate or antepenultimate stress), but we also withheld information about syllabic stress (i.e., whether, at the articulatory level, a syllable has to be implemented as stressed or not). Had the computation of the prime proceeded up to the phonetic encoding, the prime segments should have activated either the stressed or the unstressed version of the corresponding syllabic unit, and thus different results should have been expected for penultimate- and antepenultimate-stress targets, with e.g., an unstressed syllable facilitating the reading of penultimate-stress targets (which start with an unstressed syllable), but not the

reading of antepenultimate-stress targets (which start with a stressed unit) and *vice versa*. This is clearly not what we found, since the same pattern characterized both penultimate- and antepenultimate-stress words, and this is an indication that the segmental prime exerts its effect at an abstract phonological level.

Penultimate- and antepenultimate-stress targets exhibited the same pattern also when we manipulated stress priming. For both types of targets, reading times in Experiment 2 were slower in the incongruent than in the congruent (and the control) condition. Previous studies on word reading with visible primes have shown that the metrical structure of a word may be primed independently from its segmental content, and that such a priming occurs for both penultimate- and antepenultimate-stress targets. The results were taken as evidence that an abstract representation of the words' metrical structure is available during reading, and can intervene during the stage of word phonological encoding (cf. Colombo and Zevin, 2009; Sulpizio et al., 2012b; Sulpizio and Job, 2013). The results of Experiment 2 allow us to better qualify those findings, by showing that: (a) the stress priming effect may be interpreted as inhibitory on the bases of the present experiment that included a control condition; (b) the effect is automatic, i.e., it is not driven by strategic mechanisms, since it emerges with masked primes as well.

We posit that the prime-target stress interference arises within the phonological output buffer of the reading system, and postulate a mechanism that activates stress information and specifies the position the stress takes in the word – for Italian words the three possibilities being either the antepenultimate, the penultimate, or the final syllable. During word reading, in the phonological output buffer, information about stress position coming from lexical and sub-lexical processing is collected and, as soon as activation for a stress pattern is reached, the system specifies the stressed syllable among those available. As for the time dynamics of the stress mismatch interference, when the stress pattern activated by the prime differs from the stress pattern required by the target there is a delay in specifying the position of stress within the available segmental sequence (i.e., the segment-to-frame association), since the currently available, incorrect stress pattern must be disengaged and the correct one must be activated. This proposal is similar to that put forward by Perry et al. (2010), who implemented a detailed system for stress assignment in their CDP++ model of bi-syllabic reading that we will further discuss below (see also Perry et al., 2014).

The finding of a pure metrical priming stands in contrast with the results reported by Roelofs and Meyer (1998) for speech production. Using a form-priming paradigm with Dutch words, unlike the present experiment, Roelofs and Meyer (1998) did not find a pure stress priming effect, and argued that the absence of such effect follows from the fact that metrical and segmental spell-out run in parallel and take the same amount of time (Roelofs and Meyer, 1998). Methodological differences between the two studies may account for the pattern. In the implicit form-priming paradigm, participants are required to learn cue-target word pairs, and to produce the target word upon presentation of a

cue word. This differs from the present research that investigated reading aloud by adopting a masked priming procedure, in which participants have to read stimuli aloud within the frame activated by the prime information. Thus, the discrepancies in the results may reflect the processes involved in performing the tasks. In particular, in Roelofs and Meyer's (1998) study episodic memory is heavily involved.

The joint manipulation of segmental and suprasegmental information of Experiments 3 and 4 shows an asymmetric pattern between antepenultimate-stress words, on the one hand, and penultimate-stress and final-stress words, on the other, and this allows us to better articulate the operations carried out by the phonological output buffer of the reading system. Moreover, such findings allow us to rule out the possibility that the dissociation is due to the asymmetric distribution of the stress patterns in Italian as final-stress words are quite infrequent, thus patterning with antepenultimate-stress words on this dimension, but showing no difference between the congruent and the incongruent condition, just like penultimate-stress words. As a consequence, the asymmetry among different stress patterns has to be ascribed to the temporal dynamics of the operations the reading system carries out for the stimulus.

The pattern we found may be accounted for by the UIASAP proposal, which makes three assumptions about the functioning of the phonological output buffer: (a) for words with unpredictable stress, the phonological encoding requires specifying which of the syllables receives stress among the available segmental material (see Roelofs, 2015) (b) the phonological-to-phonetic mapping takes place through a rightward incremental process (Levelt et al., 1999; for the same proposal in reading: Kinoshita, 2000; cf. Carreiras et al., 2005), with the minimal planning unit that goes from the word beginning up to (at least) the stressed syllable; (c) the reading system starts the planning of articulation as soon as the relevant information for the to-be-planned unit is active. Taken together these assumptions allow for the different temporal dynamics in the phonological output buffer for stress-related word classes. For antepenultimate-stress words, the first syllable comprises the activation of both its phonemes and the stress pattern, the latter being specified in the segmental sequence by the stress system; instead, for penultimate-stress words the first syllable comprises the activation of only its phonemes and it is the second syllable that requires the activation of both its phonemes and the stress pattern; for the final stress, the syllable requiring stress activation will be the last one. If information about the stressed syllable is needed to start articulation, for antepenultimate-stress targets articulation may start as soon as the first syllable is encoded since both segmental and suprasegmental information is available, while for penultimate- and final-stress targets articulation may only begin when information about the second or the third syllable, respectively, becomes available. Therefore, the inconsistent stress prime would affect differently the three types of words: for antepenultimate-stress words the interference would be stronger as it would directly impact on the to-be-articulated unit while for penultimate- and final-stress words there would be time to mitigate the impact of the incongruent

stress prime for articulation cannot start until the information about stress become available on the second or third syllable, respectively.

The UIASAP proposal would predict an advantage for stimuli bearing earlier stress compared with those stimuli bearing later stress. The empirical evidence on this issue is scanty, with studies generally reporting contrasting evidence on reading times of penultimate- and antepenultimate-stress words. Recently, in a pseudoword reading study, an advantage for antepenultimate-over penultimate-stress targets has been reported by Sulpizio et al. (2015), who found that participants read pseudowords faster when they assigned antepenultimate than penultimate stress. Sulpizio et al. (2015) proposed that stress computation affects naming speed at the stage of articulatory planning, as readers may buffer a partial articulatory representation of stimuli that proceeds from the first syllable up to the stressed syllable (for a similar perspective, see also Sternberg et al., 1988; Laudanna et al., 1989; Sulpizio and Colombo, 2013; Sulpizio et al., 2013). For words, a similar result has been reported by Burani and Arduino (2004, Experiment 2), who showed that low-frequency antepenultimate stress words were read faster than low-frequency penultimate stress words. Note, however, that an opposite pattern has also been reported (Colombo, 1992). Finally, Burani et al. (2014) reported no difference between words with penultimate and antepenultimate stress. Thus, an antepenultimate-stress advantage appears to be elusive and difficult to detect.

Although a difference between antepenultimate- and penultimate-stress targets might have been expected in our study, we believe that there are at least two reasons to account for its absence. First, in our study the system has to process a prime-target event instead of a simple target event, with the consequence that the operations involved in the former are partly different from those involved in the latter (Kinoshita and Norris, 2012). Specifically, the "disengagement" from the prime might (globally) interfere more with words that can be articulated faster than with words that requires more time to be articulated. Therefore, the prime-target computation may obscure or take away the possible advantage of antepenultimate words. Second, and more generally, the process of lexicalization (reading aloud) is affected by several concurrent factors that weight differently for the different stress patterns. Thus, the presumed advantage for antepenultimate stress may be diminished or eliminated by the distributional properties of stress (80% of polysyllable words bear penultimate stress) and/or explicit stress marks (available only for final stress words).

A computational account for our results may be offered by the CDP++ Italian model of polysyllable reading (Perry et al., 2014), which is the Italian version of the CDP++ (Perry et al., 2010). The model implements a detailed phonological output buffer composed of two distinct mechanisms for segmental and suprasegmental computation, i.e., *Phonological Output Nodes* (henceforth PONs) and *Stress Output Nodes* (henceforth SONs). Both PONs and SONs receive activation from the lexical and the sub-lexical route in parallel and combine the two sources of information through competitive interactions. The activation within the SONs is also regulated by a lateral inhibition parameter

and the activation of a stress node inhibits the other node. The CDP++ model could deal quite easily with the pure segmental and suprasegmental priming effects we reported in Experiments 1 and 2. The English version of the model has already successfully simulated the MOPE (Perry et al., 2010; see also Perry et al., 2007); the CDP++.Italian has simulated the stress priming effect reported by Sulpizio et al. (2012b) for visible priming experiments and for this reason we assume the model should be able to simulate the prime-target stress interference we reported in Experiment 2. However, *prima facie*, the CDP++.Italian does not seem equipped to account for the asymmetric pattern arising from the joint manipulation of stress and phonemes, mainly because the current implementation does not specify how the PONs and the SONs communicate with each other and this underspecifies how phonemes and stress information are assembled together. On this issue, one possibility is that the segment-to-frame association would work rightward incrementally and the phonological-to-phonetic mapping may start as soon as the relevant usable information becomes available. Therefore, while the activation within the PONs and the SONs may proceed in parallel and quite independently, the phonological-to-phonetic interface would require all the relevant (segmental and suprasegmental) information for the to-be-articulated unit to be available for the system.

On a related issue, as it stands now, the phonological output buffer of the CDP++ binds the start of articulation only to the activation of the correct stress pattern of the stimulus: according to the *stress naming criterion* parameter, independently of how easy and/or fast the word's phonemes are being processed, reading aloud can start only after the stress has been assigned. However, the results we obtained in Experiments 3 and 4 suggest that the timing of word articulation is affected not only by stress

activation, but also by phonemic activation and by the interaction between the two types of information.

CONCLUSION

Our findings shed new light on the stages of phonological and phonetic encoding in word reading. We have shown that readers may compute stress apart from phonemes and that the two types of information may be independently primed as we obtained both pure segmental priming and pure suprasegmental priming in our first two experiments. The data are consistent with previous findings reported in literature (e.g., Forster and Davis, 1991; Colombo and Zevin, 2009; Dimitropoulou et al., 2010; Sulpizio et al., 2012b, for the suprasegmental priming) and provide further support for the assumption that the latest stages of reading aloud include a process of segment-to-frame association that drives the stimulus phonetic encoding (see also the speech planning account: Kinoshita, 2000; Malouf and Kinoshita, 2007). Moreover, we propose that the phonological buffer of the reading system acts as the locus of the phonological-to-phonetics interface, that is the locus where the abstract phonological word is converted into its phonetic representation as soon as the relevant information for the to-be-planned unit becomes available.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01612>

REFERENCES

- Arciuli, J., Monaghan, P., and Ševa, N. (2010). Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. *J. Mem. Lang.* 63, 180–196. doi: 10.1016/j.jml.2010.03.005
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., et al. (2005). *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS) [CoLFIS. Corpus and Frequency Lexicon of Written Italian]*. Available at: http://www.istc.cnr.it/material/database/colfis/index_eng.shtml
- Burani, C., and Arduino, L. S. (2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. *Brain Lang.* 90, 318–325. doi: 10.1016/S0093-934X(03)00444-9
- Burani, C., Paizi, D., and Sulpizio, S. (2014). Stress assignment in reading Italian: friendship outweighs dominance. *Mem. Cogn.* 42, 662–675. doi: 10.3758/s13421-013-0379-5
- Caramazza, A., and Miceli, G. (1990). The structure of graphemic representations. *Cognition* 37, 243–297. doi: 10.1016/0010-0277(90)90047-N
- Carreiras, M., Ferrand, L., Grainger, J., and Perea, M. (2005). Sequential effects of phonological priming in visual word recognition. *Psychol. Sci.* 16, 585–590. doi: 10.1111/j.1467-9280.2005.01579.x
- Chetail, F., Scaltritti, M., and Content, A. (2014). Effect of the consonant–vowel structure of written words in Italian. *Q. J. Exp. Psychol.* 67, 833–842. doi: 10.1080/17470218.2014.898668
- Colombo, L. (1992). Lexical stress effect and its interaction with frequency in word pronunciation. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 987–1003. doi: 10.1037/0096-1523.18.4.987
- Colombo, L., and Zevin, J. (2009). Stress priming in reading and the selective modulation of lexical and sub-lexical pathways. *PLoS ONE* 4:e7219. doi: 10.1371/journal.pone.0007219
- 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
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* 93, 283–321. doi: 10.1037/0033-295X.93.3.283
- Dell, G. S. (1988). The retrieval of phonological forms in production: tests of predictions from a connectionist model. *J. Mem. Lang.* 27, 124–142. doi: 10.1016/0749-596X(88)90070-8
- Dimitropoulou, M., Duñabeitia, J. A., and Carreiras, M. (2010). Influence of prime lexicality, frequency, and pronounceability on the masked onset priming effect. *Q. J. Exp. Psychol.* 63, 1813–1837. doi: 10.1080/17470210903540763
- Forster, K. I., and Davis, C. (1991). The density constraint on form-priming in the naming task: interference effects from a masked prime. *J. Mem. Lang.* 30, 1–25. doi: 10.1016/0749-596X(91)90008-8
- Grainger, J., and Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: cross-task comparisons. *J. Mem. Lang.* 35, 623–647. doi: 10.1006/jmla.1996.0033

- Kinoshita, S. (2000). The left-to-right-nature of the masked onset priming effect in naming. *Psychol. Bull. Rev.* 7, 133–141. doi: 10.3758/BF03210732
- Kinoshita, S., and Norris, D. (2012). Task-dependent masked priming effects in visual word recognition. *Front. Psychol.* 3:178. doi: 10.3389/fpsyg.2012.00178
- Kinoshita, S., and Woollams, A. (2002). The masked onset priming effect in naming: computation of phonology or speech-planning. *Mem. Cogn.* 30, 237–245. doi: 10.3758/BF03195284
- Laudanna, A., Burani, C., Cermele, A., and Parisi, D. (1989). Effetti sillabici e accentuali nella lettura di parole nuove [Syllabic and stress effects in reading new words aloud]. *G. Ital. Psicol.* 16, 119–140.
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–75. doi: 10.1017/S0140525X99001776
- Malouf, T., and Kinoshita, S. (2007). Masked onset priming effect for high-frequency words: further support for the speech-planning account. *Q. J. Exp. Psychol.* 80, 1155–1167. doi: 10.1080/17470210600964035
- Mousikou, P., Coltheart, M., and Saunders, S. (2010). Computational modeling of the masked onset priming effect in reading aloud. *Eur. J. Cogn. Psychol.* 22, 725–763. doi: 10.1080/09541440903052798
- Paizi, D., Zoccolotti, P., and Burani, C. (2011). Lexical stress assignment in Italian development dyslexia. *Read. Writ. Interdiscip. J.* 24, 443–461. doi: 10.1007/s11145-010-9236-0
- Peressotti, F., Job, R., and Cubelli, R. (2003). On recognizing proper names: the orthographic cue hypothesis. *Cogn. Psychol.* 47, 87–116. doi: 10.1016/S0010-0285(03)00004-5
- Perry, C., Ziegler, J. C., and Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychol. Rev.* 114, 273–315. doi: 10.1037/0033-295X.114.2.273
- Perry, C., Ziegler, J. C., and Zorzi, M. (2010). Beyond single syllables: large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cogn. Psychol.* 61, 106–151. doi: 10.1016/j.cogpsych.2010.04.001
- Perry, C., Ziegler, J. C., and Zorzi, M. (2014). CDP++. Italian: modelling sublexical and supralexical inconsistency in a shallow orthography. *PLoS ONE* 9:e94291. doi: 10.1371/journal.pone.0094291
- Rastle, K., and Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *J. Mem. Lang.* 42, 342–364. doi: 10.1006/jmla.1999.2687
- Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Mem. Cogn.* 32, 212–222. doi: 10.3758/BF03196853
- Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: from Germanic languages to mandarin Chinese and Japanese. *Jpn. Psychol. Res.* 57, 22–37. doi: 10.1111/jpr.12050
- Roelofs, A., and Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 922–939. doi: 10.1037/0278-7393.24.4.922
- Schiller, N. O. (2004). The onset effect in word naming. *J. Mem. Lang.* 50, 477–490. doi: 10.1016/j.jml.2004.02.004
- Schiller, N. O., and Kinoshita, S. (2007). “The effect of mismatching segmental information on the masked onset priming effect (MOPE),” in *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, eds J. Trouvain and W. J. Barry (Saarbrücken: Universität der Saarlandes), 1885–1888.
- Sternberg, S., Knoll, R. L., Monsell, S., and Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica* 45, 175–197. doi: 10.1159/000261825
- Sulpizio, S., Arduino, L. S., Paizi, D., and Burani, C. (2013). Stress assignment in reading Italian polysyllabic pseudowords. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 51–68. doi: 10.1037/a0028472
- Sulpizio, S., Bureaux, M., Burani, C., Deguchi, C., and Colombo, L. (2012a). “Stress assignment in the development of reading aloud: nonword priming effects on Italian children,” in *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, eds N. Miyake, D. Peebles, and R. P. (Austin, TX: Cognitive Science Society), 2369–2374.
- Sulpizio, S., Job, R., and Burani, C. (2012b). Priming lexical stress in reading Italian aloud. *Lang. Cogn. Proc.* 27, 808–820. doi: 10.1080/01690965.2011.584820
- Sulpizio, S., and Colombo, L. (2013). Lexical stress, frequency, and stress neighbourhood effects in the early stages of Italian reading development. *Q. J. Exp. Psychol.* 66, 273–284. doi: 10.1080/17470218.2013.785577
- Sulpizio, S., and Job, R. (2013). “Syllable frequency and stress priming interact in reading Italian aloud,” in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Science Society), 1402–1407.
- Sulpizio, S., Spinelli, G., and Burani, C. (2015). Stress affects articulatory planning in reading aloud. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 453–461. doi: 10.1037/a0038714
- Thornton, A. M., Iacobini, C., and Burani, C. (1997). *BDVDB: Una Base di Dati Sul Vocabolario di Base Della Lingua Italiana. [A database for the Italian basic dictionary]*. Rome: Bulzoni.
- Van Selst, M., and Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Q. J. Exp. Psychol.* 47, 631–650. doi: 10.1080/14640749408401131

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

Copyright © 2015 Sulpizio and Job. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Distinguishing Target From Distractor in Stroop, Picture–Word, and Word–Word Interference Tasks

Xenia Schmalz^{1,2*}, Barbara Treccani³ and Claudio Mulatti¹

¹ Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università degli Studi di Padova, Padova, Italy, ² ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, NSW, Australia, ³ Dipartimento di Storia, Scienze dell'Uomo e della Formazione, Università degli Studi di Sassari, Sassari, Italy

Lexical selection—both during reading aloud and speech production—involves selecting an intended word, while ignoring irrelevant lexical activation. This process has been studied by the use of interference tasks. Examples are the Stroop task, where participants ignore the written color word and name the color of the ink, picture–word interference tasks, where participants name a picture while ignoring a super-imposed written word, or word–word interference (WWI) tasks, where two words are presented and the participants need to respond to only one, based on an pre-determined visual feature (e.g., color, position). Here, we focus on the WWI task: it is theoretically impossible for existing models to explain how the cognitive system can respond to one stimulus and block the other, when they are presented by the same modality (i.e., they are both words). We describe a solution that can explain performance on the WWI task: drawing on the literature on visual attention, we propose that the system creates an object file for each perceived object, which is continuously updated with increasingly complete information about the stimulus, such as the task-relevant visual feature. Such a model can account for performance on all three tasks.

Keywords: word–word interference, picture–word interference, Stroop test, lexical selection by competition, mental lexicon, selective attention

OPEN ACCESS

Edited by:

Simone Sulpizio,
University of Trento, Italy

Reviewed by:

Kyran Tsapkini,
Johns Hopkins Medicine, USA
F-Xavier Alario,
Centre National de la Recherche
Scientifique – Aix-Marseille Université,
France

*Correspondence:

Xenia Schmalz
xenia.schmalz@gmail.com

Specialty section:

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

Received: 22 May 2015

Accepted: 16 November 2015

Published: 15 December 2015

Citation:

Schmalz X, Treccani B and Mulatti C
(2015) Distinguishing Target From
Distractor in Stroop, Picture–Word,
and Word–Word Interference Tasks.
Front. Psychol. 6:1858.
doi: 10.3389/fpsyg.2015.01858

INTRODUCTION

The cognitive system is often confronted with a set of stimuli, where one stimulus requires a response while others need to be ignored. This phenomenon is relevant to the process of lexical selection (Levelt et al., 1999): here, a target word needs to be produced, while irrelevant information (e.g., a semantically related word, or the word's translation for multi-linguals) is ignored. This is only one step in the complex process of speech production, but it has been subject to some attention and controversy (e.g., Lupker, 1979; Finkbeiner and Caramazza, 2006; La Heij et al., 2006; Mahon et al., 2007).

Here, we consider whether existing models of lexical selection can adequately account for performance on three tasks that have been used to study the process of word selection in speech production: the Stroop task (Stroop, 1935; Klein, 1964; MacLeod, 1991), the picture–word interference (PWI) task (La Heij, 1988; Schriefers et al., 1990; Mahon et al., 2007), and the word–word interference (WWI) task (Glaser and Glaser, 1989; Waechter et al., 2011; Mulatti et al., 2015). These experimental tasks have in common the process of selecting a target, to which the participant needs to respond (e.g., by reading aloud, a lexical decision, or semantic categorization), and the need to ignore an irrelevant stimulus, the distractor. In the Stroop task the target is usually the font color and the distractor is the written color word, for the PWI task the target is a picture and the distractor a

super-imposed written word, and for the WWI task the participants are presented with two words and need to respond to one based on a pre-determined characteristic (e.g., color, position).

A model of how the cognitive system performs selection should be able to explain performance on all three of these tasks. We argue that contemporary theories fail to account for performance on the WWI task, as it is theoretically impossible for the system in these models to ignore a distractor of the same type as a target (i.e., when both are words). We describe a model that can account for performance on all three tasks by creating a token, which combines, for each visual object, its identity with task-relevant visual features. We conclude with a brief discussion of how this model may account for phenomena in the more ecologically valid tasks of speech production and text reading.

In the current paper, we address an issue that arises in interference tasks: how does the system know which potentially activated lexical node belongs to the target, and which to the distractor? This is different—and logically preceding—from asking how relevant lexical entries are activated. The problem here is understanding how a given pattern of activation in memory is linked back to the stimulus evoking it. Ultimately, the task is to respond to only one of the two stimuli simultaneously presented, and so, the system needs to know that a given response corresponds to a given stimulus to decide what to process and what to gate.

PREVIOUS SOLUTIONS

Any explanation of performance on the Stroop and PWI interference tasks relies on the concept of mental lexicons (Coltheart, 2004; but see Elman, 2004, 2009, 2011, for an alternative account of lexical knowledge). To explain the Stroop and PWI tasks, the mental lexicon needs to include three different domain-specific input modules: a color system (CS), a picture lexicon (PL), and an orthographic lexicon (OL). In addition, it needs a semantic system and a phonological output module. Each input module comprises a collection of domain-specific units, where each unit corresponds to a given element in that domain (e.g., each unit in the PL represents the structural description of an object), and is activated if that element is presented as input stimulus. Once a unit in one of the input modules is activated, it sends activation to the connected units in the semantic and phonological modules. In contrast to the units in the CS and PL, units in the OL also directly activate units in the phonological output lexicon, rather than only indirectly via the semantic system.

The existing proposals of performance on the PWI and Stroop are intrinsically linked to the notion of modality-specific input lexicons. These models achieve selective target activation based on a simple principle: in a PWI or Stroop task, the system needs to block the information from the wrong module. The system needs to monitor the activation in the input modules, because monitoring the activation in the later stages (i.e., the semantic system or phonological lexicon) would not provide the means to distinguish between information from different modes of input. Then, the system could deactivate the distractor activation, if it detects that it is sent from the distractor module which, in the case

of the Stroop and PWI tasks, is the OL. Such deactivation could be achieved by disrupting processing of a stimulus that is provided by the “wrong” module.

Proposals along these lines have been made by several authors. Cohen et al. (1990) describe a parallel-distributed-processing computational model that can simulate results from Stroop-like tasks. Task instructions (ignore the written word vs. ignore the color of the font) are implemented as two input units that, via a set of hidden units, increase the activation for their respective target mode, and inhibit the stimulus provided by the distractor mode. WEAVER++ (Levelt et al., 1999), a leading computational model of word production, has been programmed to account for results on both Stroop and PWI tasks (Roelofs, 2003). Like the model of Cohen et al. (1990) the system tracks the input source of each stimulus: When activation spreads along the connections of the model's network, it leaves activation tags at each node (Roelofs, 1993). These tags specify the source of the activation, and thus, in a PWI experiment, there are tags for both the picture stimulus and for the printed word stimulus: a response is selected only if its source tag corresponds to the picture.

These mechanisms rest on the same basic intuition, that pictures, colors, and printed words are inherently different. If the system can track the nature of a given item, it can distinguish targets from distractors. The identity of each stimulus does not influence these processes, since this would imply that the system knows the identity of the item before recognizing the item itself; instead, it only needs to classify the item in input as a member of the category of pictures (or colors, or printed words).

This family of explanations works when the two stimuli are processed through different input modules, but when the stimuli are of the same nature, it runs into fatal trouble. In a WWI task, participants are presented with two words simultaneously and are required to read one word while ignoring the other. Target and distractor can be distinguished because of their relative spatial position (La Heij et al., 1990; Mulatti et al., 2015), because of the different colors (Waechter et al., 2011), or their temporal order (Glaser and Glaser, 1989). Akin to the PWI task, evidence suggests that distractor affects target processing: unrelated low frequency distractors interfere more than unrelated high frequency distractor (Mulatti et al., 2015), target and distractor frequency exert additive effects on target processing (Mulatti et al., 2015), and semantically related distractors facilitate target processing (Waechter et al., 2011; Mulatti et al., 2015). This demonstrates that the distractors activate their orthographic and semantic representations to some extent. Therefore, accounting for the performance in the WWI task requires a mechanism that traces the source of the activation so that the system knows what has been activated by the distractor and what has been activated by the target. This mechanism, however, cannot be monitoring, tagging or biasing activation of a specific input module, because both stimuli in the WWI task are printed words, and activate nodes in the same module.

AN ALTERNATIVE APPROACH

A model that could perform input control in the Stroop, PWI and WWI tasks would need to achieve the following: (1) at an early

processing stage, it needs to assign the task-relevant visual feature to the stimulus, (2) the distractor is processed to some degree, and (3) when articulation occurs, the distractor has been suppressed (in the behavioral data, it is very rare for errors to occur, where the distractor is articulated instead of the target). Furthermore, to reflect psychologically valid mechanisms, the model should be applicable to all three tasks, as well as the extensive literature on visual attention and object recognition (Carr, 1999).

This problem has been described by Allport (1977), who stated, about Morton's (1969) logogen model, that it "lacks a specific mechanism for relating particular logogen outputs to the particular stimuli that evoked them. In particular where more than one word, or nameable item, is presented at the same time, a mechanism is clearly required to integrate appropriately the nominal identities of the items—their logogen output—with their other physical attributes—location, color, size, etc." (p. 525). Allport's (1977) proposed solution is a mechanism which binds the word's pre-categorical perceptual features with the word's identity, or orthographic features, to form an episode. Once the task-dependent visual characteristics are linked to their respective orthographic information, the system knows which of the two lexical representations correspond to the target and which corresponds to the distractor, and the appropriate decision of what to read and what to ignore can be made. This approach is different from those explicitly proposed to account for interference tasks, because it does not require the tracking the input modes of each stimulus. Importantly, the idea of binding various attributed of the stimulus could be applied to explain how participants perform the WWI task as well as the PWI and Stroop tasks. In the following section, we describe a specified model based on Allport's (1977) suggestion, and how it could account for performance on interference as well as reading tasks.

Creating Proto-Words: Binding Visual Features

Upon stimulus presentation, the first step for the model is to detect that the display consists of two objects. In the WWI task, the system perceives the words as objects due to their visual distinctiveness compared to the background, and creates abstract representations for each of these objects. This lower-level selection process has been studied in great detail by researchers of visual attention. According to object file theories (e.g., Kahneman et al., 1992; Xu and Chun, 2009; Hayworth et al., 2011), a "file" is created for each object, which can be subsequently filled with continuously updated information about the object's characteristics. At this stage, the objects have not yet been identified as words, but instead are organized bundles of the visual features of the word ("proto-words" in their terms).

Orthographic Processing

As soon as proto-words are created, orthographic processing can be initiated, as two functionally independent sets of letter detectors—one for each proto-word—are constructed. After the creation of the letter sets, lexical processing can be initiated. The lexical processing stage creates a bottleneck, as only one word can undergo lexical processing at a given point in time (Coltheart et al., 2001). When the system is faced with multiple written words, it is

assumed that the foveated word is prioritized (Engbert et al., 2005; Mulatti et al., 2015). This attentional gradient reflects the anatomy of the retina, where increasing distance from the fovea results in poorer spatial resolution. A further assumption of the model is that lexical processing is ballistic: once lexical processing of the item is initiated, it cannot be deactivated until identification has occurred.

During lexical processing, entries in the OL are activated, and this activation propagates—in an interactive and cascaded fashion—forward to the subsequent processing levels (Coltheart et al., 2001). The model posits the presence of an identification threshold in the orthographic input lexicon: as soon as this threshold is reached, the word can be treated as a tokenized instance of the type activation in the OL.

Creating a Token: The Binding Visual and Orthographic Information and the Transfer to Verbal Working Memory

A token thus serves to bind the orthographic information to the specific instance of its occurrence, including the word's non-orthographic characteristics. This process is based on the Simultaneous Type, Serial Token (TS²) model of Bowman and Wyble (2007). In the TS² model, the token does not contain the information of the corresponding type: in the case of the WWI task, the token is created once an activation threshold in the OL is reached, meaning that subsequent cascaded processing is still required to activate semantic or phonological information. Thus, the token, rather than containing all of the information that is relevant for word production and semantic processing, acts as a pointer to where this information can be found. Subsequent processing is required to bind the newly created token to the activation in the phonological and semantic lexicons, as well as to its visual, pre-categorical representation. At this stage the system can continue processing that stimulus if it occupies the position of the target, or trigger deactivation if it occupies the position of the distractor. Once the relevant information associated with the token is bound, the task-relevant information is transferred to the phonological loop of working memory (Saito and Baddeley, 2004; Bowman and Wyble, 2007). From there, articulation of the target is initiated, and the correct response can be articulated.

BEYOND THE WWI TASK: RELEVANCE OF THE MODEL TO OTHER SETTINGS

By using object files and tokens, the model described above proposes a mechanism by which the system can perform the WWI task. As we argue, it is theoretically impossible within existing proposals to account for the fact that the human participants are capable of ignoring a distractor while processing a target when these stem from the same source of input. Furthermore, the model allows for greater flexibility in incorporating visuo-attentional processes which may affect performance on interference tasks. This would provide a fruitful avenue for future research.

Future research is needed to establish how the model can account for performance on the PWI and Stroop tasks. Due to the similar nature of the three tasks, a mechanism explaining

performance on one should be applicable to the other task, with relatively minor, task-specific modifications. The principle of creating object files and tokens could theoretically also work for the Stroop and PWI tasks. However, it would be a challenge for the model to create two files for a single visual object. In the case of the Stroop task, for example, the stimulus is a word written in a specific color, and the system needs to create a separate file for two aspects of the same stimulus. Beyond experimental scenarios, it is also worth considering whether the model could be applied to more naturalistic scenarios, and specifically, how it relates to word production and sentence reading.

Interference Tasks and Word Production

The PWI interference task plays a central role in studying lexical selection in speech production (see Levelt et al., 1999, for a review). It is argued that the system, when translating a concept node to a phonological word form, needs to block competing word forms, thus posing a similar problem to the system as a PWI task. This view is not uncontroversial: it has been pointed out that in addition, the PWI task requires visuo-attentional, decision and selection processes that are not employed during speech production (Lupker, 1979; Carr, 1999; Finkbeiner and Caramazza, 2006). The degree to which the selection process involved in the PWI, WWI, and Stroop tasks—and in the model—reflects the selection process underlying lexical selection remains an open question. From a methodological perspective, a model which explains at least a proportion of the selection processes underlying the PWI task can help to isolate the task's non-linguistic components from those that are directly related to the selection of a lexical node during speech production.

Selecting Words During Reading

We argue that the WWI and the model in particular capture a cognitive mechanism that is particularly useful for text reading: namely, selecting a target word while ignoring the information

provided by the surrounding words. Generally speaking, a well-specified model which incorporates such visuo-attentional mechanisms as well as higher-level orthographic processing can provide valuable insights and testable predictions about how these processes interact.

Text reading is generally studied with the use of eye-movement tracking. In the literature on reading and eye-movements, the degree to which all words in the visual field are processed is still under debate (e.g., Kliegl et al., 2006; Schotter et al., 2012; Angele et al., 2015). Several studies report when a word is fixated, the subsequent word influences its processing, especially when the fixated word is short. As in the WWI task (Mulatti et al., 2015), high frequency of the fixated and non-fixated words are facilitatory for target processing (Kennedy and Pynte, 2005; Kliegl et al., 2006) and have an additive effect (Schroyens et al., 1999; Kliegl et al., 2006). Future research could further explore the similarities between performance in the WWI task and in the task of text reading. Given a sufficiently high overlap in the underlying cognitive processing, the WWI could serve as an experimental task to study the processes underlying text reading.

CONCLUSION

In summary, performance on the Stroop, PWI and WWI tasks reflects an important problem that is relevant to speech production and text reading. In all three tasks, information about the stimulus identity needs to be bound to the task-relevant visual information. We describe a specified model, based on a previous proposal by Allport (1977), that is capable of performing these tasks, and draws from literature of visual attention (Kahneman et al., 1992; Xu and Chun, 2009), object recognition (Bowman and Wyble, 2007; Hayworth et al., 2011), written word recognition (Coltheart et al., 2001), speech production (La Heij, 1988; La Heij et al., 1990; Levelt et al., 1999), and working memory (Saito and Baddeley, 2004).

REFERENCES

- Allport, D. A. (1977). "On knowing the meaning of words we are unable to report: the effects of visual masking," in *Attention and Performance*, Vol. 6, ed. S. Dornic (London: Academic Press), 505–533.
- Angele, B., Schotter, E. R., Slattery, T. J., Tenenbaum, T. L., Bicknell, K., and Rayner, K. (2015). Do successor effects in reading reflect lexical parafoveal processing? Evidence from corpus-based and experimental eye movement data. *J. Mem. Lang.* 79, 76–96. doi: 10.1016/j.jml.2014.11.003
- Bowman, H., and Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychol. Rev.* 114, 38. doi: 10.1037/0033-295X.114.1.38
- Carr, T. H. (1999). How does WEAVER pay attention? *Behav. Brain Sci.* 22, 39–40. doi: 10.1017/S0140525X99231779
- Cohen, J. D., Dunbar, K., and McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychol. Rev.* 97, 332. doi: 10.1037/0033-295X.97.3.332
- Coltheart, M. (2004). Are there lexicons? *Q. J. Exp. Psychol. A* 57, 1153–1171. doi: 10.1080/02724980443000007
- 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
- Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends Cogn. Sci.* 8, 301–306. doi: 10.1016/j.tics.2004.05.003
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. *Cogn. Sci.* 33, 547–582. doi: 10.1111/j.1551-6709.2009.01023.x
- Elman, J. L. (2011). Lexical knowledge without a lexicon? *Ment. Lex.* 6, 1. doi: 10.1075/ml.6.1.01elm
- Engbert, R., Nuthmann, A., Richter, E. M., and Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychol. Rev.* 112, 777. doi: 10.1037/0033-295X.112.4.777
- Finkbeiner, M., and Caramazza, A. (2006). Lexical selection is not a competitive process: a reply to La Heij et al. (2006). *Cortex* 42, 1032–1036. doi: 10.1016/S0010-9452(08)70210-7
- Glaser, W. R., and Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *J. Exp. Psychol. Gen.* 118, 13. doi: 10.1037/0096-3445.118.1.13
- Hayworth, K. J., Lescroart, M. D., and Biederman, I. (2011). Neural encoding of relative position. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1032. doi: 10.1037/a0022338
- Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* 24, 175–219. doi: 10.1016/0010-0285(92)90007-0
- Kennedy, A., and Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vis. Res.* 45, 153–168. doi: 10.1016/j.visres.2004.07.037
- Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *Am. J. Psychol.* 77, 576–588. doi: 10.2307/1420768

- Kliegl, R., Nuthmann, A., and Engbert, R. (2006). Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *J. Exp. Psychol. Gen.* 135, 12–35. doi: 10.1037/0096-3445.135.1.12
- La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Mem. Cogn.* 16, 400–410. doi: 10.3758/BF03214220
- La Heij, W., Happel, B., and Mulder, M. (1990). Components of Stroop-like interference in word reading. *Acta Psychol.* 73, 115–129. doi: 10.1016/0001-6918(90)90074-P
- La Heij, W., Kuipers, J.-R., and Starreveld, P. A. (2006). In defense of the lexical-competition account of picture–word interference: a comment on Finkbeiner and Caramazza (2006). *Cortex* 42, 1028–1031. doi: 10.1016/S0010-9452(08)70209-0
- Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/S0140525X99001776
- Lupker, S. J. (1979). The semantic nature of response competition in the picture–word interference task. *Mem. Cogn.* 7, 485–495. doi: 10.3758/BF03198265
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychol. Bull.* 109, 163. doi: 10.1037/0033-2909.109.2.163
- Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., and Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture–word interference paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 503. doi: 10.1037/0278-7393.33.3.503
- Morton, J. (1969). Interaction of information in word recognition. *Psychol. Rev.* 76, 165. doi: 10.1037/h0027366
- Mulatti, C., Ceccherini, L., and Coltheart, M. (2015). What can we learn about visual attention to multiple words from the word–word interference task? *Mem. Cogn.* 43, 121–132. doi: 10.3758/s13421-014-0450-x
- Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: retrieval of verbs. *Cognition* 47, 59–87.
- Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychol. Rev.* 110, 88. doi: 10.1037/0033-295X.110.1.88
- Saito, S., and Baddeley, A. D. (2004). Irrelevant sound disrupts speech production: exploring the relationship between short-term memory and experimentally induced slips of the tongue. *Q. J. Exp. Psychol. A* 57, 1309–1340. doi: 10.1080/02724980343000783
- Schotter, E. R., Angele, B., and Rayner, K. (2012). Parafoveal processing in reading. *Attent. Percept. Psychophys.* 74, 5–35. doi: 10.3758/s13414-011-0219-2
- Schriefers, H., Meyer, A. S., and Levelt, W. J. (1990). Exploring the time course of lexical access in language production: picture–word interference studies. *J. Mem. Lang.* 29, 86–102. doi: 10.1016/0749-596X(90)90011-N
- Schroyens, W., Vitu, F., Brysbaert, M., and d'Ydewalle, G. (1999). Eye movement control during reading: foveal load and parafoveal processing. *Q. J. Exp. Psychol. A* 52, 1021–1046. doi: 10.1080/713755859
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643. doi: 10.1037/h0054651
- Waechter, S., Besner, D., and Stolz, J. A. (2011). Basic processes in reading: spatial attention as a necessary preliminary to orthographic and semantic processing. *Vis. Cogn.* 19, 171–202. doi: 10.1080/13506285.2010.517228
- Xu, Y., and Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends Cogn. Sci.* 13, 167–174. doi: 10.1016/j.tics.2009.01.008

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

Copyright © 2015 Schmalz, Treccani and Mulatti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



What can Written-Words Tell us About Lexical Retrieval in Speech Production?

Eduardo Navarrete^{1*}, Bradford Z. Mahon^{2,3,4*}, Anna Lorenzoni¹ and Francesca Peressotti¹

¹ Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova, Padova, Italy, ² Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA, ³ Department of Neurosurgery, University of Rochester Medical Center, Rochester, NY, USA, ⁴ Center for Language Sciences, University of Rochester, Rochester, NY, USA

OPEN ACCESS

Edited by:

Sachiko Kinoshita,
Macquarie University, Australia

Reviewed by:

Jon Andoni Dunabeitia,
Basque Center on Cognition, Brain
and Language, Spain
Elin Runnqvist,
Laboratoire Parole et Langage,
France

*Correspondence:

Eduardo Navarrete
enavarrete2007@gmail.com;
Bradford Z. Mahon
mahon@rcbi.rochester.edu

Specialty section:

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

Received: 15 August 2015

Accepted: 11 December 2015

Published: 06 January 2016

Citation:

Navarrete E, Mahon BZ, Lorenzoni A
and Peressotti F (2016) What can
Written-Words Tell us About Lexical
Retrieval in Speech Production?
Front. Psychol. 6:1982.
doi: 10.3389/fpsyg.2015.01982

In recent decades, researchers have exploited semantic context effects in picture naming tasks in order to investigate the mechanisms involved in the retrieval of words from the mental lexicon. In the blocked naming paradigm, participants name target pictures that are either blocked or not blocked by semantic category. In the continuous naming task, participants name a sequence of target pictures that are drawn from multiple semantic categories. Semantic context effects in both tasks are a highly reliable phenomenon. The empirical evidence is, however, sparse and inconsistent when the target stimuli are printed-words instead of pictures. In the first part of the present study we review the empirical evidence regarding semantic context effects with written-word stimuli in the blocked and continuous naming tasks. In the second part, we empirically test whether semantic context effects are transferred from picture naming trials to word reading trials, and from word reading trials to picture naming trials. The results indicate a transfer of semantic context effects from picture naming to subsequently read within-category words. There is no transfer of semantic effects from target words that were read to subsequently named within-category pictures. These results replicate previous findings (Navarrete et al., 2010) and are contrary to predictions from a recent theoretical analysis by Belke (2013). The empirical evidence reported in the literature together with the present results, are discussed in relation to current accounts of semantic context effects in speech production.

Keywords: speech production, word reading, semantic context effect, lexical access, picture naming

INTRODUCTION

One of the most widely investigated issues by researchers interested in language production concerns the nature of the processes involved in the retrieval of words from the speaker's memory system. Researchers agree on two general architectural parameters about lexical access. The first universally agreed-upon parameter is that what determines word retrieval is the level of activation of the corresponding lexical representation, in the sense that the word that is ultimately produced was the most highly activated lexical representation at the moment that it was retrieved for production. The second assumption shared by models of speech production is that activation spreads from the semantic system to the lexical system, and a single concept cannot be activated alone, but rather spreads activity to a cohort of related concepts. The implication is that not only

the target word is activated, but also a cohort of semantically related words; thus, the target word must be retrieved against a backdrop of activated but non-target words (e.g., Dell, 1986; Caramazza, 1997; Levelt et al., 1999; Rapp and Goldrick, 2000).

Based on these assumptions, a straightforward empirical approach for exploring lexical retrieval consists in manipulating the semantic context within which speakers retrieve words in naming tasks. Many empirical studies have adopted this approach and there is compelling evidence that speaking is modulated by semantic context: the time and the accuracy of word production is affected by the semantic relationship between the targets, the to-be uttered words, and non-target and potentially task irrelevant stimuli. Broadly speaking two types of semantic context manipulations can be distinguished, one in which semantic context is manipulated at the *intra-trial level* and one in which semantic context is manipulated at the *inter-trial level*. For *intra-trial* semantic context manipulations, participants have to name a target stimulus while ignoring the presentation of a distractor element that can be semantically related or unrelated with the target. The distractor element appears within the same trial, that is, simultaneously with, or slightly before or after, the target itself. This is the type of manipulation behind Stroop-like interference paradigms, such as the picture-word interference (e.g., Rosinski, 1977; Lupker, 1979; La Heij, 1988; Schriefers et al., 1990; Damian and Bowers, 2003; Roelofs, 2003; Finkbeiner and Caramazza, 2006) and the word-word interference tasks (Glaser and Glaser, 1989; La Heij et al., 1990; Roelofs et al., 2013; Treccani and Mulatti, 2015).

One of the most widely exploited experimental approaches for investigating lexical retrieval has been done in the tradition of *intra-trial* manipulations, and in particular, the picture-word interference task. In this task, participants are required to name pictures while ignoring the presentation of a distractor word (for reviews see Mahon et al., 2007; Navarrete and Mahon, 2013; Spalek et al., 2013). One concern with that paradigm is that in order to understand how target words are retrieved, bridging assumptions are required about how distractor stimuli are processed in the system. To date, there is still no consensus on this issue, as debate continues regarding (i) how distractors are excluded from production, (ii) and whether, and if so how, distractor exclusion affects retrieval of the target words (e.g., Damian and Bowers, 2003; Dhooge and Hartsuiker, 2010; Roelofs et al., 2011; Mahon et al., 2012; Mulatti and Coltheart, 2012, 2014; Finocchiaro and Navarrete, 2013; Hantsch and Mädebach, 2013; Hutson et al., 2013; Mädebach and Hantsch, 2013; Navarrete and Mahon, 2013; Roelofs and Piai, 2013; Starreveld et al., 2013; Mahon and Navarrete, 2014). More recently, researchers have focused on *inter-trial* semantic manipulations.

In *inter-trial* manipulations, the goal is to explore how lexical retrieval (i.e., on trial n) is modulated by having retrieved semantically related or unrelated words on preceding trials (e.g., on trials $n-1$, $n-2$, $n-3$, ...). It has been shown that picture naming requires more time (and is more prone to errors) when semantic coordinate words have been retrieved some trials before. For instance, participants name pictures (e.g., shark) more slowly when some trials before a semantic coordinate word (e.g., whale) is named as a response to a written definition, compared to

when a non-semantic coordinate was previously named (e.g., volcano) (Wheeldon and Monsell, 1994). Semantic costs are also reported when, instead of naming written definitions, participants name pictures (e.g., Brown, 1981; Howard et al., 2006). In contrast, when semantic coordinate pictures are presented on two *consecutive* trials, semantic facilitation instead of a semantic cost (i.e., interference) is observed. That is, a picture on trial n is named faster when a semantic coordinate picture is named on trial $n-1$ compared to when a semantically unrelated picture was named on trial $n-1$ (e.g., Huttenlocher and Kubicek, 1983; Lupker, 1988; Biggs and Marmurek, 1990). While semantic interference is a long lasting phenomenon that persists over several trials, semantic facilitation is a short lasting phenomenon that is observed only for consecutive trials (e.g., Damian and Als, 2005; Navarrete et al., 2012, 2014).

Aside from the distinction between *intra-trial* and *inter-trial* semantic context manipulations, another critical dimension concerns the format of the target stimuli, which can be either pictures or printed words. The distinction between picture and word format is of critical relevance because while naming a target picture is a semantically mediated task, in the sense that the lexicon is accessed through visual and semantic processing, word reading can be achieved independently of semantic mediation and be resolved through direct links between orthographic and phonological representations (i.e., via sub-lexical processing; see for instance Coltheart, 2004). This does not mean that there is no semantic activation in reading and that some amount of activation will be propagated through to the semantic system (Coltheart et al., 1993, 2001; Perry et al., 2007); rather, word naming can be accomplished without semantic mediation. Perhaps the neuropsychological pattern of word-meaning blindness is the most intuitive evidence of this. Patient JO reported by Lambon Ralph et al. (1996), was able to read aloud words and non-words and was perfect on visual lexical decision tasks. Access to meaning was also normal for spoken words and from objects. Interestingly, JO was severely impaired at comprehending written words during silent reading. Further evidence that printed word stimuli can bypass semantic mediation comes from word translation tasks in bilingual speakers (Kroll and Stewart, 1994; Navarrete et al., 2015).

The advantage of *inter-trial* manipulations is that, as described above, there is no distractor element that needs to be excluded from production. In this case, the semantic context manipulation must be understood in terms of temporal extension in the activation dynamics of the lexical system, i.e., as in a form of memory. Elsewhere we have reviewed the principal phenomena observed with *inter-trial* manipulations in picture naming tasks, that is, semantic facilitation and semantic interference effects with target pictures (Navarrete et al., 2012, 2014). In the current review, we focus on experimental results with target words and compare them with those obtained with target pictures. As we will see, empirical evidence from experiments in which the target stimuli are printed words is sparse and inconsistent. In the last section of this paper, we aim to empirically resolve, at least in part, this inconsistency by replicating the results of a previous study (Experiment 3, Navarrete et al., 2010). Before, and in order to

better situate the implications of the reviewed empirical findings, in the next section we briefly review the main theoretical accounts of long lasting *inter-trial* semantic interference.

Theoretical Accounts of Long Lasting Inter-Trial Semantic Interference

Two main approaches have been proposed to account for long lasting *inter-trial* semantic interference. Oppenheim et al. (2007, 2010) have implemented an incremental learning mechanism by which semantic-to-lexical connection weights are adjusted after each naming event (see also, see Damian and Als, 2005). The production of a word as a response to a target picture strengthens the connections between the semantic and lexical representations of that word (e.g., *cat*) and, at the same time, weakens the connections between semantic and lexical representations of semantic coordinates of that word (e.g., *dog*, *horse*). When on a subsequent trial a semantic coordinate item has to be retrieved (e.g., *dog*), naming latencies will be longer because of the weakened semantic-to-lexical connections (see also Vitkovitch and Humphreys, 1991; Navarrete et al., 2010, 2012, 2014; Kleinman et al., 2015). The second approach is based on the hypothesis that lexical retrieval (i.e., selection) is a competitive process, so that the time required to retrieve a word depends on the levels of activation of other activated but non-target words (e.g., Roelofs, 1992, 1993; Levelt et al., 1999). When a word is produced as a response to a picture stimulus, it retains lexical activation for a certain period of time, making it a stronger competitor when, on subsequent trials, a semantic coordinate has to be retrieved. For instance, according to Howard et al. (2006), long lasting *inter-trial* semantic interference arises due to the convergence of three properties: priming, shared activation and competitive lexical selection. In their model, Howard et al. (2006) implement competition by lateral inhibition between lexical candidates, that is, each lexical unit (i.e., lemma in their model) inhibits other lexical units in proportion to its own activation level.

Within the competition account, some researchers argue that semantic interference should emerge in all circumstances which require the retrieval of a lexical representation (e.g., Damian et al., 2001; Vigliocco et al., 2002); other researchers argue that semantic interference emerges only when lexical selection is conceptually mediated (e.g., Howard et al., 2006; Belke, 2013). For instance, somewhat different from Oppenheim and colleagues' view, Belke (2013) implemented the incremental learning mechanism at the conceptual level in the links between semantic features and lexical semantic representations. In that framework, semantic interference originates at the conceptual level, although its locus remains at the lexical level by a mechanism of selection by competition.

Long Lasting Inter-Trial Semantic Interference with Target Words: Empirical Findings and Theoretical Implications

There exists compelling evidence that long lasting *inter-trial* semantic interference emerges when lexical retrieval is

semantically mediated, as for instance in picture naming or in definition naming tasks (Wheeldon and Monsell, 1994; for review see, Navarrete et al., 2014). By contrast, studies exploring semantic effects using printed word stimuli as targets show a more complex pattern. In their influential study, Damian et al. (2001) explored semantic interference in the blocked naming paradigm. In this task, first introduced by Kroll and Stewart (1994), participants are slower to name pictures if they were grouped into a block of all within-category items (e.g., *cat*, *dog*, *horse*) compared to blocks of items from different categories (e.g., *cat*, *table*, *lemon*). Damian et al. (2001) introduced several important changes to the original blocked naming task devised by Kroll and Stewart (1994). Of particular relevance, target words were presented instead of target pictures and participants were asked to name them either accompanied by the corresponding grammatical gender-marked determiner or in a standard reading task. Grammatical gender is a syntactic feature of nouns and cannot be predicted from conceptual properties (e.g., Navarrete and Costa, 2009), except when it correlates with conceptual properties as in the case of natural gender (e.g., Vigliocco and Franck, 1999). Therefore, even though a printed word is presented, the lexical-syntactic representation corresponding to the word must be retrieved in order to retrieve its gender. Of critical relevance, the modification introduced by Damian et al. (2001) allows for the testing of different accounts regarding semantic interference effects in speech production. The logic underlying the use of this task is that if semantic interference is ascribed to lexical processes, the effect should be present in a determiner + word naming task. In contrast, it may be argued that if semantic interference emerges because of adjustments to the mappings from semantic representations to a specific lexical representation (i.e., the target word), and such adjustments are not required when the target is a word, word stimuli should not elicit semantic interference, even in the case of a determiner + word naming task (for extended discussion of these issues, see Navarrete et al., 2010). Furthermore, it may also be argued that semantic interference emerges through lexical selection by competition only when the task at hand implies conceptual mediation; therefore, if determiner + word naming does not mandatorily require conceptual retrieval, word stimuli should not elicit semantic interference (Howard et al., 2006; Belke, 2013).

Damian et al. (2001) observed interference with picture stimuli, replicating Kroll and Stewart (1994). Critically, semantic interference was also observed with printed words in the determiner + word naming task, but not in a bare noun naming task. Indeed, when participants read the words without the determiner, a semantic facilitation effect emerged: response times were faster in the related blocks than in unrelated blocks. Damian et al. (2001) interpreted this pattern as congruent with the explanation of semantic interference in terms of competition during the selection of the syntactically specified word representation (i.e., lemma selection, in their model; for similar arguments see also Vigliocco et al., 2002; Roelofs, 2006). Because grammatical gender is a syntactically specified lexical feature, lemma selection is required to perform the task and interference would emerge as a consequence of increased

competition for lemma selection in blocks containing within-category word stimuli, compared to blocks containing word stimuli from different categories. In contrast, competition would not arise in a bare noun naming task because speakers can read target words by accessing word form representations via a route that bypasses semantic and lemma representations (Damian et al., 2001). In sum, the two printed word manipulations introduced by Damian et al. (2001) produce opposing semantic effects: semantic facilitation in bare noun production and semantic interference in determiner + word naming production. Below, we focus on these two effects.

Somewhat in contrast to Damian et al. (2001) conclusion, other studies have reported semantic interference induced by printed words even when syntactic information is not required to perform the task at hand, as in bare noun naming. For instance, participants in the study by Vitkovitch et al. (2010) first read a sequence of words and then, shortly afterwards, they named a sequence of pictures. The results showed a semantic interference effect on picture naming, such that picture naming times were longer if some trials before semantic coordinate words had been read, compared to when unrelated words had previously been read (see also Tree and Hirsh, 2003; Vitkovitch et al., 2006; Vitkovitch and Cooper-Pye, 2012). Vitkovitch et al. (2006) interpreted this word-to-picture semantic interference as congruent with competition during name retrieval, so that the relative activation of competitors (i.e., the word stimuli) slows down the selection of the picture name. Critically, as this was a bare noun-reading task, no syntactic information had to be retrieved in order to perform the task; thus, according to the framework of Damian et al. (2001), no semantic interference should be expected. In other words, there should be no transfer of semantic interference from word reading to picture naming. Furthermore, and in contrast with the facilitation effect reported by Damian et al. (2001), Janssen et al. (2011) have reported a small, but reliable, semantic interference effect in a blocked naming task with bare-noun reading.

On the other hand, the semantic interference effect observed by Damian et al.'s (2001) study in the determiner + word naming can be contrasted with the findings of Navarrete et al. (2010) and Belke (2013). Belke (2013) did not observe semantic interference in the determiner + word naming task using the same language, procedure and task as was used by Damian et al. (2001). Navarrete et al. (2010) also failed to observe semantic interference using target words in the continuous naming paradigm, a paradigm similar to the one used by Vitkovitch et al. (2010). In the continuous naming task, participants are presented with a sequence of items (pictures or words) from diverse semantic categories in a (seemingly) random order. A reliable phenomenon is the cumulative semantic cost: picture naming times increase for every successive within-category item that is named. That is, the naming latency for each item is determined by the total number of items from the same category that have been already named (for early work see Brown, 1981; for more recent work see Howard et al., 2006; Costa et al., 2009; Alario and Martin, 2010; Runnqvist et al., 2012; Belke and Stielow, 2013; Schnur, 2014; Navarrete et al., 2015). Navarrete et al. (2010) observed the cumulative semantic

cost using pictures as targets, but no cumulative cost in a determiner + word naming task. In a further experiment, participants were presented with a sequence of intermingled words and pictures and named them (all) along with the corresponding gender-marked determiner. In that experiment, a semantic interference effect was obtained for both words and pictures, but only when the preceding within-category items were pictures, and not when the preceding within-category items were words (but see Belke, 2013, and below). Navarrete et al. (2010) concluded that naming a picture entails adjustments to semantic-to-lexical connections, specifically, incremental weakening of the semantic-to-lexical connections for semantic coordinates of the target word. Such adjustments affect the time required to access lexical representations on subsequent within-category trials, irrespective of their format (i.e., picture or word). In contrast, Navarrete et al. (2010) argued that, naming a word does not entail incremental weakening adjustments to semantic-to-lexical mappings, and therefore, the time required to access lexical representations on subsequent within-category trials is unaffected by semantic context (again, irrespective of its format, i.e., picture or word).

Interim Summary

As outlined above, semantic context effects in picture and word naming experiments is a common and straightforward approach to explore lexical retrieval during speech production. While empirical evidence regarding picture naming research is relatively congruent (see for instance, Navarrete et al., 2014), this is not so within word naming research. It is evident from the previous paragraphs that there is no simple answer as to whether word stimuli are able to elicit long lasting *inter-trial* semantic interference in language production. Certainly, differences between blocked naming and continuous naming designs may be relevant for explaining divergent findings. However, the lack of replication within the same paradigm remains problematic. For instance, in contrast with what was observed by Damian et al. (2001) in the semantic blocked naming paradigm, Belke (2013) did not report semantic interference in the determiner + word naming task. In addition, Janssen et al. (2011) using the same paradigm, reported semantic interference in bare noun production. Further experimental evidence is therefore needed in order to pinpoint which are the relevant factors in determining whether semantic interference with word targets is observed. Here we seek to provide some of this evidence by focusing on the contrasting experimental results within the continuous naming task.

Recently, Belke (2013) failed to replicate the transfer of semantic interference from pictures to words that we reported in a previous determiner + word/picture naming study (Navarrete et al., 2010; see above). In our original experimental design, for each semantic category, four items were presented within the same format (e.g., picture format) and one in a different format (in this case, word format). The deviant condition referred to those items presented in a different format than the other four within-category items, while the non-deviant condition referred to those items presented in the same format as the other four within-category items. The results of our Experiment

3 indicated semantic interference for both pictures and words when the preceding within-category items were pictures, and no semantic interference effect when the preceding within-category items were words. Belke (2013) argued that those results may have been due to uncontrolled switching costs. As picture and word stimuli were presented randomly intermingled within the naming sequence, there were, within the sequence, switch and non-switch trials, as a function of whether the previous trial contained a same-format stimulus or a different-format stimulus, respectively. Belke did not explain how the switch cost might account for our results; nevertheless it might be surmised that semantic interference could be ‘confused’ with switch costs. In order to control for such possible confusion, here we first reanalyze our data (Navarrete et al., 2010; Experiment 3), distinguishing switch from non-switch trials.

Half of the trials were switch trials and the other half were non-switch trials. At the same time, half of the pictures were presented on switch trials and half on non-switch trials. The same was the case for word stimuli. Mean latencies, split by Switch Type, are reported in **Table 1**. As can be seen, transfer effects indeed were modulated according to whether or not the trial was a switch trial—but importantly, the critical finding is present for non-switch trials and absent on switch trials. Deviant determiner + word naming trials (i.e., word naming that followed within-category picture naming trials) were slower than non-deviant determiner + word naming trials (i.e., word naming that followed within-category word naming trials), but only when the trials were non-switch trials [22 ms, $t(19) = 3.98$, $p < 0.002$]. By contrast, such an effect was absent for switch trials (3 ms, $t < 1$), indicating that switching the format of target presentation canceled out (or decreased) the transfer of semantic interference from pictures to words. Switching was also a relevant factor in the determiner + picture naming condition. In our original study, the lack of semantic interference transfer from within-category words to pictures was defined as the difference between deviant and non-deviant determiner + picture naming trials. Specifically, if there is no transfer from determiner + word naming to determiner + picture naming trials, deviant determiner + picture

naming trials (those that followed within-category word trials) should be named faster than non-deviant determiner + picture naming trials (those that followed within-category picture trials). This prediction was confirmed in the Navarrete et al. (2010) study. The re-analysis performed here shows that switching was again a critical variable, such that the difference between deviant and non-deviant determiner + picture naming trials was reliable for non-switch trials only [54 ms, $t(19) = -4.93$, $p < 0.001$]. No differences between deviant and non-deviant determiner + picture naming trials were observed for switch trials (2 ms, $t < 1$).

In sum, the re-analysis suggests that switching the format of the target may indeed affect the transfer/absence of semantic interference in the continuous naming task. However, contrary to the hypothesis of a confound between switch costs and semantic interference, semantic interference was obtained only for non-switch trials, leaving intact the conclusion reached in our previous study, and undermining the concerns raised by Belke (2013). However, and nonetheless, further empirical work is called for in order to understand how switching between formats could modulate semantic interference in the continuous naming tasks. Altogether, our re-analysis contrasts with the results reported by Belke (2013), who did not find transfer of semantic interference effect in any direction. In the present study, we aim to replicate the interaction between the transfer of semantic interference and the target format (picture or word) in the continuous naming task, while attending to any effects of switch costs.

Testing the transfer of semantic interference has relevant implications for models of lexical retrieval. Under the assumption that semantic interference effects originate from competition at the lexical level (e.g., Damian et al., 2001; Vigliocco et al., 2002; Vitkovitch et al., 2010; Vitkovitch and Cooper-Pye, 2012), there should be transfer of semantic interference in both directions, that is, independently of stimuli format (i.e., pictures or words). In contrast, according to the competitive account of Belke (2013), no transfer of semantic interference effects are expected because determiner + word naming trials do not involve conceptual processing. Finally, according to Navarrete et al. (2010), there should be transfer of semantic interference from picture to word naming trials but not from word to picture naming trials.¹

THE PRESENT STUDY

The goal of this experiment was to replicate the interaction between stimulus format and the transfer of semantic cost in the continuous naming task, originally reported by Navarrete et al. (2010, Experiment 3). Four conditions were included. In two conditions four within-category items were presented in the same format: in a picture format in the PPP-P condition,

¹Howard et al. (2006) do not make specific predictions regarding transfer effects between word and picture targets. It is however worth mentioning that Howard et al. (2006) account would predict a transfer of semantic interference from picture to word naming. This would be so because if prior picture naming events modify the accessibility of a given lexical unit, this should generate interference when the subsequent within-category trial requires lexical access, as it is the case in a determiner + word naming task.

TABLE 1 | Mean naming averages (RT in ms), standard errors (SE, in ms) for Non-deviant and Deviant trials in the determiner + picture naming and determiner + word naming trials, broken down by trial type (i.e., Non-switch and switch), of the Experiment 3 of Navarrete et al. (2010).

Trial type	Non-deviant		Deviant		
	RT	SE	RT	SE	
Determiner + picture naming					Transfer Effect
Non-switch trials	747	18	693	18	54
Switch trials	714	19	712	17	2
Switch effect	33		−19		
Determiner + word naming					
Non-switch trials	513	13	535	14	−22
Switch trials	519	13	516	16	3
Switch effect	−6		19		

The Transfer Effect is calculated as the difference between Non-deviant minus Deviant per each of the 4 sub-conditions (see Navarrete et al., 2010, per details).

and in a word format in the WWW-W condition². In the other two conditions, the item located in the fourth within-category position was presented with a different format than the previous three within-category items: In the PPP-W condition the first three within-category items were presented in a picture format and the fourth item in a word format, and in the WWW-P condition the first three within-category items were presented in a word format and the fourth item in a picture format. In order to avoid potential effects of switching, care was taken that all critical trials were in the same format as the preceding trial of the sequence. In other words, all critical trials were non-switch trials, such that issues having to do with switch costs simply do not arise.

Method

Participants

Forty native Italian speakers (undergraduate students at the University of Padova) gave written informed consent to take part in the experiment. Ethical approval was granted by the Ethical Committee for the Psychological Research of the University of Padova (protocol number: 1361, title: *Mechanisms of Word Retrieval in Spoken Language Production*).

Materials

One hundred and forty-eight items were selected. Items were presented in black upper case letters (Times Roman, Regular, 24 point) and as color photographs taken from the Internet and sized to fit within a square of 400 × 400 pixels, for the word and picture format respectively. Eighty of the items belonged to 20 different semantic categories, with four items from each category. The remaining items were fillers that did not belong to the same categories as the experimental items.

Design

The 148 items were randomly inserted into a sequence with the following constraints. Items from each category were separated by lags of 2, 4, or 6 intervening items. Each lag was used the same number of times in the sequence (i.e., 20). The first five items of the sequence were filler items. Filler items and the order of the categories in the sequence were randomly assigned with the following constraints. Five categories were assigned to each of the four experimental conditions (i.e., PPP-P, PPP-W, WWW-W, WWW-P). Half of the filler items were presented in picture format and the other half in word format. Thus, the sequence contained 74 trials in picture format and 74 trials in word format. There were a total of forty-eight switch trials (i.e., formats in trial n and $n-1$ were different) and ninety-nine non-switch trials (i.e., same format on trials n and $n-1$). Switch trials always entailed filler items: the preceding trial of a critical trial was always presented in the same format as the critical trial.

Once the first sequence was created, the same structure was used to generate nineteen new sequences. In generating these first 20 experimental sequences, it was ensured that each specific category occupied a different position across the 20 sequences. The critical items within each category were presented equally at each of the four ordinal positions (i.e., across this group of 20 sequences each critical item was presented a total of five times in each within-category ordinal position, 1 to 4). Finally, a new group of 20 sequences was created from this first group of 20 sequences by changing the format of presentations of all the items (so that each sequence had a paired sequence with the same item presentation but varying only the format of presentation of the items). There were a total of 40 experimental sequences. Each participant was presented with two sequences, one sequence of the first group and one sequence of the second group. Care was taken that the same participant was not presented with paired sequences. Each of the 40 experimental sequences was used twice across all the participants and was used the same number of times as the first and second sequence (i.e., 2).

Procedure

Participants were seated approximately 60 cm from the screen. Participants were required to name the items (pictures or words) preceded by the corresponding definitive determiner. There was no familiarization phase. The experimental session consisted of a total of two sequences of 148 trials; there was a short pause between sequences. Participants were not corrected by the experimenter throughout the experimental session. An experimental trial consisted of the following events. A fixation cross was shown in the center of the screen. In order to prevent participants from falling into a rhythm about when they were producing responses, the duration of the fixation cross was (randomly) varied among four durations: 300, 400, 500, and 600 ms. The fixation cross was followed by a blank screen for 500 ms. Following the blank screen the target picture or word was presented for 2000 ms or until the participant's response. Response latencies were measured from the onset of the picture. The next trial began 1500 ms after the onset of participants' response. Stimulus presentation, response times and response recording were controlled by the program DMDX (Forster and Forster, 2003). Naming latencies and accuracy were determined using the CheckVocal software (Protopapas, 2007).

Analysis

Analyses were performed on critical items only and separately for pictures and words. Production of clearly erroneous picture names and verbal dysfluencies were excluded from the analysis of response times and considered as errors. A total of 3.7% of the data points were excluded following these criteria. In addition, voice-key failures (0.1%) were removed from the analysis. Mean naming latencies and error rates by condition are reported in **Table 2**. Naming latencies were log-transformed using the natural logarithm to reduce skewness and approximate a normal distribution. We performed two analyses on the naming latencies. We first tested for the presence of a cumulative semantic cost in determiner + picture naming and determiner + word naming separately. This analysis was performed on non-deviant trials

²P stands for 'picture format' and W for 'word format'. The names of the conditions (e.g., PPP-P) do not refer to the presentation format of four consecutive trials within the sequence, but the presentation format of the four within-category items (e.g., in PPP-P all within-category items were pictures).

TABLE 2 | Mean naming latencies (RT in ms), standard deviations (SD in ms) and percentage of error rates (E) by Ordinal Position Within-Category and Condition.

Position	Conditions											
	PPP-P			PPP-W			WWW-W			WWW-P		
	RT	SD	E	RT	SD	E	RT	SD	E	RT	SD	E
(1)	724	183	5.8	707	160	5.8	521	102	0.3	517	99	0.5
(2)	751	211	5.6	754	217	5.6	517	96	0.8	522	100	1.3
(3)	762	221	9	769	207	9.6	517	98	0.8	517	112	1
(4)	787	224	8.7	523	94	0.3	510	99	0.8	747	193	8.1
Mean	755		7.3	685		5.3	516		0.6	516		2.7

only. That is, for determiner + picture naming trials, the first, second and third ordinal positions for the conditions PPP-P and PPP-W, and the fourth ordinal position of the condition PPP-P were included in the analysis. For determiner + word naming trials, the first, second and third ordinal position of the conditions WWW-W and WWW-P and the fourth ordinal position of the condition WWW-W were included in the analysis. We expected to replicate the cumulative semantic cost with target pictures (e.g., Howard et al., 2006) but not with word targets (Navarrete et al., 2010; Belke, 2013).

In a second analysis, we explored the transfer of semantic interference from determiner + picture naming to determiner + word naming trials and from determiner + word naming to determiner + picture naming trials. To this end, for pictures, we compared naming latencies in the fourth ordinal position in the PPP-P condition (i.e., non-deviant trials) to naming latencies in the fourth ordinal position in the WWW-P condition (i.e., deviant trials). The same was done for words, that is, we compared naming latencies in the fourth ordinal position in the WWW-W condition (i.e., non-deviant trials) to naming latencies in the fourth ordinal position in the PPP-W condition (i.e., deviant trials). We expected to replicate our previous study. That is, for picture trials, faster naming latencies for deviant picture trials than for non-deviant picture trials; for word trials, slower naming latencies for deviant word trials than for non-deviant word trials.

Analyses were performed employing linear mixed-effects models (LMM) with crossed random effects for participant and items. We used the lme4 package (Bates et al., 2011) with the R program (R Development Core Team, 2011). For the first analysis, the following LMM models were tested and compared separately for picture naming trials and word naming trials. The null models (see 0_Cumulative models) contained intercepts only and no predictors. First we added the predictor Condition (see 1_Cumulative models). We then included the predictor Lag (see 2_Cumulative models). Finally, we added the critical predictor Order within-category in order to explore the cumulative semantic cost (see 3_Cumulative models), separately for pictures and word naming trials. For the second type of analysis, that is, for trials from the fourth within-category ordinal positions, the same logic was applied. The null model

(see 0_Transfer models) contained intercepts only and no predictors. First the predictor Lag was added (see 1_Transfer models) and afterward, the critical Condition (see 2_Transfer models) was included in order to explore differences between deviant and non-deviant naming trials, again separately for picture and word trials. In each of these models the same random effects were set: participants and items. The comparison between models was performed on the likelihood ratio test and took into consideration the Bayesian information criterion (BIC; Schwarz, 1978). We calculated the BIC difference (ΔBIC) between the null model and the other models. We then used the Bayes factor (BF) approximation formula [$\exp(\Delta\text{BIC}/2)$; Raftery, 1995] to compare the relative evidence of the different models. In general, the higher the BF and the Δbic , the more evidence there is for the model compared to the null model.

Results

Picture Naming Trials

Analysis were performed on 2981 data points. As shown in **Table 3**, model P1_Cumulative does not improve the fit in relation to the null model. In contrast, the inclusion of the predictor Lag (P2_Cumulative) improves the model fit. We therefore kept Lag as a critical predictor and explored the influence of Order within-category. The model with the predictor Order within-category (P3_Cumulative) shows a better fit in relation to the model with Lag as the unique predictor. This last result replicated the cumulative semantic cost in determiner + picture naming (Navarrete et al., 2010; Belke, 2013), naming latencies increase with each additional within category item that is named (see **Table 2**). In relation to transfer effects, as can be seen in **Table 2**, naming latencies were faster for deviant trials (747 ms, fourth ordinal position in the WWW-P condition) than in non-deviant trials (787, fourth ordinal position in the PPP-P condition). Models testing this effect are reported at the bottom of the **Table 3**. The model with the critical predictor condition, P2_Transfer, shows a better fit in relation to the null model and the model with the predictor Lag (P1_Transfer).

Word Naming Trials

Analysis were performed on 3177 data points. As shown in **Table 4**, the null model (W0_Cumulative) with only participants and items as random predictors shows a better fit in relation to the models that contained Condition, Lag and Order within-category as predictors (W1_Cumulative, W2_Cumulative and W3_Cumulative models). In other words, this result suggests that there is no cumulative semantic cost in determiner + word naming task, replicating previous findings (Navarrete et al., 2010; Belke, 2013). In relation to transfer effects, as can be seen in **Table 2**, naming latencies were slower for deviant trials (523 ms, fourth ordinal position in the PPP-W condition) than for non-deviant trials (510 ms, fourth ordinal position in the WWW-W condition). The model with the critical predictor of condition, W2_Transfer, shows a better fit in relation to the null model, indicating that there was a

TABLE 3 | The fit indices on the analysis with picture trials.

Fixed effects		Model Df	Chisq (df)	p	BIC	Δbic	Approx. BF
P0_Cumulative		4			−1184		
P1_Cumulative	Condition	5	4 (1)	=0.03	−1181	−3	0.2
P2_Cumulative	Lag	5	47 (1)	<0.001	−1224	40	>10000
P3_Cumulative	Lag + order within-category	6	71(2)	<0.001	−1241	57	>10000
P0_Transfer		4			−175		
P1_Transfer	Lag	5	0.5(1)	=0.49	−169	−6	0.04
P2_Transfer	Condition	5	7	<0.01	−176	1	1.7

Df, degree of freedom; Chisq (df), chi-squared and degree of freedom; p, probability value; BIC, Bayesian Information Criterion; Δbic, differences between the Null Models (i.e., P0_) and other models; Approximately BF, Bayes Factor's (BF) approximation, $\exp(\Delta bic/2)$.

transfer of semantic interference from determiner + picture naming to determiner + word naming trials (see bottom at **Table 4**).

Discussion

This pattern of results replicates the main findings of Experiment 3 in Navarrete et al. (2010). In our previous Experiment, we showed that semantic interference is transferred from determiner + picture naming to determiner + word naming, but not from determiner + word naming to determiner + picture naming. We explained this pattern by suggesting that determiner + word naming does not involve semantically driven lexical access, but does require lexical access. This would explain why determiner + word naming trials are affected when previous within-category items were determiner + picture naming trials. At the same time, determiner + word naming does not involve incremental weakening of the semantic-to-lexical connections for semantic coordinate items, explaining the lack of transfer of the semantic interference in determiner + picture naming trials when the preceding within-category trials are determiner + word naming trials. This pattern was replicated in the present experiment. As can be seen in **Figure 1**, the conditions showing a transfer of interference were the non-deviant picture condition (determiner + picture following determiner + picture naming trials) and the deviant word condition (determiner + word following determiner + picture naming trials). No transfer of interference was reported in the deviant picture condition (determiner + picture following determiner + word naming trials) or the non-deviant word

condition (determiner + word following determiner + word naming trials; for a similar pattern see Figure 4 in Navarrete et al., 2010).

GENERAL DISCUSSION

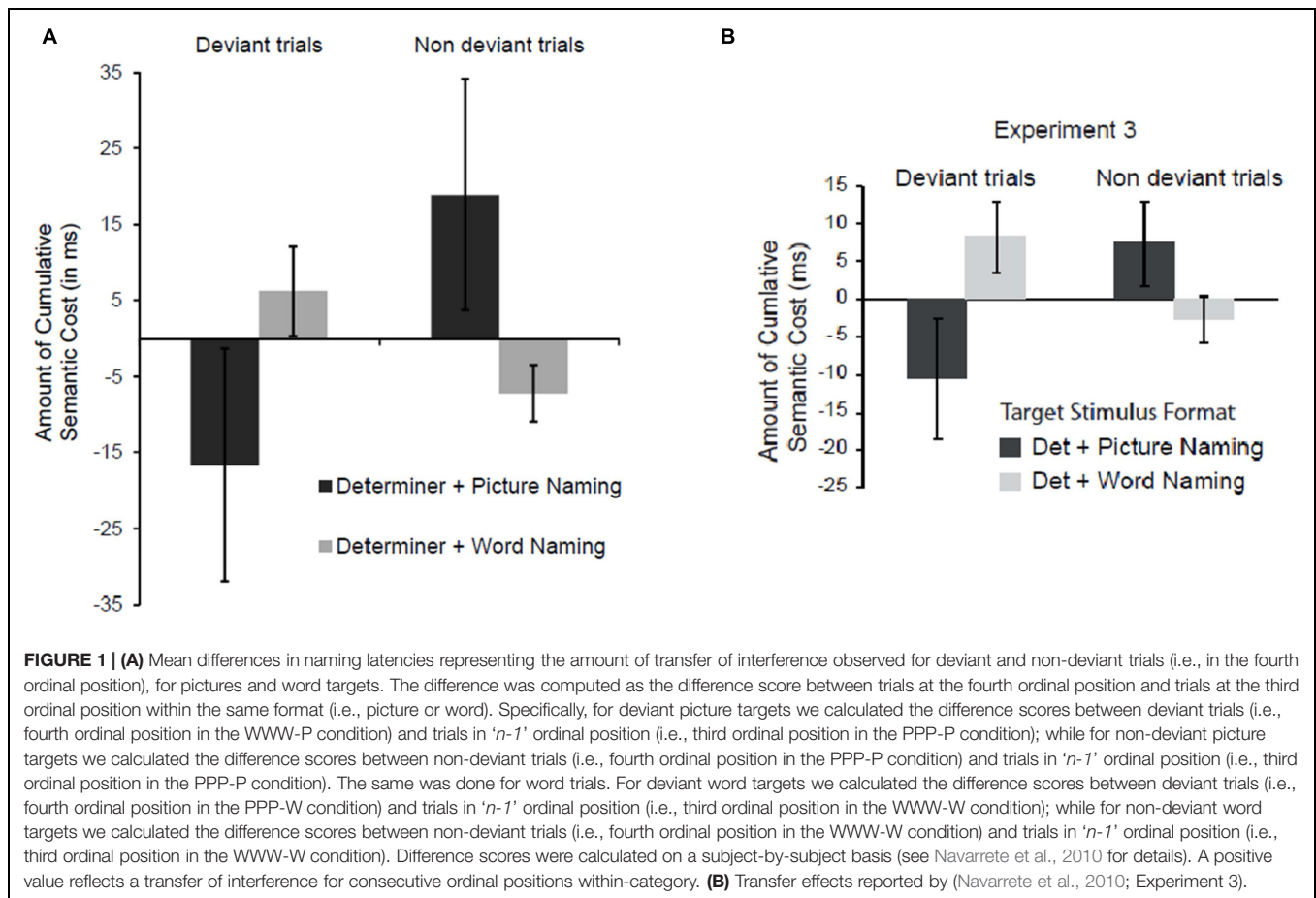
An important open issue in the field of lexical access concerns the origin(s) of semantic context effects in language production tasks. As reviewed in the Introduction, studies exploring semantic effects have revealed a consistent pattern with picture stimuli, but the evidence for words is sparse and inconsistent. In the experiment reported here we have explored whether pictures and words elicit similar long lasting *inter-trial* semantic interference effects in language production. Participants named pictures and words along with their gender marked definite determiner. The presentation format (picture or word) of the target stimuli varied within the semantic categories. For some categories, all the within-category items were presented in picture format, while for other categories, all the within-category items were presented in word format. A cumulative semantic cost with picture stimuli was observed, such that naming times increased with each additional within-category item that was named, replicating previous findings (e.g., Brown, 1981; Howard et al., 2006). In contrast, no cumulative semantic cost was reported with word stimuli, that is, when all of the within-category items were word targets, replicating previous findings (Navarrete et al., 2010; Belke, 2013).

Furthermore, two other conditions were tested. For some categories, the last within-category item was presented in a different format (picture or word) from the other previously

TABLE 4 | The fit indices on the analysis for word trials.

Fixed effects		Model Df	Chisq (df)	p	BIC	Δbic	Approx. BF
W0_Cumulative		4			−3419		
W1_Cumulative	Condition	5	0.54 (1)	=0.46	−3412	−7	0.03
W2_Cumulative	Lag	5	17(1)	= 68	−3411	−8	0.01
W3_Cumulative	Order within-category	5	5.6 (1)	<0.02	−3416	−3	0.22
W0_Transfer		4			−910		
W2_Transfer	Lag	5	0.2(1)	=0.61	−904	−6	0.04
W3_Transfer	Condition	5	10 (1)	<002	−914	4	7

For the fit statistics, see the note to **Table 3**.



presented within-category items (picture or word). These two conditions allow us to explore the transfer of semantic interference from picture to word naming trials and vice-versa, from word to picture naming trials. As stated above, whether or not there is semantic interference transfer between words and pictures is a critical issue for current models of lexical retrieval. The results demonstrated a transfer from picture to word naming trials, but not in the other direction, from word to picture naming trials. That is, while determiner + picture naming trials induce a cumulative semantic cost for subsequent within-category named determiner + word naming trials, determiner + word naming trials do not induce a semantic cost for subsequent within-category determiner + picture naming trials, replicating previous findings (Navarrete et al., 2010; see Experiment 3). This pattern suggests that lexical retrieval can be accomplished differently, depending on the format of the target stimuli. When the target is a picture, lexical retrieval is a semantically mediated process and semantic-to-lexical connections will be adjusted: connections from semantics to words are strengthened for the target word and weakened for semantic coordinates of the target word (Oppenheim et al., 2010). In contrast, when the target is a word, connections between semantics to words will (probably) be strengthened for the target word but not weakened for semantic coordinates of the target word. As a consequence, there is no semantic interference (i.e., cumulative semantic cost) for the

subsequent within-category item, regardless of its format (picture or word). In sum, the transfer of semantic interference from picture to word trials suggests that determiner + word naming involves lexical access, while the lack of transfer of semantic interference from word to picture (and word) trials suggest that determiner + word naming does not weaken the semantic-to-lexical connections of semantic coordinates of the target word (Navarrete et al., 2010)³.

Interference induced by having previously retrieved semantically related information is a broader phenomenon.

³Although grammatical gender in Italian nouns is basically a lexical property and is not related with noun meaning and phonological forms, there are some regularities between the phonological noun endings and grammatical gender. For example, Italian nouns ending in -o are predominantly masculine while nouns ending in -a are predominantly feminine. Based on this, it could be argued that on some word trials, participants might be accessing the appropriate determiner based on word ending information only. If this were the case, this could explain the lack of transfer of semantic interference with word targets, because accessing determiners from word ending information bypasses lexical retrieval. Critically, this would be independent of the experimental condition. That is, it would be independent of whether a target word is located at the fourth within-category ordinal position in condition WWW-W or at the fourth within-category ordinal position in condition PPP-W. Therefore, no semantic interference effects would be expected in either condition. That is not what the results showed. That is, the transfer of semantic interference from pictures to words in the PPP-W condition ensures that participants were accessing the corresponding lexical representation in the word target trials.

For instance, retrieval-induced forgetting is a phenomenon in which the recall of a previously studied word is hampered if, between the learning and the recall phase, participants are required to ‘actively’ retrieve other exemplars from the same semantic category (e.g., Anderson, 2003). That is, retrieval of unpracticed items from practiced categories is worse than retrieval of unpracticed items from unpracticed categories. Critically, a factor that determines retrieval-induced forgetting is the format in which the practice exemplars are presented. The phenomenon appears, for instance, when participants are presented with ‘part of the word’, as in the case of category-stem cues such as ‘FRUIT-or___’, and have to ‘actively’ retrieve the word (orange). However, when participants do not ‘actively’ retrieve the word but simply read it, by means of, for instance, category-stem cues such as ‘FRUIT-orange’, no retrieval-induced forgetting is observed (Anderson et al., 2000). The origin of long lasting *inter-trial* semantic interference is generally considered within the somewhat narrow scope of language production processes. However, the parallels with other phenomena, such as retrieval-induced forgetting, suggest that it may be promising to take a broader view. In this line, the results reported here would be congruent with the notion that only when lexical retrieval entails an ‘active’ process, as in the case of picture naming trials, there are incremental learning adjustments to the semantic-to-lexical connections and long lasting semantic interference (i.e., cumulative semantic cost) is propagated to subsequent within-category items. Such adjustments are not present when lexical retrieval does not require ‘active’ lexical retrieval, as in the case of word naming trials (Navarrete et al., 2010; see also Oppenheim et al., 2010).

In comparison to language comprehension research, it is more difficult to control the input stimulus in speech production experiments. While it is relatively easy to control relevant variables of the input words in comprehension tasks, it is harder to elicit the expected responses in production tasks. This could be a reason why written words have been extensively used as target stimuli in speech production research. Another reason could be that the most influential model of lexical retrieval in

speech production, the model developed by Levelt et al. (1999), has localized semantic interference effects at the lexical level. As written words have a direct link to the lexical system (i.e., lemma and/or lexeme), they modulate the processes occurring at the lexical level of processing (for discussion see Roelofs et al., 1996). This theoretical approach has resulted in an increased use of written words as experimental stimuli, in both the *inter-trial* and *intra-trial* semantic context manipulations. However, our findings indicate that the format of the target stimuli is a critical factor that has to be taken into consideration.

CONCLUSION

Our results suggest that long lasting *inter-trial* semantic interference is caused by adjustments to the semantic-to-lexical connections that occur in picture naming, but which do not occur in word naming. Consistent with prior arguments (Vitkovitch and Humphreys, 1991; Navarrete et al., 2010; Oppenheim et al., 2010; see also Kleinman et al., 2015) we suggest that long lasting *inter-trial* semantic interference in language production arises as a consequence of incremental adjustments to semantic-to-lexical connections, and that such adjustments do not obligatorily occur for word reading.

ACKNOWLEDGMENTS

This research was supported in part, by NSF Grant 1349042 to BM, EN, and FP; preparation of this ms was supported, in part, by NIH grant R01NS089609 to BM.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01982>

REFERENCES

- Alario, F.-X., and Martin, F. M. (2010). On the origin of the “cumulative semantic inhibition” effect. *Mem. Cognit.* 38, 57–66. doi: 10.3758/MC.38.1.57
- Anderson, M. C. (2003). Rethinking interference theory: executive control and the mechanisms of forgetting. *J. Mem. Lang.* 49, 415–445. doi: 10.1016/j.neuroscience.2008.08.026
- Anderson, M. C., Bjork, E. L., and Bjork, R. A. (2000). Retrieval-induced forgetting: evidence for a recall-specific mechanism. *Psychon. Bull. Rev.* 7, 522–530. doi: 10.3758/BF03214366
- Bates, D., Maechler, M., and Bolker, B. (2011). *lme4: Linear Mixed-Effects Models Using Eigen and S4*. R Package Version 0.999375-39. Available at: <http://CRAN.R-project.org/package=lme4>
- Belke, E. (2013). Long-lasting inhibitory semantic context effects on object naming are necessarily mediated: implications for models of lexical-semantic encoding. *J. Mem. Lang.* 69, 228–256. doi: 10.1016/j.jml.2013.05.008
- Belke, E., and Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming: evidence from blocked and continuous manipulations of semantic context. *Q. J. Exp. Psychol.* 66, 2135–2160. doi: 10.1080/17470218.2013.775318
- Biggs, T., and Marmurek, H. (1990). Picture and word naming: is facilitation due to processing overlap? *Am. J. Psychol.* 103, 81–100. doi: 10.2307/1423260
- Brown, A. S. (1981). Inhibition in cued retrieval. *J. Exp. Psychol. Hum. Learn. Mem.* 7, 204–215.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cogn. Neuropsychol.* 14, 177–208. doi: 10.1080/026432997381664
- Coltheart, M. (2004). Are there lexicons? *Q. J. Exp. Psychol. Sec. A* 57, 1153–1171. doi: 10.1080/02724980443000007
- Coltheart, M., Curtis, B., Atkins, P., and Haller, M. (1993). Models of reading aloud: dual-route and parallel-distributed-processing approaches. *Psychol. Rev.* 100, 589–608. doi: 10.1037/0033-295X.100.4.589
- 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
- Costa, A., Strijkers, K., Martin, C., and Thierry, G. (2009). The time-course of word retrieval revealed by event-related brain potentials during overt speech. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21442–21446. doi: 10.1073/pnas.0908921106
- Damian, M. F., and Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1372–1384.

- Damian, M. F., and Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychon. Bull. Rev.* 10, 111–117. doi: 10.3758/BF03196474
- Damian, M. F., Vigliocco, G., and Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition* 81, B77–B86. doi: 10.1016/S0010-0277(01)00135-4
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* 3, 283–321. doi: 10.1037/0033-295X.93.3.283
- Dhooze, E., and Hartsuiker, R. J. (2010). The distractor frequency effect in picture-word interference: evidence for response exclusion. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 878–891. doi: 10.1037/a0019128
- Finkbeiner, M., and Caramazza, A. (2006). Now you see it, now you don't: on turning semantic interference into facilitation in a Stroop-like task. *Cortex* 42, 790–796. doi: 10.1016/S0010-9452(08)70419-2
- Finocchiaro, C., and Navarrete, E. (2013). About the locus of the distractor frequency effect: evidence from the production of clitic pronouns. *J. Cogn. Psychol.* 25, 861–872. doi: 10.1080/20445911.2013.832254
- Forster, K. I., and Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35, 116–124. doi: 10.3758/BF03195503
- Glaser, W. R., and Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *J. Exp. Psychol. Gen.* 118, 13–42. doi: 10.1037/0096-3445.118.1.13
- Hantsch, A., and Mädebach, A. (2013). What does the articulatory output buffer know about alternative picture names? Evidence against the response-exclusion hypothesis. *Lang. Cogn. Process.* 28, 684–700. doi: 10.1080/01690965.2011.595725
- Howard, D., Nickels, L., Coltheart, M., and Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition* 100, 464–482. doi: 10.1016/j.cognition.2005.02.006
- Hutson, J., Damian, M. F., and Spalek, K. (2013). Distractor frequency effects in picture-word interference tasks with vocal and manual responses. *Lang. Cogn. Process.* 28, 615–632. doi: 10.1080/01690965.2011.605599
- Huttenlocher, J., and Kubicek, L. F. (1983). The source of relatedness effects on naming latency. *J. Exp. Psychol. Learn. Mem. Cogn.* 9, 486–496.
- Janssen, N., Carreiras, M., and Barber, H. A. (2011). Electrophysiological effects of semantic context in picture and word naming. *Neuroimage* 57, 1243–1250. doi: 10.1016/j.neuroimage.2011.05.015
- Kleinman, D., Runqvist, E., and Ferreira, V. S. (2015). Single-word predictions of upcoming language during comprehension: evidence from the cumulative semantic interference task. *Cogn. Psychol.* 79, 68–101. doi: 10.1016/j.cogpsych.2015.04.001
- Kroll, J. F., and Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *J. Mem. Lang.* 33, 149–174. doi: 10.1006/jmla.1994.1008
- La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Mem. Cogn.* 16, 400–410. doi: 10.3758/BF03214220
- La Heij, W., Happel, B., and Mulder, M. (1990). Components of Stroop-like interference in word reading. *Acta Psychol.* 73, 115–129. doi: 10.1016/0001-6918(90)90074-P
- Lambon Ralph, M. A., Sage, K., and Ellis, A. W. (1996). Word meaning blindness: a new form of acquired dyslexia. *Cogn. Neuropsychol.* 13, 617–639. doi: 10.1080/026432996381863
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–75. doi: 10.1017/S0140525X99001776
- Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Mem. Cogn.* 7, 485–495. doi: 10.3758/CABN.9.3.260
- Lupker, S. J. (1988). Picture naming: an investigation of the nature of categorical priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 444–455.
- Mädebach, A., and Hantsch, A. (2013). Explaining semantic facilitation and interference effects in the picture-word interference task-A rejoinder to Navarrete and Mahon (2013). *Lang. Cogn. Process.* 28, 717–722. doi: 10.1080/01690965.2013.770891
- Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., and Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 503–535.
- Mahon, B. Z., Garcea, F. E., and Navarrete, E. (2012). Picture-word interference and the response exclusion hypothesis: a response to Mulatti and Coltheart. *Cortex* 48, 373–377. doi: 10.1016/j.cortex.2011.10.008
- Mahon, B. Z., and Navarrete, E. (2014). The CRITICAL DIFFERENCE in models of speech production: a response to Roelofs and Piai. *Cortex* 52, 123–127. doi: 10.1016/j.cortex.2013.12.001
- Mulatti, C., and Coltheart, M. (2012). Picture-word interference and the response-exclusion hypothesis. *Cortex* 48, 363–372. doi: 10.1016/j.cortex.2011.04.025
- Mulatti, C., and Coltheart, M. (2014). Color naming of colored non-color words and the response-exclusion hypothesis: a comment on Mahon et al. and on Roelofs and Piai. *Cortex* 52, 120–122. doi: 10.1016/j.cortex.2013.08.018
- Navarrete, E., Caccaro, A., Pavani, F., Mahon, B. Z., and Peressotti, F. (2015). With or without semantic mediation: retrieval of lexical representations in sign production. *J. Deaf Stud. Deaf Educ.* 20, 163–171. doi: 10.1093/deafed/enu045
- Navarrete, E., and Costa, A. (2009). The naming of gender-marked pronouns supports interactivity in models of lexical access. *Psicologica* 30, 301–321.
- Navarrete, E., Del Prato, P., and Mahon, B. Z. (2012). Factors determining semantic facilitation and interference in the cyclic naming paradigm. *Front. Psychol.* 3:38. doi: 10.3389/fpsyg.2012.00038
- Navarrete, E., Del Prato, P., Peressotti, F., and Mahon, B. Z. (2014). Lexical selection is not by competition: evidence from the blocked naming paradigm. *J. Mem. Lang.* 76, 253–272. doi: 10.1016/j.jml.2014.05.003
- Navarrete, E., and Mahon, B. Z. (2013). A rose by any other name is still a rose: reinterpreting the picture-word experiment of Hantsch and Mädebach. *Lang. Cogn. Process.* 28, 701–706. doi: 10.1080/01690965.2012.682071
- Navarrete, E., Mahon, B. Z., and Caramazza, A. (2010). The cumulative semantic cost does not reflect lexical selection by competition. *Acta Psychol.* 134, 279–289. doi: 10.1016/j.actpsy.2010.02.009
- Oppenheim, G. M., Dell, G. S., and Schwartz, M. F. (2007). Cumulative semantic interference as learning. *Brain Lang.* 103, 175–176. doi: 10.1016/j.bandl.2007.07.102
- Oppenheim, G. M., Dell, G. S., and Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition* 114, 227–252. doi: 10.1016/j.cognition.2009.09.007
- Perry, C., Ziegler, J. C., and Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychol. Rev.* 114, 273–315. doi: 10.1037/0033-295X.114.2.273
- Protopapas, A. (2007). CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behav. Res. Methods* 39, 859–862. doi: 10.3758/BF03192979
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociol. Methodol.* 25, 111–163. doi: 10.2307/271066
- Rapp, B., and Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychol. Rev.* 107, 460–499. doi: 10.1037/0033-295X.107.3.460
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition* 42, 107–142. doi: 10.1016/0010-0277(92)90041-F
- Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: retrieval of verbs. *Cognition* 47, 59–87. doi: 10.1016/0010-0277(93)90062-Z
- Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychol. Rev.* 110, 88–125. doi: 10.1037/0033-295X.110.1.88
- Roelofs, A. (2006). Context effects of pictures and words in naming objects, reading words, and generating simple phrases. *Q. J. Exp. Psychol.* 59, 1764–1784. doi: 10.1080/17470210500416052
- Roelofs, A., Meyer, A. S., and Levelt, W. J. (1996). Interaction between semantic and orthographic factors in conceptually driven naming: comment on Starreveld and La Heij (1995). *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 246–251. doi: 10.1037/0278-7393.22.1.246
- Roelofs, A., and Piai, V. (2013). Associative facilitation in the Stroop task: comment on Mahon et al. (2012). *Cortex* 49, 1767–1769. doi: 10.1016/j.cortex.2013.03.001
- Roelofs, A., Piai, V., and Schriefers, H. (2011). Selective attention and distractor frequency in naming performance: comment on Dhooze and Hartsuiker

- (2010). *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1032–1038. doi: 10.1037/a0023328
- Roelofs, A., Piai, V., and Schriefers, H. (2013). Context effects and selective attention in picture naming and word reading: competition versus response exclusion. *Lang. Cogn. Process.* 28, 655–671. doi: 10.1080/01690965.2011.615663
- Rosinski, R. (1977). Picture–word interference is semantically based. *Child Dev.* 48, 643–647. doi: 10.2307/1128667
- Runnqvist, E., Strijkers, K., Alario, F. X., and Costa, A. (2012). Cumulative semantic interference is blind to language: implications for models of bilingual speech production. *J. Mem. Lang.* 66, 350–869. doi: 10.1016/j.jml.2012.02.007
- Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *J. Mem. Lang.* 75, 27–44. doi: 10.1016/j.jml.2014.04.006
- Schriefers, H., Meyer, A. S., and Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: picture–word interference studies. *J. Mem. Lang.* 29, 86–102. doi: 10.1016/0749-596X(90)90011-N
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Ann. Statist.* 6, 461–464. doi: 10.1214/aos/1176344136
- Spalek, K., Damian, M. F., and Bölte, J. (2013). Is lexical selection in spoken word production competitive? Introduction to the special issue on lexical competition in language production. *Lang. Cogn. Process.* 28, 597–614. doi: 10.1080/01690965.2012.718088
- Starreveld, P. A., La Heij, W., and Verdonchot, R. G. (2013). Time course analysis of the effects of distractor frequency and categorical relatedness in picture naming: an evaluation of the response exclusion account. *Lang. Cogn. Process.* 28, 633–654. doi: 10.1080/01690965.2011.608026
- Treccani, B., and Mulatti, C. (2015). Semantic(effects) in the word–word interference task: a comment on Roelofs, Piai, and Schriefers (2013). *Lang. Cogn. Neurosci.* 30, 700–703. doi: 10.1080/23273798.2014.1002797
- Tree, J. J., and Hirsh, K. W. (2003). Sometimes faster, sometimes slower: associative and competitor priming in picture naming with young and elderly participants. *J. Neurol.* 16, 489–514. doi: 10.1016/S0911-6044(02)00005-2
- Vigliocco, G., and Franck, J. (1999). When sex and syntax go hand in hand: gender agreement in language production. *J. Mem. Lang.* 40, 455–478. doi: 10.1006/jmla.1998.2624
- Vigliocco, G., Lauer, M., Damian, M. F., and Levelt, W. J. M. (2002). Semantic and syntactic forces in noun phrase production. *J. Exp. Psychol. Learn. Mem. Lang.* 28, 46–58.
- Vitkovitch, M., and Cooper-Pye, E. (2012). My word! Interference from reading object names implies a role for competition during picture name retrieval. *Q. J. Exp. Psychol.* 65, 1229–1240. doi: 10.1080/17470218.2012.655699
- Vitkovitch, M., Cooper-Pye, E., and Ali, L. (2010). The long and the short of it! Naming a set of prime words before a set of related picture targets at two different intertrial intervals. *Eur. J. Cogn. Psychol.* 22, 161–171. doi: 10.1080/09541440902743348
- Vitkovitch, M., Cooper-Pye, E., and Leadbetter, A. (2006). Semantic priming over unrelated trials: evidence for different effects in word and picture naming. *Mem. Cogn.* 34, 715–725. doi: 10.3758/BF03193590
- Vitkovitch, M., and Humphreys, G. W. (1991). Perseverant naming errors in speeded picture naming: it's in the links. *J. Exp. Psychol. Learn. Mem. Cogn.* 17, 664–680.
- Wheeldon, L. R., and Monsell, S. (1994). Inhibition of spoken word production by priming a semantic competitor. *J. Mem. Lang.* 33, 332–356. doi: 10.1006/jmla.1994.1016

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

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



“When” Does Picture Naming Take Longer Than Word Reading?

Andrea Valente^{1,2*}, Svetlana Pinet¹, F.-Xavier Alario¹ and Marina Laganaro²

¹ Aix-Marseille Université, CNRS, LPC UMR 7290, Marseille, France, ² Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland

OPEN ACCESS

Edited by:

Simone Sulpizio,
University of Trento, Italy

Reviewed by:

Francesca Peressotti,
Università degli Studi di Padova, Italy
Claudio Mulatti,
Università degli Studi di Padova, Italy

*Correspondence:

Andrea Valente
valentea78@gmail.com

Specialty section:

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

Received: 03 August 2015

Accepted: 08 January 2016

Published: 25 January 2016

Citation:

Valente A, Pinet S, Alario F-X
and Laganaro M (2016) “When” Does
Picture Naming Take Longer Than
Word Reading? *Front. Psychol.* 7:31.
doi: 10.3389/fpsyg.2016.00031

Differences between the cognitive processes involved in word reading and picture naming are well established (e.g., visual or lexico-semantic stages). Still, it is commonly thought that retrieval of phonological forms is shared across tasks. We report a test of this second hypothesis based on the time course of electroencephalographic (EEG) neural activity, reasoning that similar EEG patterns might index similar processing stages. Seventeen participants named objects and read aloud the corresponding words while their behavior and EEG activity were recorded. The latter was analyzed from stimulus onset onward (stimulus-locked analysis) and from response onset backward (response-locked analysis), using non-parametric statistics and the spatio-temporal segmentation of ERPs. Behavioral results confirmed that reading entails shorter latencies than naming. The analysis of EEG activity within the stimulus-to-response period allowed to distinguish three phases, broadly successive. Early on, we observed identical distribution of electric field potentials (i.e., topographies) albeit with large amplitude divergences between tasks. Then, we observed sustained cross-task differences in topographies accompanied by extended amplitude differences. Finally, the two tasks again revealed the same topographies, with significant cross-task delays in their onsets and offsets, and still significant amplitude differences. In the response-locked ERPs, the common topography displayed an offset closer to response articulation in word reading compared with picture naming, that is the transition between the offset of this shared map and the onset of articulation was significantly faster in word reading. The results suggest that the degree of cross-task similarity varies across time. The first phase suggests similar visual processes of variable intensity and time course across tasks, while the second phase suggests marked differences. Finally, similarities and differences within the third phase are compatible with a shared processing stage (likely phonological processes) with different temporal properties (onset/offset) across tasks. Overall, our results provide an overview of when, between stimulus and response, word reading and picture naming are subtended by shared- versus task-specific neural signatures. This in turn is suggestive of when the two tasks involve similar vs. different cognitive processes.

Keywords: ERPs, topographic analysis, word production, reading aloud, object naming

INTRODUCTION

It is well known that the time required to name an object is greater than the time required to read aloud its name (Cattell, 1885). This effect can resist even intensive training (Fraisse, 1964; Theios and Amrhein, 1989; Ferrand, 1999). This appears to be one of the clearest and most ubiquitously replicated pieces of evidence in psycholinguistics, and it has been object of scientific investigation since the very early stages of psycholinguistics.

Given that no controversy exists on such observation, efforts have been conveyed toward the understanding of its causes. Some accounts relied on differences at the level of the visual discriminability of the stimuli (see for instance Theios and Amrhein, 1989). Words are more easily processed perceptually if compared to a pictorial representation of the object they refer to. Nonetheless, several studies demonstrated that words and pictures are equally discriminable stimuli, and therefore discriminability cannot really account for differences in response speed (e.g., Fraisse, 1984; Theios and Amrhein, 1989).

Alternatively, it has been argued that pictures can be named in different ways, while only one response is possible for a written word (Fraisse, 1969; Theios, 1975), the so called “uncertainty factor” (see Ferrand, 1999).

If we narrow the discussion down to specific processing accounts, it has been submitted that naming a picture differs from word reading in some fundamental cognitive aspects. For instance, naming a depicted object would require some additional processing steps reading a word does not call for. When presented with a picture, the speaker has to recognize the object it represents. This is achieved by retrieving its visuo-semantic properties, prior to selection of the corresponding lexical item (e.g., Glaser, 1992). Conversely, reading a word aloud (in the alphabetic languages most commonly studied), can in principle be done by performing a conversion of the graphemes (the written form of a phoneme) in a phonological output, dispensing with the need of an extensive retrieval of semantic information necessary for object recognition (Coltheart et al., 1993; Indefrey and Levelt, 2004; Price et al., 2006).

Evidence has also been reported of an early activation of semantic information on presentation of both auditory (e.g., Pulvermüller et al., 2005) and written (Dell’Acqua et al., 2007) words and non-words. Furthermore, a thorough retrieval of semantic information is expected in reading aloud tasks involving semantic categorization and decision (see for instance Mulatti et al., 2009). Nonetheless, the involvement of semantics should occur with different weights in picture naming vs. word reading tasks in which participants are instructed to read aloud words appearing on a screen, and this has been repeatedly used as one of the arguments to explain, why naming a picture takes longer than reading the corresponding word (Theios and Amrhein, 1989, but see Janssen et al., 2011).

Despite these differences, some processing steps do appear to be equally needed for successful performance in both picture naming and word reading. For instance, it is commonly thought that phonological processing – that is, the retrieval of the word’s phonological form necessary to implement the articulatory gestures – is shared in both tasks (e.g., Coltheart et al., 2001).

In accordance to this, picture naming and word reading are assumed to involve similar outputs triggered by distinct inputs. Both behavioral and neuroimaging data have been marshaled in support of this hypothesis. Roelofs (2004) investigated initial segment planning – measured as the facilitation in reaction times – by using mixed vs. blocked picture naming and word reading trials. The idea behind the paradigm was that if the phonological processing stage were common to both picture naming and word reading, then a phonological facilitation should be observed not only when the tasks are blocked, but also when they are mixed (i.e., pictures and words alternated). Results supported the authors’ hypothesis. Price et al. (2006) investigated the neuronal basis of object naming compared with word reading in a functional neuroimaging (fMRI) study. Results revealed that the areas of speech production selectively activated during object naming were the same that were recruited during the reading of words, though in word reading the activation was comparatively enhanced.

The converging anatomical substrate supports the assumption that retrieval of the phonological form of the word to be uttered is comparable, whether one has to name an object or to read the corresponding word. In this context, the primary aim of the present study is to characterize the contrastive temporal signatures of picture naming and word reading; we do so by comparing directly the electroencephalographic (EEG) correlates of the two tasks, while taking into account their typically different response latencies. These contrasts can ultimately clarify the contrasts between shared and specific processes underlying the two tasks.

As noted above, picture naming and word reading differ on at least three major aspects: the specific cognitive processes they are assumed to involve, the moment in time in which such processes are triggered, and the speed at which responses are given. For this reason, ERP waveform analysis will be associated with topographic pattern (microstate) analysis (Murray et al., 2008). Topographic analysis is a reference-free methodology useful to partition the evoked potentials in periods of stable topographic activity, corresponding to “a period of coherent synchronized activation of a large-scale neuronal network,” which represent “the basic building blocks of information processing” (Brunet et al., 2011). It is useful to clarify that highlighting a serial succession of periods of stable topographic activity does not constitute an endorsement of a serial organization of the processing stages envisaged by some cognitive models. Each period of topographic stability can surely conglomerate the brain’s parallel processing of different types of information. Still, it is thought to represent a functional integrative step necessary to accomplish the cognitive task at hand (Brunet et al., 2011). This point has a particular relevance given the issues addressed in the present study; topographic analysis can inform us on whether specific topographic maps are present both in picture naming and in word reading, with additional information on their temporal signatures, which is useful to draw conclusions on cross-task differences in stages of information processing.

According to the theoretical accounts concerning the cognitive processes underlying picture naming and word reading, cross-task ERP differences should be detectable in early time

windows following visual encoding, since this specific time window is thought to involve cross-task specificities (extensive lexico-semantic vs. primarily ortho-phonological processing). In previous ERP studies, evidence has been reported that these task-specific processing stages are engaged in the time window following visual encoding and preceding retrieval of the phonological codes (Bentin et al., 1999; Simon et al., 2004; Indefrey, 2011; Hauk et al., 2012). In contrast, between-task similarities in the electrophysiological signatures are expected in later time windows approaching response articulation, in which it is likely to expect a primary involvement of phonological codes in both tasks.

MATERIALS AND METHODS

Method

The present study was approved by the local ethics committee of the faculty of Psychology and Educational Sciences of the University of Geneva, and carried out in accordance with the recommendations of the Swiss Federal Act on Research involving Human Beings. All participants gave their written informed consent in accordance with the Declaration of Helsinki.

A total of 24 healthy participants recruited among university students (aged between 18–30, mean: 22.8, *SD*: 3.5; three men) participated in the study. The scoring and data analysis led to the exclusion of seven participants, while 17 were retained (see below for details).

All participants gave informed consent and received monetary compensation for their participation in the study. They were right-handed as assessed by the Edinburgh Handedness Scale (Oldfield, 1971). They were all native French speakers.

Material

We used a set of 120 black-and-white line drawings and their corresponding words extracted from two French databases (Alario and Ferrand, 1999; Bonin et al., 2003).

All pictures had a name agreement above 75% (mean = 92.5%). This was done to minimize the odds of atypical responses. The stimuli were monosyllabic ($N = 40$), bisyllabic ($N = 60$), and trisyllabic ($N = 20$) words, of lexical frequency ranging from 0.13 to 227 occurrences per million words (mean = 17.3) according to the French database Lexique (New et al., 2004). The same 120 items (i.e., words and pictures) were presented in each task. Pictures consisted of 280×280 pixels black-line drawings, while the corresponding words were displayed in Courier New 18-point font, in white color on a gray background.

Procedure

Participants were tested individually in a soundproof cabin. They sat at about 60 cm from the computer screen. The stimuli were presented with the E-Prime software (Psychology Software Tools, Inc., [E-Studio], 2012) and appeared in a pseudo-randomized order, that is semantic and phonological neighbors were never presented in strict succession.

All participants were familiarized with the pictures before performing the task. They were shown the set of pictures associated with the corresponding written words in order to resolve any doubts or non-recognitions. To familiarize participants with the task, a training part was administered before the experimental session involving trials with the same temporal sequence than those used in the experiment. The order of picture naming and word reading blocks was counterbalanced across participants.

Picture Naming

Each trial started with a fixation cross presented for 500 ms, followed by a 200 ms blank screen and finally by the picture which was displayed for 1500 ms on a gray background.

Participants were instructed to name the pictures overtly, as quickly and accurately as possible while responses were recorded with a microphone.

The maximum delay conceded for articulation was 2000 ms. Responses not given within this time interval were classified as “no responses.”

Word Reading

The timing sequence of the trials was identical to the picture naming task with words, instead of the pictures, presented on the screen for 1000 ms. This shorter duration was chosen, because faster response latencies were expected in word reading.

Processing of Verbal Responses

Behavioral analyses were conducted on the sample of participants retained for the ERP analysis, after exclusion of participants with artifact-contaminated EEG signal. Seventeen participants (aged 18–29, mean: 22.6, *SD*: 3.1) were finally retained.

Response latencies, defined as the time elapsing between stimuli (picture and word) onset and the acoustic onset of response articulation, were estimated with Check Vocal (Protopapas, 2007). This software allows to visualize, both speech waveforms and spectrograms of each response in order to identify the speech onset.

EEG Acquisition and Processing

We used the Active-Two Biosemi EEG system (Biosemi V.O.F., Amsterdam, Netherlands) in its high-density montage, with 128 channels covering the scalp. Signals were sampled at 512 Hz with a band-pass filtering set between 0.16 and 100 Hz.

The EEG signal was calculated against the average reference and bandpass-filtered to 0.1–40 Hz.

Each trial in both tasks and for each participant was inspected visually for various forms of artifact contamination (blinks, eye movements, skin, or muscular artifacts) and noisy channels. An automated selection criterion, highlighting channels displaying amplitudes oscillations reaching $\pm 100 \mu\text{V}$, was also applied. Trials containing artifacts were excluded from ERP averaging. As a heuristic criterion, only participants with at least 60 usable trials in each task were retained for further analyses. For the waveform analysis (detailed below), stimulus-aligned epochs were extracted with a baseline correction of 100 ms; for the topographic analysis, no baseline correction was applied to the ERPs.

In both picture naming and word reading, stimulus-aligned and response-aligned epochs of 400 ms were averaged across participants in both conditions in order to obtain a Grand-Mean of ERPs for each task. Stimulus-aligned epochs were locked to picture onset in the picture naming task and to word onset in the word reading task. Response-aligned epochs were locked to the onset of articulation in both tasks.

EEG Analyses

Electroencephalographic analyses were performed in two main steps: waveform analysis and topographic pattern analysis.

Waveform analysis

First, a sample-wise ERP waveform analysis was performed on both stimulus- and response-locked ERPs in order to assess at which time points significant amplitude differences were present between picture naming and word reading. We compared both conditions time-locked to the stimulus onward (from -100 to 400 ms) and to vocal onset backward (up to -400 ms).

ERP waveforms were analyzed by means of a cluster-based non-parametric analysis (Maris and Oostenveld, 2007). This technique allows to compare each point in time (~every 2 ms) and channel between two conditions while correcting for multiple comparisons by taking into account spatial (four neighboring channels) and temporal (two successive time-points) adjacency: only clusters over a given significance level were kept. The level of significance was determined by building a distribution stemming from the data itself by successive random permutations of the two experimental conditions (picture naming and word reading).

Topographic pattern analysis

Amplitude differences between experimental conditions within a given time-window may have different causes. This is because the clustering algorithm used to partition the ERP data only considers the spatial configuration of activity (relative intensity across electrode sites), but not the absolute intensity at each electrode site. Significant differences may be detected when the same scalp electric fields are present in overlapping time intervals but with different intensity. Amplitude differences can also occur when spatial differences in the distribution of the potentials at the scalp are present (different topographic maps, revealing different brain generators), or even if the same scalp electric fields are present, but appear in different time-windows between conditions (i.e., if they are shifted in time). To assess the precise origin of the amplitude differences determined above, we performed two analyses.

First, we ran a sample-wise topographic analysis of variance (TANOVA). This method identifies the time periods during which topographic differences were present between tasks. The TANOVA is a non-parametric randomization test of the global dissimilarity measures between different experimental conditions or groups (Murray et al., 2008), useful to determine at which time points between stimulus and response different scalp topographies are present across the conditions of interest. As an empirical criterion, only topographic differences lasting more than 30 ms were considered and an alpha value of 0.01 was adopted.

Secondly, a microstate analysis (spatio-temporal segmentation) was performed on the ERP Grand-Means to identify the electrophysiological and temporal signatures of the building blocks of information processing present in picture naming and in word reading. This methodology clusters the ERP Grand-Means in a series of periods of quasi-stable electrophysiological activity (template maps) that best explain the global variability of the dataset. Only the spatial configuration of the maps, but not their intensity, is taken into account. Additional information can be obtained concerning the duration and other dependent measures of the stable periods in different conditions or groups. Any modification of the spatial configuration of the electric field measured at the scalp is unequivocally interpreted as indicating a different pattern of cerebral sources, namely a difference in the information processing the brain is engaged in (e.g., Pascual-Marqui et al., 1995). The microstate analysis was performed as follows. A spatio-temporal segmentation was computed on the Grand-Mean ERPs of both picture naming and word reading, in both the stimulus- and response-aligned conditions separately, using an optimized agglomerative clustering algorithm, the topographic atomize and agglomerate hierarchical clustering algorithm (TAAHC; Murray et al., 2008).

In both the response- and stimulus-aligned conditions, the ERP Grand-Means of both picture naming and word reading were subjected to the clustering together, i.e., template maps were computed from a concatenation of the Grand-Means of both tasks. This was done with the purpose of maximizing information about similarities and differences in the ERP signal.

The spatio-temporal segmentation partitions the ERP Grand-Means in a series of periods of quasi stable electrophysiological activity, which summarize the data and are useful to determine which template map best explains the participants' ERPs in each experimental condition. A temporal post-processing was also performed, allowing to reassign segments with a short duration (less than 30 ms) to neighboring clusters and to merge together very highly spatially correlated maps (above 0.92).

At the end of each segmentation, we are provided with a set of quality measures indicating, which is the best segmentation among alternatives. Cross-validation and Krzanovski-Lai criterion were used to this end. Cross-validation is the ratio between global explained variance (GEV) and degrees of freedom (number of electrodes). Since this measure gets less reliable as the number of electrodes increases, it is associated with the Krzanovski-Lai criterion, which computes the dispersion of the segmentation (see Murray et al., 2008). Segmentations corresponding to both the CV minimum and Krzanovski-Lai measure peak are usually the most reliable, as they represent a reasonable compromise between compression of the data and high GEV.

Once the group-averaged ERPs is segmented into a series of template maps, these can be tested by *back-fitting* them in the individual subject-averaged ERPs. This *back-fitting* procedure assigns each time point of each of the individual subjects ERPs to the Grand-Average's template maps it best correlates with. This yields a measure of the template maps' presence in each condition and allows to establish how well a cluster map explains individual

patterns of activity (GEV). Moreover, it provides information on map duration, first onset and last offset, Global Field Power and other dependent measures, which can subsequently be used for statistics.

These analyses were performed with the Cartool software (Brunet et al., 2011).

RESULTS

Behavioral Results

In both tasks, atypical responses (i.e., errors) and non-responses were excluded from further analysis (1.3% of the data). Response latencies above and below 3 SD were calculated for each participant in each task and excluded from further analysis (2% of the data).

On average, participants named pictures slower (mean RTs = 872 ms, $SD = 205$ ms) than they read the corresponding words (mean RTs = 560 ms, $SD = 101$ ms). The 312 ms difference was significant [$t(16) = 18,799$, $p < 0.001$].

ERP analysis

Stimulus-Aligned

In the stimulus-aligned condition (from 0 to 400 ms after stimulus onset) significant differences in amplitudes ($p < 0.05$) were observed between word reading and picture naming throughout the whole time-window of processing. These differences were particularly present over posterior electrodes and bilaterally from 100 to 400 ms post-stimulus (**Figure 1A**).

Results of the TANOVA showed that topographic differences between tasks also stretched across the whole time-window of processing, with the exception of the period comprised between about 75 and 150 ms after stimulus onset (see **Figure 1B**), corresponding to the temporal signature of the P1 component map.

The spatio-temporal segmentation of the stimulus-aligned Grand-Means explained 95.81% of the Global Variance, and revealed the presence of a total of six template maps. In **Figure 1B**, the five template maps starting from the P1 component map onward (map labeled “A”) are shown. In picture naming, the topographic configurations present in the P1 range (map “A”) and later in the 200–300 ms time window (map “D”) were highly correlated spatially and therefore labeled with the same template map by the clustering algorithm. When the same template map appears repeatedly in different non-overlapping time windows of the same Grand-Mean, it does not reflect comparable neuronal activity (e.g., Michel et al., 2009). For this reason, the later map has been relabeled differently in the figure, as it likely reflects a qualitatively different step of information processing following early visual encoding.

The application of the clustering algorithm resulted in a sequence of topographic maps, depicted in **Figure 1B** for the grand-averages of each task. Results of the spatio-temporal segmentation revealed that in an early time-window (comprised between about 75 and 150 ms after stimulus onset and thus compatible with visual encoding), the same topographic map (labeled “A”) was present in the grand-averages of both tasks.

In the waveform analysis, higher amplitudes were detected in word reading compared with picture naming (**Figure 1A**). The TANOVA corroborated the results of the spatio-temporal segmentation, revealing that the same topographic maps were predominant across tasks in the considered time-window (75–150 ms). A back-fitting was performed in the time window comprised between 0 and 400 ms from stimulus onset to test for the onsets, offsets and durations of map “A” across participants in both tasks. Results revealed that map “A” had a slightly later onset in picture naming (mean onset: 66 ms after picture onset) with respect to word reading (mean onset: 50 ms after word onset). A Wilcoxon signed-rank test proved the difference to be marginal ($z = -1,818$, $p = 0.07$). Map “A” also displayed a later offset in picture naming (mean offset: 155 ms after picture presentation) compared with word reading (mean offset: 132 ms after word presentation). The difference proved to be significant ($z = -2,301$, $p < 0.05$). Finally, no differences were found in map duration across tasks ($z = -1,086$, $p = 0.278$). **Figure 1C** illustrates the distributions of the individual onsets and offsets of map “A” in both picture naming and word reading.

The time window following visual encoding (starting from about 150 ms onward) was characterized by extensive amplitude differences, mainly located on posterior sites. In this time window, substantial topographic cross-task differences were detected. A back-fitting performed on the time window comprised between 160 and 300 ms after stimulus onset revealed that map “D,” characterized by posterior positivity and anterior negativity (**Figure 1B**), was significantly more present in picture naming compared with word reading (Pearson Chi Square computed on map presence across individuals: $\chi^2 = 14.43$, $p < 0.001$). In picture naming, Map “D” explained the 10% of the variance in the considered time-window (160–300 ms). The posterior characterization of amplitude differences as revealed by the waveform analysis seems consistent with the fact that in picture naming, map “D” was predominant in the considered time-window. Conversely, map named “B” was significantly more present in the word reading task ($\chi^2 = 6.10$, $p < 0.05$) and explained only the 3% of the variance in the time window comprised between 160 and 300 ms after word onset. The low explained variance can be attributed to the rapidly changing spatial configuration of map “B,” which is likely to be due to the unstable and transitory nature of the ERP activity in the considered time-window.

The back-fitting revealed that map named “C” had a negligible presence in individual ERPs. This is probably due to the transitional and unstable nature of this topographic map (see **Figure 1B**).

Amplitude differences were then sustained in the time window from about 250 ms to the end of the stimulus-locked analysis, corroborated by topographic differences identified by the TANOVA. These differences are however likely to be due to the very different time course of the processing stages specific of each task. In fact, the spatio-temporal segmentation performed on this time window revealed the presence, in both tasks, of the same period of topographic stability (map labeled “E”) characterized by posterior positivity and anterior negativity. This common map displayed noticeably different time courses

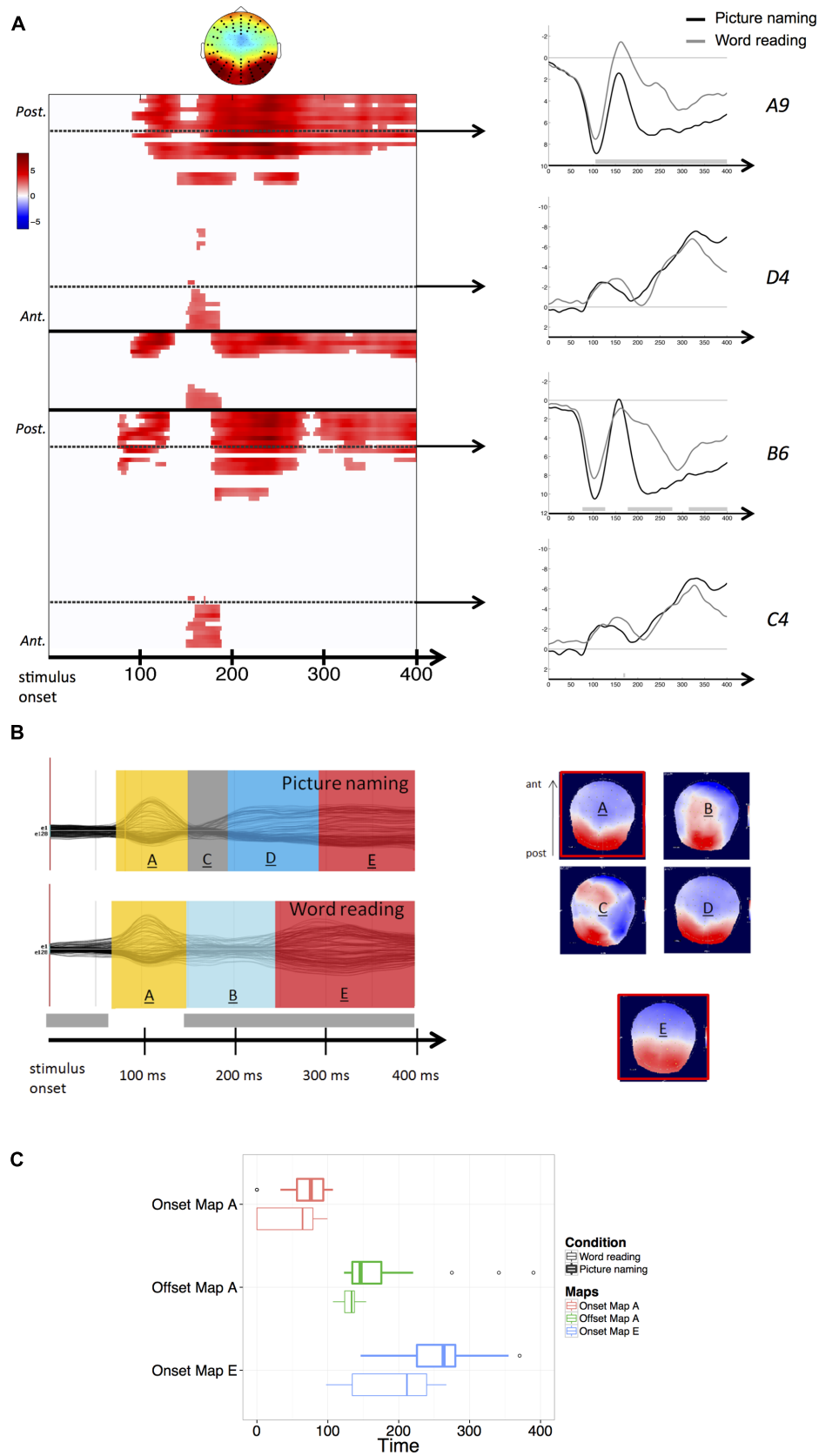


FIGURE 1 | Continued

FIGURE 1 | Continued

(A) Results of the stimulus-aligned waveform analysis. Values are masked by results of cluster-based non-parametric analysis: only significant values are plotted. Within the left panel, upper part corresponds to left hemisphere electrodes, middle part to midline electrodes, and lower part to right hemisphere electrodes; within each part, electrodes are ordered from posterior to anterior. Dashed lines outline representative electrodes, which time course is plotted separately (picture naming in black, word reading in gray). The topography represents the spatial distribution of the effect over each cluster (black dots outline electrodes within each cluster). **(B)** Results of the spatio-temporal segmentation on the stimulus-locked ERP Grand-Means of both tasks. Each period of topographic stability is displayed in the color bars with the information about its time course. The corresponding topographies are listed on the right (positive values in red, negative values in blue), with the common topographies marked in red. The gray bar on the temporal axis represents the periods of topographic difference between tasks, as revealed by the TANOVA. **(C)** Boxplots of distributions of individual onsets of maps A and E and offsets of map A, extracted from the back-fitting procedure for both picture naming (bold lines) and word reading (thin lines). Zero of times represent stimulus presentation.

between tasks. The back-fitting performed on the time window comprised between 100 and 400 ms after stimulus onset revealed that map “E” (explaining the 22% of the variance across tasks in the considered time-window) displayed an earlier onset in word reading (mean onset: 187 ms after word presentation) with respect to picture naming (mean onset: 252 ms after picture presentation). **Figure 1C** illustrates the distribution of the individual onsets of the common map “E” between tasks. A Wilcoxon signed-rank test proved the cross-task difference in the onset to be significant across participants ($z = -2,342$, $p < 0.05$).

Response-Aligned

In the response-aligned condition (from -400 ms to the vocal response onset), significant amplitude differences ($p < 0.05$) were observed between word reading and picture naming throughout the whole time-window of interest. Differences were observed earlier over anterior electrodes (from -400 to -300 ms) and more posteriorly in the following time-window closer to articulation. Again, effects were bilateral (**Figure 2A**).

The TANOVA revealed an extended period of topographic difference, stretching across the whole time-window of processing with the exception of the last period starting about 100 ms prior to the onset of articulation.

The spatio-temporal segmentation revealed the presence of three template maps (**Figure 2B**) – labeled “F,” “G,” and “H” – explaining 94.5% of the Global variance.

The template map labeled “F” corresponds to the common map (“E”) in the stimulus-aligned condition. These maps were, in fact, spatially correlated above 0.99.

All the three maps were common to both tasks, but maps “F” and “G” displayed different time courses. A back-fitting procedure was carried out in the time-window comprised between -400 and -60 ms before response articulation, revealing that in word reading the map labeled “F” explaining the 20% of the variance in the considered time-window, displayed an offset much closer to response articulation (mean map offset: 184 ms before articulation) compared with picture naming (mean map offset: 257 ms before articulation). This result proved to be significant across participants ($z = -2,580$, $p = 0.01$). A second back-fitting was performed in the time-window comprised between -380 and -60 ms to test for the duration of map “G” across tasks. The results revealed that map “G” had a longer duration in picture naming (mean duration: 243 ms) compared with word reading (mean duration: 113 ms). The result was significant ($z = -3,297$, $p < 0.01$).

Figure 2C illustrates the distribution of the offsets of map “F” and the duration of map “G” across participants and for each task. It is worthy of notice that the mean maps offset and duration calculated across participants might be different when compared to the mean onsets of the same maps in the ERP Grand-Means, because of variability across participants.

DISCUSSION

Participants named pictures slower than they read the corresponding words. This result is consistent with a vast literature (e.g., Cattell, 1885; Fraisse, 1964, 1969; Theios and Amrhein, 1989; Ferrand, 1999; Price et al., 2006; Riès et al., 2012).

Results of the ERP analysis will be discussed by focusing on three successive phases, tentatively defined on the basis of the degree of cross-task similarities and differences in the observed EEG patterns. These phases correspond roughly to the time-window between 75 and 150 ms after stimulus onset, a post-visual time-window ranging from about 150 to 250 ms, and a later time-window encompassing ERP activity close to response articulation.

No cross-task differences were detected in terms of the spatial configurations at the scalp present in the time-window between 75 and 150 ms after stimulus onset. Nevertheless, the visual topographies displayed a slightly different time course across tasks, and the waveform analysis showed that the visual ERP component in word reading displayed higher amplitude with respect to picture naming. These observations possibly stem from the different recruitment of the visual cortex due to the different types of visual stimuli (pictures vs. words), during a period in which visual encoding of the stimuli was presumably predominant. The observation of such cross-task differences in intensity and time course of the early map “A” does not stand in contrast with previous evidence that visual processing can occur in parallel with the cascaded activation of other pertinent informational codes. In this respect, evidence has been reported of an early activation of semantic information in both visual objects (e.g., Miozzo et al., 2015) and written words (e.g., Pulvermüller et al., 2005; Dell’Acqua et al., 2007). In accordance with this hypothesis, the shared “visual” map might possibly encompass a task-dependent variable degree of spreading of activation to semantics.

The time-window immediately following visual encoding (from about 150 to 250 ms) was characterized by extensive cross-task amplitude and topographic differences. In fact,

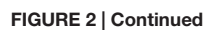


FIGURE 2 | Continued

(A) Results of the response-aligned waveform analysis. Values are masked by results of cluster-based non-parametric analysis: only significant values are plotted. Within the left panel, upper part corresponds to left hemisphere electrodes, middle part to midline electrodes, and lower part to right hemisphere electrodes; within each part, electrodes are ordered from posterior to anterior. Dashed lines outline representative electrodes, which time course is plotted separately (picture naming in black, word reading in gray). The topography represents the spatial distribution of the effect over each cluster (white dots outline electrodes within each cluster). **(B)** Results of the spatio-temporal segmentation on the response-locked ERP Grand-Means of both tasks. Each period of topographic stability is displayed in the color bars with the information about its time course. The corresponding topographies are listed on the right (positive values in red, negative values in blue), with the common topographies marked in red. The gray bar on the temporal axis represents the periods of topographic difference between tasks, as revealed by the TANOVA. **(C)** Boxplots of distributions of individual offsets of maps F and durations of map G, extracted from the back-fitting procedure for both picture naming (bold lines) and word reading (thin lines). Zero of times in the boxplot of the offset of map F represents voice onset.

two different scalp topographies were predominantly present between tasks (map “D” in picture and map “B” in word reading in **Figure 1B**). Even though map “B” was characterized by a transitory and rapidly changing spatial configuration, these results point to substantial electrophysiological differences between picture naming and word reading in the “post-visual” time window, which can in turn be interpreted as functional differences in the information processing occurring between tasks. More specifically, naming a picture is thought to require extensive retrieval of the semantic information associated with the stimulus, without which no recognition of the object would be achieved (Glaser and Glaser, 1989; Theios and Amrhein, 1989; Indefrey and Levelt, 2004; Price et al., 2006). The hypothesis that different processing stages are implied between picture naming and word reading has been proposed in a previous ERP study (Yum et al., 2011) in which the authors reported diverging ERP correlates associated with the processing of pictures vs. words in the time-window compatible with the N170 component (between 150 and 200 ms after stimulus presentation).

Amplitude and topographic differences were also detected in the later time-window from about 250 ms from stimulus onset onward. These are mainly due to the remarkable cross-task difference in response latencies and time course of the information processed between tasks. Indeed, the spatio-temporal segmentation on this time window revealed the presence in both tasks of the same period of topographic stability characterized by posterior positivity and anterior negativity. The presence of this shared topographic pattern indicates that the same underlying brain generators were active in both tasks thought with different time courses, namely significantly earlier in word reading with respect to picture naming. The fact that these similar topographic patterns were present in a time-window approaching the onset of response articulation, in which it is likely to expect retrieval of the phonological codes necessary to initiate utterance of the words, gives room to the tentative hypothesis that they, at least partially, convey retrieval of the phonological codes. This interpretation is in line with previous neuroimaging evidence reporting that picture naming and word reading rely on the activation of comparable speech production areas (e.g., Price et al., 2006) and would hence support the hypothesis that the processing of phonology is shared across picture naming and word reading (Roelofs, 2004; Price et al., 2006).

Furthermore, the fact that the shared topographic map displayed an earlier onset (about 70 ms) in word reading

compared with picture naming seems to suggest that this specific operation may occur earlier in word reading, which could account for the shorter response latencies observed in this task.

Hypotheses that retrieval of the phonological form occur earlier in word reading with respect to picture naming have been advanced in previous studies by positing a stronger connection between written words and phonological codes (Glaser and Glaser, 1989) and by assuming that contrary to picture naming, in word reading no semantic processing followed by selection of the lexical entry is required (Theios and Amrhein, 1989). In fact, in word reading the phonology could in principle be directly accessed via a grapheme-to-phoneme conversion, whereas in picture naming the retrieval of the phonological form is thought to be conditional to the retrieval of semantic information (e.g., Price et al., 2006) or concomitant to the semantic processing necessary to recognize the depicted object (Miozzo et al., 2015; see also Abdel-Rahman et al., 2003). The earlier onset of the map “E” in word reading would be compatible with theories positing a faster access to phonological codes in word reading with respect to picture naming. Whether this earlier access to phonology in word reading stems from access to phonology from the grapheme to phoneme conversion stage, or to the differential levels of recruitment of semantic information across tasks, the observation might partially account for the shorter latencies observed in word reading compared with picture naming.

It is useful to clarify that more specific interpretations concerning the information processed during the shared periods of stable topographic activity cannot be completely excluded. For instance, one can posit that the activation of phonological codes in word reading might automatically spread to semantics. However, the experimental design here adopted does not allow to extensively test these hypotheses, insofar as no factors capable of affecting phonological or semantic processing were manipulated.

In the response-locked ERPs, amplitude and topographic differences (identified by the TANOVA) are again ascribable to the different time course of the processing stages involved. The results of the segmentation allowed to identify three common periods of topographic stability, two of which displayed very different time courses.

The same topographic map identified in the final stimulus-locked period was detected in the segmentation of the response-locked ERPs. Interestingly, in this case the map displayed an offset much closer to response articulation in word reading compared

with picture naming. In other words, the transition between the offset of this topographic map and the onset of articulation was significantly shorter in word reading with respect to picture naming.

Considering that this period of quasi stable electrophysiological scalp activity was common in both tasks, and surmising that it could convey the retrieval of the phonological form of the words, then the transition between this phase and the moment in which articulation could be initiated appears to be faster in word reading compared with picture naming.

This can be explained by intervening pre-articulation monitoring processes specific of picture naming, and could for instance reflect the higher level of uncertainty one has to face when naming a picture compared with reading a word aloud (e.g., Fraisse, 1969; Ferrand, 1999), leading to more cautious responses. For instance in an overt picture naming task, Valente et al. (2014) reported effects of the variables name agreement and image agreement in the time window preceding response articulation, supporting the hypothesis of a monitoring of response before the onset of articulation.

Another possible explanation might be the cascading of information from earlier stages of encoding. Even though the present study was not aimed to tackle directly the issue of cascade processing, one could hypothesize that phonological information can be differentially activated depending on the specific task one has to perform. This could in turn affect the moment in which articulation can be undertaken. Such hypothesis is also consistent with evidence reported by Price et al. (2006) of a higher activation of speech production areas in word reading compared with picture naming. Likewise it has been posited that, in word reading, articulation can be initiated on the basis of more partial information (Hennessey and Kirsner, 1999). Our results are not inconsistent with this assumption, in so far as the transition between the offset of the common topographic map and the onset of articulation was significantly faster when participants had to read words compared with when they had to name pictures.

Although further investigation is required to directly address the issue, this assumption would also reinforce an explanation on why and when word reading is faster than picture naming.

REFERENCES

- Abdel-Rahman, R. A., van Turenhout, M., and Levelt, W. J. M. (2003). Phonological encoding is not contingent on semantic feature retrieval: an electrophysiological study on object naming. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 850–860. doi: 10.1037/0278-7393.29.5.850
- Alario, F. X., and Ferrand, L. (1999). A set of 400 pictures standardized for French: norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behav. Res. Methods Instrum. Comput.* 31, 531–552. doi: 10.3758/BF03200732
- Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., and Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *J. Cogn. Neurosci.* 11, 235–260. doi: 10.1162/089892999563373
- Bonin, P., Peereman, R., Malardier, N., Méot, A., and Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behav. Res. Methods Instrum. Comput.* 35, 158–167. doi: 10.3758/BF03195507

CONCLUSION

This study sought to investigate how picture naming and word reading differ over the time course of processing from stimulus to response. We offer evidence that the same periods of stable topographic activity were present across tasks in two time-windows, compatible with processing of visual information and retrieval of the phonological form. The latter period of stable topographic activity, close to response articulation, displayed different time courses across tasks, with an earlier onset with respect to stimulus presentation in word reading than picture naming. This result can be tentatively interpreted as a faster access to phonological codes from written words than pictorial stimuli, which do require an extra semantic stage necessary for object recognition and identification. Likewise, the common topographic map thought to partially convey phonological processing, had an offset closer to response articulation in word reading compared with picture naming, suggesting that response articulation can be initiated comparably faster in word reading, once phonological information becomes available. Altogether, our interpretation provides some indications regarding the temporal origin of faster responses in word reading compared with picture naming.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation under Grant P2GEP1_152010 to AV, and grants PP001-118969 and PP00P1_140796 to ML, and by the French Ministère de l'Enseignement Supérieur et de la Recherche under a doctoral MNRT Grant to SP by the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013 Grant agreement n 263575), and by the Brain and Language Research Institute (Aix-Marseille Université: A*MIDEX grant ANR-11-IDEX-0001-02 and LABEX grant ANR-11-LABX-0036). We thank the "Fédération de Recherche 3C" (Aix-Marseille Université) for institutional support.

- Brunet, D., Murray, M. M., and Michel, C. M. (2011). Spatiotemporal analysis of multichannel EEG: CARTOOL. *Comput. Intell. Neurosci.* 2011, 813870. doi: 10.1155/2011/813870
- Cattell, J. M. (1885). Über die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern Und Farben. *Philos. Stud.* 2, 635–650.
- Coltheart, M., Curtis, B., Atkins, P., and Haller, M. (1993). Models of reading aloud: dual-route and parallel-distributed-processing approaches. *Psychol. Rev.* 100, 589–608. doi: 10.1037/0033-295X.100.4.589
- 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
- Dell'Acqua, R., Pesciarelli, F., Jolicoeur, P., Eimer, M., and Peressotti, F. (2007). The interdependence of spatial attention and lexical access as revealed by early asymmetries in occipito-parietal ERP activity. *Psychophysiology* 44, 436–443. doi: 10.1111/j.1469-8986.2007.00514.x
- Ferrand, L. (1999). Why naming takes longer than reading? The special case of Arabic numbers. *Acta Psychol.* 100, 253–266. doi: 10.1016/S0001-6918(98)00021-3

- Fraisse, P. (1964). Le temps de reaction verbale: I. Denomination et lecture [Verbal reaction times: I. Naming and reading]. *Année Psychol.* 64, 21–46. doi: 10.3406/psy.1964.27145
- Fraisse, P. (1969). Why is naming longer than reading? *Acta Psychol.* 30, 96–103. doi: 10.1016/0001-6918(69)90043-2
- Fraisse, P. (1984). “Perceptual processing of words and drawings,” in *Perspectives in Psychological Experimentation, Toward The Year 2000*, eds V. Sarris and A. Parducci (Hillsdale, NJ: Lawrence Erlbaum), 151–163.
- Glaser, W. R. (1992). Picture naming. *Cognition* 42, 61–105. doi: 10.1016/0010-0277(92)90040-O
- Glaser, W. R., and Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *J. Exp. Psychol. Gen.* 118, 13–42. doi: 10.1037/0096-3445.118.1.13
- Hauk, O., Coutout, C., Holden, A., and Chen, Y. (2012). The time-course of single-word reading: evidence from fast behavioral and brain responses. *Neuroimage* 60, 1462–1477. doi: 10.1016/j.neuroimage.2012.01.061
- Hennessey, N. W., and Kirsner, K. (1999). The role of sub-lexical orthography in naming: a performance and acoustic analysis. *Acta Psychol.* 103, 125–148. doi: 10.1016/S0001-6918(99)00033-5
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* 2:255. doi: 10.3389/fpsyg.2011.00255
- Indefrey, P., and Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition.2002.06.001
- Janssen, N., Carreiras, M., and Barber, H. A. (2011). Electrophysiological effects of semantic context in picture and word naming. *Neuroimage* 57, 1243–1250. doi: 10.1016/j.neuroimage.2011.05.015
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Michel, C. M., Koenig, T., Brandeis, D., and Gianotti, L. R. R. (2009). *Electric Neuroimaging*. Cambridge: Cambridge University Press.
- Miozzo, M., Pulvermüller, F., and Hauk, O. (2015). Early parallel activation of semantics and phonology in picture naming: evidence from a multiple linear regression MEG study. *Cereb. Cortex* 25, 3343–3355. doi: 10.1093/cercor/bhu137
- Mulatti, C., Lotto, L., Peressotti, F., and Job, R. (2009). Speed of processing explains the picture-word asymmetry in conditional naming. *Psychol. Res.* 74, 71–81. doi: 10.1007/s00426-008-0182-2
- Murray, M. M., Brunet, D., and Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr.* 20, 249–264. doi: 10.1007/s10548-008-0054-5
- New, B., Pallier, C., Brysbaert, M., and Ferrand, L. (2004). Lexique 2: a new French lexical database. *Behav. Res. Methods Instrum. Comput.* 36, 516–524. doi: 10.3758/BF03195598
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Pascual-Marqui, R. D., Michel, C. M., and Lehmann, D. (1995). Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Trans. Biomed. Eng.* 42, 658–665. doi: 10.1109/10.391164
- Price, C. J., McCrory, E., Noppeney, U., Mechelli, A., Moore, C. J., Biggio, N., et al. (2006). How reading differs from object naming at the neuronal level. *Neuroimage* 29, 643–648. doi: 10.1016/j.neuroimage.2005.07.044
- Protopapas, A. (2007). CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behav. Res. Methods* 39, 859–862. doi: 10.3758/BF03192979
- Psychology Software Tools, Inc., [E-Studio] (2012). *Psychology Software Tools, Inc., [E-Studio]*. Available at: <http://www.pstnet.com>
- Pulvermüller, F., Shtyrov, Y., and Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *J. Cogn. Neurosci.* 17, 884–892. doi: 10.1162/0898929054021111
- Riès, S., Legou, T., Burle, B., Alario, F.-X., and Malfait, N. (2012). Why does picture naming take longer than word reading? The contribution of articulatory processes. *Psychon. Bull. Rev.* 19, 955–961. doi: 10.3758/s13423-012-0287-x
- Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Mem. Cognit.* 32, 212–222. doi: 10.3758/BF03196853
- Simon, G., Bernard, C., Largy, P., Lalonde, R., and Rebai, M. (2004). Chronometry of visual word recognition during passive and lexical decision tasks: an ERP investigation. *Int. J. Neurosci.* 114, 1401–1432. doi: 10.1080/00207450490476057
- Theios, J. (1975). “The components in response latency in simple human information processing tasks,” in *Attention and Performance V*, eds P. M. A. Rabbitt and S. Dornic (New York, NY: Academic Press), 418–440.
- Theios, J., and Amrhein, P. C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: reading, naming, and visual and conceptual comparisons. *Psychol. Rev.* 96, 5–24. doi: 10.1037/0033-295X.96.1.5
- Valente, A., Bürki, A., and Laganaro, M. (2014). ERP correlates of word production predictors in picture naming: a trial by trial multiple regression analysis from stimulus onset to response. *Front. Neurosci.* 8:390. doi: 10.3389/fnins.2014.00390
- Yum, Y. N., Holcomb, P. J., and Grainger, J. (2011). Words and pictures: an electrophysiological investigation of domain specific processing in native Chinese and English speakers. *Neuropsychologia* 49, 1910–1922. doi: 10.1016/j.neuropsychologia.2011.03.018

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

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



Asymmetric Switch Costs in Numeral Naming and Number Word Reading: Implications for Models of Bilingual Language Production

Michael G. Reynolds^{1*}, Sophie Schlöffel² and Francesca Peressotti^{3*}

¹ Visual Cognition Lab, Department of Psychology, Trent University, Peterborough, ON, Canada, ² Basque Center on Cognition, Brain and Language, San Sebastian, Spain, ³ Dipartimento di Psicologia Dello Sviluppo e Della Socializzazione, Università degli Studi di Padova, Padova, Italy

OPEN ACCESS

Edited by:

Simone Sulpizio,
University of Trento, Italy

Reviewed by:

Mikel Santesteban,
University of the Basque Country,
Spain
John George Grundy,
York University, Canada

*Correspondence:

Michael G. Reynolds
michaelchanreynolds@trentu.ca;
Francesca Peressotti
francesca.peressotti@unipd.it

Specialty section:

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

Received: 06 August 2015

Accepted: 16 December 2015

Published: 25 January 2016

Citation:

Reynolds MG, Schlöffel S and
Peressotti F (2016) Asymmetric
Switch Costs in Numeral Naming and
Number Word Reading: Implications
for Models of Bilingual Language
Production. *Front. Psychol.* 6:2011.
doi: 10.3389/fpsyg.2015.02011

One approach used to gain insight into the processes underlying bilingual language comprehension and production examines the costs that arise from switching languages. For unbalanced bilinguals, asymmetric switch costs are reported in speech production, where the switch cost for L1 is larger than the switch cost for L2, whereas, symmetric switch costs are reported in language comprehension tasks, where the cost of switching is the same for L1 and L2. Presently, it is unclear why asymmetric switch costs are observed in speech production, but not in language comprehension. Three experiments are reported that simultaneously examine methodological explanations of task related differences in the switch cost asymmetry and the predictions of three accounts of the switch cost asymmetry in speech production. The results of these experiments suggest that (1) the type of language task (comprehension vs. production) determines whether an asymmetric switch cost is observed and (2) at least some of the switch cost asymmetry arises within the language system.

Keywords: bilingualism, speech production, lexical decision, language switching, language comprehension, controlled processing

INTRODUCTION

How do individuals who speak more than one language (hereafter referred to as bilinguals) coordinate their language systems so as to produce continuous speech in a single language and yet switch to an appropriate language as required? One approach used to investigate this issue examines the costs that arise when bilinguals switch between languages (e.g., Von Studnitz and Green, 1997; Meuter and Allport, 1999; Thomas and Allport, 2000; Costa and Santesteban, 2004; Orfanidou and Sumner, 2005; Peeters et al., 2014). This approach utilizes the methods and reasoning developed in the task switching literature in order to gain insight into the processes underlying control over the bilingual language system (e.g., Allport et al., 1994; Rogers and Monsell, 1995; see Kiesel et al., 2010; Vandierendonck et al., 2010, for recent reviews). These methods examine whether there is a cost to switching languages on a trial-by-trial basis in response to discrete stimuli by comparing performance on trials where a language repeats (non-switch trials) to performance on trials where the language changes (switch trials). Evidence for control comes from worse performance on switch trials (longer response times and/or increased error rates) compared to non-switch trials.

One often reported finding in the context of language switching is that for unbalanced bilinguals the cost of switching to a stronger language (L1) is larger than the cost of switching to a weaker language (L2). This switch cost asymmetry was initially reported by Meuter and Allport (1999) and has since been replicated in a number of studies (e.g., Costa and Santesteban, 2004; Costa et al., 2006; Finkbeiner et al., 2006; Philipp et al., 2007; Verhoef et al., 2009; Macizo et al., 2012; Peeters et al., 2014). Presently, the switch cost asymmetry in unbalanced bilinguals is attributed to differences in relative language strength (Grainger and Beauvillain, 1987; Dijkstra and Van Heuven, 1998; Green, 1998; Van Heuven et al., 1998; Meuter and Allport, 1999; Costa and Santesteban, 2004; Finkbeiner et al., 2006; Peeters et al., 2014). Consistent with these accounts, Costa and colleagues (Costa and Santesteban, 2004; Costa et al., 2006) rule out absolute language strength, age of L2 acquisition, and language similarity as factors that give rise to the switch cost asymmetry. Furthermore, the asymmetry is reduced after extended practice (Meuter and Allport, 1999) and when languages are balanced in strength such that larger differences in relative proficiency are required in order to observe asymmetric switch costs when individuals are exceptionally proficient in two languages, to the point where highly proficient bilinguals will not show a switch cost asymmetry for unbalanced languages (Costa and Santesteban, 2004; Costa et al., 2006; Martin et al., 2013).

If the switch cost asymmetry in unbalanced bilinguals arises from differences in the relative strength of the two languages, then it seems to follow that asymmetric switch costs will be observed in any task where the two languages are unbalanced in strength. However, as can be seen in **Table 1**, there are many studies with unbalanced bilinguals that report symmetric switch costs, inconsistent with relative language strength being the sole determining factor for a switch cost asymmetry. A quick examination of **Table 1** reveals one possible source of such inconsistencies, namely the type of task. In general, asymmetric switch costs tend to be observed when the task is speech production and symmetric switch costs tend to be observed when language comprehension tasks are used (e.g., lexical decision or semantic categorization). Consistent with this possibility speech production and visual language comprehension differ in at least three important ways that could explain the different patterns of switch costs. First, different types of stimuli are used. Typically, speech production studies involve object naming whereas language comprehension studies use written words. This distinction is critical because objects (and numerals) are encoded differently than words (e.g., Humphreys et al., 1999; Damian, 2004). For instance, the visual representations of objects and words are thought to be stored in two separate lexicons: a pictogen system for objects (e.g., Humphreys et al., 1999; Coltheart, 2004) and an orthographic lexicon for words (e.g., Coltheart et al., 2001; Perry et al., 2007). Second, speech production relies on the retrieval of a phonological-lexical representation for production, whereas comprehension tasks such as lexical decision and semantic categorization rely on a search of orthographic-lexical and semantic information for a binary (yes/no) response. Finally, evidence suggests that language comprehension tasks such as lexical decision are more susceptible

to decision processes than naming tasks (Chumbley and Balota, 1984).

Although it is possible that the presence and absence of a switch cost asymmetry in unbalanced bilinguals is due to the type of task, it is necessary to rule out other competing explanations. As can be seen in **Table 1**, two other factors that are confounded with the presence/absence of the switch cost asymmetry are (1) the type of stimuli (univalent vs. bivalent) and (2) the predictability of the language switches.

The present experiments serve two goals. The first goal is to assess whether the presence of the switch cost asymmetry for unbalanced bilinguals in speech production tasks and its absence in language comprehension tasks, such as lexical decision, are a consequence of methodological differences or differences in how languages are activated during speech production and language comprehension tasks. The second goal is to discriminate between accounts of the switch cost asymmetry. In order to accomplish these goals, Experiments 1 and 2 assess whether naming yields symmetric switch costs like lexical decision when methodological differences are eliminated. Finally, Experiment 3 assesses whether languages are activated differently in naming and lexical decision by examining whether stimulus valence influences switch costs under mixed list presentation conditions in the naming task¹.

EXPERIMENTS 1A AND 1B

One factor that may determine whether a switch cost asymmetry is observed is the type of stimuli. The stimuli used in language switching experiments differ in two important ways. The first concerns stimulus valence, which refers to the correspondence between a stimulus and a task. Bivalent stimuli are those that are used to respond in both tasks, whereas univalent stimuli are those that correspond to a single task. The second concerns how likely it is that a stimulus belongs to a given language. In speech production, the stimuli tend to be numerals or pictures, which are named in both languages during an experiment (bivalent) and do not contain language specific information. In contrast, in lexical decision or semantic categorization studies, the stimuli tend to be different sets of written words for each language (univalent) and the words tend to be unique to one language (contain language

¹Here we chose to examine how the switch cost asymmetry in speech production was affected by factors used in lexical decision (univalent stimuli, predictable switches) for two reasons. First, incorporating the stimuli used in lexical decision experiments (written words) into a speech production task was feasible, whereas incorporating the bivalent stimuli used in naming (numerals and objects) into lexical decision and/or semantic categorization looks quite complicated. Indeed we could not think of an appropriate forced choice task with stimuli such as pictures and numerals, which require language switches. For example making parity judgments to numerals can be done without switching languages. The same is true if bivalent words (i.e., homographs) are used given that on a switch trial the homograph would yield a correct answer whether the subject switched languages or not (see Von Studnitz and Green, 1997; Thomas and Allport, 2000). Clearly the impossibility of pairing of bivalent stimuli with forced choice tasks also rendered a direct comparison between speech production and decision tasks problematic. Second, we chose the naming task since evidence suggests that it is more appropriate for investigating the processes involved in lexical processing. This is mainly due to the fact that the naming task avoids complications from decision processes observed in alternative forced choice tasks (e.g., Balota and Chumbley, 1984).

TABLE 1 | A summary of experiments examining switch costs as a function of task, predictability of switches, stimulus type, bilingual proficiency, and language strength.

Paper	Exp.	Switch costs	Task	Predictability	Stimuli	Proficiency	Languages
Christoffels et al., 2007	1	Symmetric	Naming	Random	Pictures	Unbalanced	Unbalanced L1 vs. L2
Costa and Santesteban, 2004	1	Asymmetric	Naming	Random	Pictures	Unbalanced	Unbalanced L1 vs. L2
Costa and Santesteban, 2004	2	Symmetric	Naming	Random	Pictures	Highly proficient	Balanced L1 vs. L2
Costa and Santesteban, 2004	3	Symmetric	Naming	Random	Pictures	Highly proficient	Balanced L1 vs. L2
Costa and Santesteban, 2004	4	Symmetric	Naming	Random	Pictures	Highly proficient	Unbalanced L1 vs. L3
Costa and Santesteban, 2004	5	Symmetric	Naming	Random	Pictures	Highly proficient	Balanced L1 vs. L2
Costa and Santesteban, 2004	5	Symmetric	Naming	Random	Pictures	Highly proficient	Balanced L1 vs. L2
Costa et al., 2006	1	Symmetric	Naming	Random	Pictures	Highly proficient	Balanced L1 vs. L2
Costa et al., 2006	2	Symmetric	Naming	Random	Pictures	Highly proficient	Unbalanced L2 vs. L3
Costa et al., 2006	3	Asymmetric	Naming	Random	Pictures	Highly proficient	Unbalanced L3 vs. L4
Costa et al., 2006	4	Asymmetric	Naming	Random	Pictures	Highly proficient	Unbalanced L1 vs. New
Declerck et al., 2012	1	Symmetric	Naming	Random	Numerals	Unbalanced	Unbalanced L1 vs. L2
Declerck et al., 2012	1	Symmetric	Naming	Random	Pictures	Unbalanced	Unbalanced L1 vs. L2
Fink and Goldrick, 2015	1	Symmetric	Naming	Random	Digits	Highly proficient	Balanced L1 vs. L2
Fink and Goldrick, 2015	1	Asymmetric	Naming	Random	Digits	Unbalanced	Unbalanced L1 vs. L2
Jackson et al., 2001	1	Asymmetric	Naming	Predictable	Numerals	Unbalanced	Unbalanced L1 vs. L2
Jackson et al., 2004	1	Asymmetric	Parity	Predictable	Number words	Unbalanced	Unbalanced L1 vs. L2
Macizo et al., 2012	1	Asymmetric	Naming	Random	Words	Unbalanced	Unbalanced L1 vs. L2
Macizo et al., 2012	2	Symmetric	Categorization	Random	Words	Unbalanced	Unbalanced L1 vs. L2
Macizo et al., 2012	3	Symmetric	Categorization	Random	Words	Unbalanced	Unbalanced L1 vs. L2
Martin et al., 2013	1	Asymmetric	Naming	Random	Pictures	Highly proficient	Balanced L1 vs. L2
Martin et al., 2013	1	Symmetric	Naming	Random	Pictures	Highly proficient	Unbalanced L1 vs. L3
Martin et al., 2013	1	Asymmetric	Naming	Random	Pictures	Highly proficient	Unbalanced L1 vs. L3
Meuter and Allport, 1999	1	Asymmetric	Naming	Random	Numerals	Unbalanced	Unbalanced L1 vs. L2
Orfanidou and Sumner, 2005	1	Symmetric	LD ★	Predictable	Words	Unbalanced	Unbalanced L1 vs. L2
Orfanidou and Sumner, 2005	2	Symmetric	LD ★	Predictable	Words	Unbalanced	Unbalanced L1 vs. L2
Philipp et al., 2007	1	Asymmetric	Naming	Random	Numerals	Unbalanced	Unbalanced L1 vs. L2
Philipp et al., 2007	1	Asymmetric	Naming	Random	Numerals	Unbalanced	Unbalanced L1 vs. L2
Thomas and Allport, 2000	1	Symmetric	LD ☆	Random	Words	Unbalanced	Unbalanced L1 vs. L2
Thomas and Allport, 2000	2	Symmetric	LD ★	Predictable	Words	Unbalanced	Unbalanced L1 vs. L2
Thomas and Allport, 2000	3	Symmetric	LD ★	Predictable	Words	Unbalanced	Unbalanced L1 vs. L2
Verhoef et al., 2009	1 ◇	Symmetric	Naming	Random	Pictures	Unbalanced	Unbalanced L1 vs. L2
Verhoef et al., 2009	1 ♦	Asymmetric	Naming	Random	Pictures	Unbalanced	Unbalanced L1 vs. L2
Von Studnitz and Green, 1997	1	Symmetric	LD ★	Predictable	Words	Unbalanced	Unbalanced L1 vs. L2
Von Studnitz and Green, 1997	2	Symmetric	LD ☆	Predictable	Words	Unbalanced	Unbalanced L1 vs. L2
Von Studnitz and Green, 2002	1	Symmetric	Categorization	Predictable	Words	Unbalanced	Unbalanced L1 vs. L2

★, *Language exclusive instructions*; ☆, *language inclusive instructions*; ♦, *short cue duration*; ◇, *long cue duration*; LD, *Lexical Decision*.

specific information). It is therefore possible that univalent words containing language specific information might eliminate the switch cost asymmetry because they trigger the appropriate language more directly than pictures and numerals, and therefore do not require the same level of non-target language inhibition (Grainger et al., 2010). Indeed, Peeters et al. (2014) argued that words automatically activate the corresponding language. Consistent with this interpretation, studies that have used univalent stimuli tend to show symmetric switch costs (but see Jackson et al., 2004; Macizo et al., 2012).

Experiments 1A and 1B assess whether the switch cost asymmetry arises when bivalent stimuli (i.e., numerals) are

used and is absent when univalent stimuli (i.e., written words that appear in only one language) are used. If the switch cost asymmetry were eliminated for univalent stimuli, then this would suggest that switch costs in language comprehension and language production tasks arise from the same processes.

Examining how the switch cost asymmetry is affected by stimulus valence also has implications for accounts of language switching in speech production. For instance, Finkbeiner et al.'s (2006) response selection account of the switch cost asymmetry predicts that the switch cost asymmetry should be eliminated when univalent stimuli are used. According to this account, when stimuli are bivalent (associated with both languages), both

language systems generate a viable response. On switch trials, the response criteria change in order to select the response generated by the appropriate language, yielding a switch cost. Asymmetric switch costs arise because on a subset of switch trials the easier L1 response becomes available before the response selection criteria have been updated and is therefore rejected. This creates an additional cost on L1 switch trials because the response must be regenerated. Unlike bivalent stimuli, univalent items only activate a response in one language; consequently, when univalent stimuli are used a single response selection criterion can be used eliminating the need to switch response criteria. Data in support of this account are mixed. Consistent with the response selection account, Finkbeiner et al. (2006) reported that the switch cost asymmetry is eliminated when univalent stimuli are used. However, Peeters et al. (2014) reported a switch cost asymmetry for univalent items. Unfortunately, in both cases, language switching was confounded with task switching, which complicates interpretation of their data. For instance in Finkbeiner et al. (2006), numerals were named in both languages whereas pictures (Experiment 1) and dots (Experiment 2) were used as univalent stimuli (only named in one language). Critically, there was no cost to switching languages for the univalent items, consistent with the response selection account. However, naming pictures and numerals is unlikely to require the same cognitive processes (see Abutalebi and Green, 2007). Indeed, evidence suggests that language switching differs when numerals and pictures are used as stimuli (Declerck et al., 2012). It is therefore possible that language switch costs were not observed for univalent items because, irrespective of whether a language switch was taking place, switching from numerals to pictures (as in their Experiment 1) constituted a task switch. A similar situation occurred in Peeters et al. (2014), where subjects switched from making binary decisions about words (e.g., lexical decision) to naming pictures. Again, a task switch corresponded with switching to the univalent stimuli, rendering interpretation of their data difficult. One finding that seems to be inconsistent with the response selection account has been reported by Macizo et al. (2012). In this study, different sets of words (univalent stimuli) were named for each language. Inconsistent with Finkbeiner et al.'s (2006) response selection account, univalent stimuli produced an asymmetric switch cost.

In the present experiments, we explored these issues by having unbalanced bilinguals alternate between naming (bivalent) numerals in L1 and L2 in one block of trials and (univalent) number words in L1 and L2 in another block of trials. Language switches were random in order to match other studies that have observed a switch cost asymmetry in speech production (e.g., Meuter and Allport, 1999; Costa and Santesteban, 2004; Costa et al., 2006; Macizo et al., 2012). Stimulus valence was blocked because according to Finkbeiner et al. a univalent item should only yield a response in one language permitting a single response criterion to be used whereas under mixed list conditions different response criteria may be used for univalent and bivalent stimuli. Therefore, Finkbeiner et al.'s (2006) response selection account predicts that neither switch costs nor a switch cost asymmetry should be observed for univalent stimuli under blocked list conditions.

Method

Participants

Forty undergraduate students (32 female, 8 male) participated in the experiment². The students received course credit in an eligible psychology course as compensation. All participants had normal or corrected to normal vision. This study was carried out in accordance with the recommendations of Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, Natural Sciences and Engineering Council of Canada, with written informed consent from all subjects in accordance with the Declaration of Helsinki.

Twenty participants who were students at the University of Padova, Italy, participated in Experiment 1A. They reported Italian as their stronger first language (L1) and English as their weaker second language (L2). They began studying English at a variable age (between 5 and 13 years old; mean = 6.2). All 20 participants studied English as a second language during the 5 years of the High School. Twelve of the participants had traveled to an English speaking country in the last 5 years. They self-evaluated their proficiency in English as 3.51 on a Likert 7-point scale (1 very low—7 excellent).

Twenty participants who were students at Trent University, Canada, participated in Experiment 1B. They reported English as their stronger first language (L1) and French as their weaker second language (L2). All participants had studied French in school for a minimum of 6 years (mean = 8.45) as per Ontario education curriculum (Ministry of Education and Training, 1998, 1999). Seventy percent of the participants reported being able to produce and comprehend simple instructions and written material in L2 (Treasury Board of Canada Secretariat, 2012). None described themselves as perfectly matched in English and French.

Stimuli

Stimuli were either numerals (bivalent) or number words (univalent) ranging from 1 to 9. In Experiment 1B, the cognate *six*, which is the same word in both French and English, was replaced with 10 (see also Jackson et al., 2004). All stimuli were presented in a white 16-pt. Times New Roman font against a black background. The words were presented in lower case letters. Numerals subtended 0.6° by 0.6° visual angle. Number words subtended 0.6° visual angle vertically and from 1.7° to 3.4° horizontally.

The language cue consisted of a box subtending 6.9° × 6.9° degrees visual angle that surrounded the location of the target. The interior of the box matched the background color (black). The border of the box was 3 pixels thick (approximately 0.1° visual angle) and was always visible. The color of the box was used to indicate the appropriate response language on a given trial. Standard EPrime colors were used (Experiment 1A: green

²Two types of bilinguals were used in the present studies: English-French bilinguals and Italian-English bilinguals. The purpose of using two different groups of bilinguals was to assess whether the outcome of the experiments generalized across language combinations and whether having English (a language with highly inconsistent spelling-to-sound correspondences) as L1 or L2 affected performance during language switching. As will be seen below, there were no systematic differences between the two types of bilinguals.

for Italian, blue for English; Experiment 1B: red for English, blue for French).

Apparatus

The experiment was conducted using a computer running MicroSoft Windows XP operating system. Stimulus presentation and data collection were controlled using EPrime 2.0 software. Vocal responses were collected using a PST Response Box with a voice-key assembly.

Procedure

Participants were tested individually in a sound attenuated, dimly lit room. They were seated approximately 50 cm from the computer monitor with the microphone placed directly in front of them. Written instructions were presented on the computer screen in the participants' native language (L1). Participants were required to name each stimulus as quickly and accurately as possible in the appropriate language. Depending on the block, participants were informed that numerals or number words would appear on the computer screen. Order of block presentation (numerals vs. words) was counterbalanced across subjects where the assignment of subject to counterbalance was determined pseudorandomly based on the order in which they were tested.

Within each block, subjects were presented with 9 set lists of pseudorandomized trials. Each list consisted of 46 trials with predetermined slots for switch and non-switch trials. *Switch* trials were preceded by a stimulus to be named in the other language and *non-switch* trials were preceded by a same-language stimulus. In order to match the conditions under which the switch cost asymmetry was initially observed (Meuter and Allport, 1999; Costa and Santesteban, 2004; Costa et al., 2006), the lists were constructed so that the probability of a switch was 0.3 and a non-switch was 0.7. The run length ranged from 1 to 7. The assignment of language to the first trial of a list was counterbalanced across subjects. The first trial in each list was then coded as a null switch trial and excluded from subsequent analyses. The assignment of a stimulus on a given trial was determined randomly without replacement. The same stimulus

was permitted to occur on consecutive trials. The number of switches per list ranged from 11 to 20. The order of lists was randomized for each subject, with the first list treated as practice trials. Overall, there were 256 switch and 464 non-switch experimental trials per subject.

Each trial began with the presentation of the box cue in a neutral silver color. After 250 ms, the target stimulus was presented inside the box cue simultaneously with the color of the box cue changing to indicate the response language. The target stimulus and box cue remained visible until a vocal response was made. As soon as a vocal response was initiated, the target stimulus disappeared and the box color changed to white. This screen remained while the experimenter coded the vocal response via button press as correct, incorrect (incorrect language vs. within-language error), or a voice-key failure. Once the response was coded, a new stimulus appeared. Participants were given the opportunity to take a break every 46 trials. The experiment took approximately 40 min to complete.

Results

The mean correct response time (RT) and percentage error data were analyzed separately using analysis of variance (ANOVA) with Language (L1 vs. L2), Stimulus Type (Numerals vs. Words), and Trial Type (Switch vs. Non-switch) as within subjects factors and Experiment (Italian-English: 1A vs. English-French: 1B) as a between subjects factor. Mean response times and percentage error are displayed in **Table 2**.

RT

Prior to analyzing the RT data, trials with incorrect responses (3.05%) or voice-key errors (0.59%) were first removed. RTs to correct responses were subjected to a recursive trimming procedure in which the criterion cutoff for outlier removal was established independently for each condition for each subject by reference to the sample size in that cell (Van Selst and Jolicoeur, 1994). This resulted in the removal of an additional 1.57% of the data.

As can be seen in **Table 2**, there was a main effect of stimulus type where participants took 165 ms longer to respond

TABLE 2 | Mean RT (ms) and percentage errors from Experiments 1A and 1B (random switches) as a function of stimulus type (numerals vs. words), trial type (switch vs. non-switch) and language (L1 vs. L2).

	Response time				Percentage error			
	Numerals		Words		Numerals		Words	
	L1	L2	L1	L2	L1	L2	L1	L2
ITALIAN/ENGLISH BILINGUALS								
Switch	674	646	500	499	5.3	5.0	2.1	2.3
Non-switch	592	608	468	488	3.6	4.3	1.0	1.6
Cost	82	38	32	11	1.7	0.6	1.1	0.7
ENGLISH/FRENCH BILINGUALS								
Switch	731	753	530	532	7.8	7.1	1.3	1.3
Non-switch	659	710	516	526	3.7	4.3	1.0	1.6
Cost	72	43	14	6	4.2	2.8	0.2	-0.4

to numerals compared to number words, $F_{(1, 38)} = 420.1$, $p < 0.001$, $MSE = 5127$, $\eta_p^2 = 0.92$. Overall, the effect of stimulus type is comparable with other studies that have manipulated stimulus valence by changing the orthographic characteristics of words in lexical decision (e.g., 160 ms in Thomas and Allport, 2000 and 180 ms in Orfanidou and Sumner, 2005). Consistent with our participants being unbalanced bilinguals, there was a main effect of language where responses in L1 were faster than responses in L2 [$F_{(1, 38)} = 6.58$, $p < 0.05$, $MSE = 1759$, $\eta_p^2 = 0.15$]. Consistent with participants adjusting processing in response to changes in language, there was a main effect of trial type where switch trials were 37 ms slower than non-switch trials, $F_{(1, 38)} = 126.2$, $p < 0.001$, $MSE = 885$, $\eta_p^2 = 0.77$. There was an interaction between trial type and stimulus type where the size of the switch cost was larger for numerals (77 ms) than for number words (16 ms), $F_{(1, 38)} = 82.0$, $p < 0.001$, $MSE = 449$, $\eta_p^2 = 0.68$. This replicates Orfanidou and Sumner (2005) finding in lexical decision and extends it to speech production.

As expected, there was an interaction between trial type and language where the cost of switching languages was larger for L1 (50 ms) than for L2 (24 ms), replicating the switch cost asymmetry obtained by Meuter and Allport (1999; see also Costa and Santesteban, 2004; Costa et al., 2006; Verhoeve et al., 2009; Macizo et al., 2012), $F_{(1, 38)} = 33.0$, $p < 0.001$, $MSE = 410$, $\eta_p^2 = 0.46$. Also, the three way interaction between stimulus type, trial type and language was significant, $F_{(1, 38)} = 8.13$, $p < 0.01$, $MSE = 298$, $\eta_p^2 = 0.18$. This suggests that the switch cost asymmetry is affected by stimulus valence, since the asymmetry was smaller for the number words (15 ms) compared to the numerals (37 ms).

Since the main goal of the present experiment was to assess whether the switch cost asymmetry would be eliminated by the use of univalent stimuli, separate ANOVAs were performed for numerals and number words with Language (L1 vs. L2) and Trial Type (switch vs. non-switch) as factors. Both analyses showed significant interactions between the two factors [Numerals: $F_{(1, 38)} = 24.4$, $p < 0.001$, $MSE = 560.67$, $\eta_p^2 = 0.39$; Words: $F_{(1, 38)} = 15.23$, $p < 0.001$, $MSE = 147.28$, $\eta_p^2 = 0.29$], suggesting that for both numerals and number words, the switch cost was larger in L1 than in L2. The small switch cost observed for the univalent stimuli in L2 was also reliable [$F_{(1, 38)} = 9.54$, $p < 0.01$, $MSE = 151$, $\eta_p^2 = 0.20$] and was unaffected by Experiment ($F < 1$).

There was a main effect of Experiment, where responses in Experiment 1A were 60 ms faster than the responses in Experiment 1B, $F_{(1, 38)} = 6.56$, $p < 0.05$, $MSE = 44271$, $\eta_p^2 = 0.18$. Experiment interacted with stimulus type, where the effect of stimulus type was smaller for the Italian/English bilinguals in Experiment 1A (141 ms) compared to the English/French bilinguals in Experiment 1B (187 ms), $F_{(1, 38)} = 8.25$, $p < 0.01$, $MSE = 5127$, $\eta_p^2 = 0.18$. Experiment also interacted with language where the effect of language was primarily due to Experiment 1B [Italian-English: 558 vs. 560 ms; English-French: 609 vs. 630 ms; $F_{(1, 38)} = 4.12$, $p < 0.05$, $MSE = 1759$, $\eta_p^2 = 0.14$]. Finally, there was a three way interaction between experiment, language and stimulus type, $F_{(1, 38)} = 12.8$, $p < 0.05$, $MSE = 824$, $\eta_p^2 = 0.25$. The L1 advantage for

English-French bilinguals was larger for numerals (37 ms) than for words (6 ms), $F_{(1, 19)} = 7.142$, $p < 0.05$, $MSE = 1283$, $\eta_p^2 = 0.273$. The advantage for words was not reliable ($F < 1$). In contrast, the L1 advantage for Italian-English bilinguals was smaller for numerals (−2 ms) than for words (10 ms), $F_{(1, 19)} = 6.727$, $p < 0.05$, $MSE = 365$, $\eta_p^2 = 0.261$. The advantage for numerals was not reliable ($F < 1$).

Percent Error

There was nothing in the error data that compromised the interpretation of the RT data. There was a main effect of Trial Type [$F_{(1, 38)} = 34$, $p < 0.001$, $MSE = 0.001$; $\eta_p^2 = 0.47$], reflecting a higher error rate for switch (4%) than for non-switch (2.4%) trials and a main effect of Stimulus Type [$F_{(1, 38)} = 73.85$, $p < 0.001$, $MSE = 0.001$, $\eta_p^2 = 0.66$], reflecting more errors for numerals than for words (4.9 vs. 1.5%). There was an interaction between Stimulus Type and Experiment, where the difference between numerals and words was larger in Experiment 1A (5.7 vs. 1.3%) than in Experiment 1B [4.2 vs. 1.7%, $F_{(1, 38)} = 5.47$, $p = 0.025$, $MSE = 0.001$, $\eta_p^2 = 0.13$]. There was an interaction between Stimulus Type and Trial Type where the switch cost was larger for numerals (switch: 6.3%, non-switch: 3.6%) than for words [switch: 1.7%, non-switch: 1.3%; $F_{(1, 38)} = 11.2$, $p = 0.002$, $MSE = 0.001$, $\eta_p^2 = 0.23$]. There was also an interaction between Trial Type and Experiment where the switch cost was larger in Experiment 1A than in Experiment 1B [$F_{(1, 38)} = 4.8$, $p = 0.034$, $MSE = 0.001$, $\eta_p^2 = 0.11$].

Discussion

Given that the presence/absence of the switch cost asymmetry reported in previous studies largely co-varied with stimulus valence, Experiments 1A and 1B assessed whether the switch cost asymmetry in speech production is present for (bivalent) numerals but absent for (univalent) number words. The observation of a switch cost asymmetry in speech production, despite the use of univalent stimuli (i.e., number words) is inconsistent with bivalent stimuli (e.g., numerals and pictures) being required to observe the switch cost asymmetry. The absence of a switch cost asymmetry in lexical decision studies therefore cannot be due to the use of written words.

The observation that the switch cost asymmetry was not eliminated for univalent stimuli is also inconsistent with Finkbeiner et al.'s (2006) response selection account of the switch cost asymmetry. According to this account, univalent stimuli should not yield a switch cost when stimulus valence is blocked because a single response selection criterion can be used for all of the univalent stimuli. Given that each univalent word was tied to a response in only one language, there was no need to change the response criteria across language switch trials. The response selection account of switch costs therefore not only predicts the absence of a switch cost asymmetry, but also the absence of a switch cost. Neither of these predicted outcomes were observed.

EXPERIMENTS 2A AND 2B

As can be seen in **Table 1**, a second factor that co-varies with the presence/absence of the switch cost asymmetry is switch

predictability. Studies that report a switch cost asymmetry have typically used random switches between categories (similar to Meuter and Allport, 1999). In contrast, studies that do not report a switch cost asymmetry (e.g., Von Studnitz and Green, 1997; Thomas and Allport, 2000) have tended to use predictable switches between languages, based on a variation of Rogers and Monsell's (1995) alternating runs paradigm in which switches occur in a predictable AABB pattern. It is therefore possible that random switches between languages are required (necessary, but not sufficient) in order to observe a switch cost asymmetry. For instance, the switch cost asymmetry may not be observed when switches are predictable because advanced knowledge of the language switch provides the opportunity to endogenously prepare for the language switch prior to the presentation of the stimulus³. Otherwise, it might be that symmetric switch costs depend on the conjunction of predictable switches and univalent stimuli. The aim of the present experiment was to test these hypotheses. Similar to Experiments 1A and 1B, participants in Experiment 2A were Italian / English bilinguals, and participants in Experiment 2B were English/French bilinguals. Again, the task was naming (bivalent) numerals and (univalent) number words, with valence assigned to different blocks of trials as in the previous experiment. Following Rogers and Monsell's (1995) alternating runs paradigm, the assignment of language for responding followed an AA/BB pattern (a run length of 2).

Method

Participants

Thirty-six undergraduate students (32 female, 4 male) participated in the experiment. They all received credit in an eligible psychology course as compensation. All participants had normal or corrected to normal vision.

Sixteen participants were students from the University of Padova and participated in Experiment 2A. They were unbalanced Italian-English bilinguals with Italian as stronger first language (L1) and English as their weaker second language (L2). They met the same criteria as subjects in Experiment 1A.

Twenty participants were undergraduate students at Trent University and participated in Experiment 2B. They were unbalanced English-French bilinguals with English as their stronger first language (L1) and French as their weaker second language (L2). They met the same criteria as participants Experiment 1B.

Stimuli

Target stimuli were identical to Experiments 1A and 1B. Following Rogers and Monsell (1995), a 2×2 grid was used

to help subjects keep track of the predictable AABB pattern⁴. Each square in the grid subtended 6.9° by 6.9° visual angle as in Experiments 1A and 1B.

Apparatus

The apparatus was identical to Experiments 1A and 1B.

Procedure

The procedure was similar to Experiments 1A and 1B. Once again, the experiment consisted of two blocks, with numerals being presented in one block and number words presented in the other. The assignment of Stimulus Type to block was counterbalanced pseudorandomly across subjects based on the order in which they were tested. Within each block, subjects were presented with 9 lists of 44 trials with a predictable switch after two consecutive trials in the same language (i.e., a run length of 2). As in Experiments 1A and 1B, the assignment of stimulus to trial was determined randomly without replacement. The first list in each block was treated as practice.

The 2×2 display grid was visible throughout the presentation of a list of trials. A trial started with an empty display grid. After 250 ms, the target stimulus was presented in the center of one of the four boxes. The stimulus remained visible until a vocal response was made. The accuracy of the vocal response was then coded via button press by the researcher before the beginning of the next trial. The location of a target on successive trials moved to the adjacent clockwise location in the grid. Adjacent horizontal locations always corresponded to a single language and the assignment of language to position (top vs. bottom) was determined randomly for each subject. Participants were given the opportunity to take a break every 44 trials. The experiment took approximately 40 min to complete.

Results

The data were analyzed in the same way as Experiment 1. Mean response latencies and accuracies are displayed in **Table 3**.

RT

Prior to analyzing the RT data, trials with incorrect responses (2.9%) or voice-key errors (0.75%) were removed. RTs to correct responses were subjected to the same recursive trimming procedure used in Experiment 1 (Van Selst and Jolicoeur, 1994). This resulted in the removal of an additional 2.03% of the data.

As can be seen in **Table 3**, there was a main effect of stimulus type where participants took 142 ms longer to name numerals than number words, $F_{(1, 34)} = 128.1, p < 0.001, MSE = 11,149, \eta_p^2 = 0.79$. The effect of stimulus valence is comparable in magnitude to Experiment 1 [$F_{(1, 74)} = 1.61, p = 0.209, MSE = 8994, \eta_p^2 = 0.02$]⁵ and with previous studies that have used the lexical decision task (Von Studnitz and Green, 1997; Thomas and

³There may be additional relevant differences other than predictability. For instance, Altmann (2007) has noted that switch costs are generally larger in the alternating runs paradigm because the switch costs in this paradigm include costs associated with switching and costs associated with decoding the cue. Further, the use of long response-stimulus intervals (RSI) may eliminate endogenous contributions to the switch cost (c.f. Rogers and Monsell, 1995). Assessing whether the switch cost asymmetry is eliminated in the alternating runs paradigm as a whole therefore tests whether the absence of an asymmetry is due to any number of known and unknown differences between these two methods.

⁴The use of an AABB trial pattern increases the probability of a switch from 0.3 in Experiment 1 to 0.5 in Experiment 2. The observation of a switch cost asymmetry for the univalent stimuli in the present experiment suggests that the probability of a switch is not a determining factor for when a switch cost asymmetry is observed.

⁵All comparisons across studies (e.g., Experiment 1 vs. Experiment 2) were conducted using a mixed model ANOVA with Language (L1 vs. L2), Stimulus Type (numerals vs. words) and Trial Type (switch vs. non-switch) as repeated factors and Study (e.g., Experiment 1 vs. Experiment 2) as a between subjects factor.

TABLE 3 | Mean RT (ms) and percentage errors from Experiments 2A and 2B (predictable switches) as a function of stimulus type (numerals vs. words), trial type (switch vs. non-switch) and language (L1 vs. L2).

	Response time				Percentage error			
	Numerals		Words		Numerals		Words	
	L1	L2	L1	L2	L1	L2	L1	L2
ITALIAN/ENGLISH BILINGUALS								
Switch	660	661	512	525	6.9	7.1	2.1	2.0
Non-switch	549	600	472	499	4.5	5.0	1.2	0.7
Cost	111	61	40	26	2.4	2.1	0.9	1.3
ENGLISH/FRENCH BILINGUALS								
Switch	693	727	507	518	7.2	7.8	2.0	0.9
Non-switch	594	666	481	500	3.1	4.1	0.5	0.5
Cost	99	61	26	18	4.1	3.7	1.5	0.4

Allport, 2000; Orfanidou and Sumner, 2005). Consistent with our participants being unbalanced bilinguals, there was a main effect of language where L1 responses were 28 ms faster than L2 responses, $F_{(1, 34)} = 8.02$, $p < 0.05$, $MSE = 2684$, $\eta_p^2 = 0.32$.

Once again, there was a main effect of trial type where participants took longer to respond on switch trials compared to non-switch trials, yielding a 55 ms switch cost, $F_{(1, 34)} = 212.2$, $p < 0.001$, $MSE = 1023$, $\eta_p^2 = 0.86$. Consistent with stimulus valence influencing language selection, there was an interaction between stimulus type and trial type where the switch cost was 55 ms larger for numerals (83 ms) than for the number words (28 ms), $F_{(1, 34)} = 103.5$, $p < 0.001$, $MSE = 525$, $\eta_p^2 = 0.75$. This replicates the pattern observed in Experiment 1 and previously reported by Orfanidou and Sumner (2005) in lexical decision. Also consistent with stimulus valence influencing language selection, there was an interaction between language and stimulus type where L1 responses were 40 ms faster than L2 responses for numerals, but only 17 ms faster than L2 responses for number words, $F_{(1, 34)} = 8.96$, $p < 0.05$, $MSE = 1003$, $\eta_p^2 = 0.21$. The interaction between trial type and language was significant, where the switch costs were asymmetric, with a 28 ms larger switch cost in L1 (69 ms) than in L2 (41 ms), $F_{(1, 34)} = 26.0$, $p < 0.001$, $MSE = 536$, $\eta_p^2 = 0.43$. Critically, there was an interaction between trial type, language and stimulus type where the switch cost asymmetry was larger for numerals than for number words, $F_{(1, 34)} = 12.9$, $p < 0.001$, $MSE = 389$, $\eta_p^2 = 0.28$. This replicates the pattern observed in Experiments 1A and 1B.

Since the main goal of the Experiment 2 was to assess whether the switch cost asymmetry would be eliminated by the use of predictable switches, additional repeated-measure ANOVAs were performed separately for the numeral and number word conditions with Language (L1 vs. L2) and Trial Type (switch vs. non-switch) as factors. Inconsistent with predictable switches between languages eliminating the switch cost asymmetry, a switch cost asymmetry was observed for numerals, $F_{(1, 34)} = 26.6$, $p < 0.001$, $MSE = 670$, $\eta_p^2 = 0.44$, where the switch cost was 44 ms larger in L1 (105 ms) compared to L2 (61 ms), and this effect was not qualified by experiment ($F < 1$).

Inconsistent with the conjunction of predictable switches and the univalent stimuli being necessary to eliminate the switch cost

asymmetry, there was a reliable switch cost asymmetry for the (univalent) number words, $F_{(1, 34)} = 4.4$, $p < 0.05$, $MSE = 254$, $\eta_p^2 = 0.11$. The switch cost was 11 ms larger in L1 (33 ms) compared to L2 (22 ms), and it was not qualified by experiment ($F < 1$).

No main effect of Experiment was obtained, $F_{(1, 34)} = 1.21$, $p = 0.279$, $MSE = 40,069$, $\eta_p^2 = 0.03$. However, there was an interaction between experiment and stimulus type where the effect of stimulus type was smaller for the Italian/English bilinguals (115 ms) compared to the English/French bilinguals (168 ms), $F_{(1, 34)} = 4.43$, $p < 0.05$, $MSE = 11,149$, $\eta_p^2 = 0.11$. Also, similar to Experiments 1A and 1B, there was an interaction between language, stimulus type and experiment where the language by stimulus type interaction was more pronounced in the English/French bilinguals than in the Italian/English bilinguals, $F_{(1, 34)} = 4.47$, $p < 0.05$, $MSE = 1003$, $\eta_p^2 = 0.12$.

Percent Error

There were no effects in the error data that compromised the interpretation of the RT data. There was a main effect of Stimulus Type where more errors were made to numerals than to number words, $F_{(1, 34)} = 88.4$, $p < 0.001$, $MSE = 0.001$, $\eta_p^2 = 0.72$. There was a main effect of Trial Type where more errors were made on switch compared to non-switch trials, $F_{(1, 34)} = 51.4$, $p = 0.001$, $MSE < 0.001$, $\eta_p^2 = 0.60$. As in Experiment 1, there was an interaction between trial type and Experiment, $F_{(1, 34)} = 8.12$, $p = 0.007$, $MSE < 0.001$, $\eta_p^2 = 0.19$. However, here the switch cost was larger in Experiment 2B (English-French: 2.5%) than in Experiment 2A (Italian-English: 1%). There was also an interaction between Stimulus Type and Trial Type where the switch cost was larger for numerals (switch: 6.2%; non-switch: 3.4%) than it was for number words (switch: 1.3%; non-switch: 0.6%), $F_{(1, 34)} = 14.2$, $p = 0.001$, $MSE = 0.001$, $\eta_p^2 = 0.30$. No other effects were significant ($F_s < 1.4$).

Discussion

The goal of Experiments 2A and 2B was to assess whether the absence of a switch cost asymmetry reported in previous work was due to the use of predictable switches between languages. Inconsistent with unpredictable switches being required for asymmetric switch costs, asymmetric switch costs were observed

for numerals despite predictable switches between languages. Furthermore, the switch cost asymmetry was once again observed for number words, which uniquely indicated the language to be used for the response. This suggests that the absence of the switch cost asymmetry in lexical decision studies was not due to the conjunction of predictable switches and the use of univalent stimuli. Further, this second demonstration that the switch cost asymmetry is not eliminated for univalent stimuli provides additional evidence inconsistent with Finkbeiner et al.'s (2006) response selection account of language switching, which predicts that switch costs should not be observed when the response selection criteria do not need to change.

EXPERIMENT 3

The outcomes of Experiments 1 and 2 indicate that the switch cost asymmetry is not due to methodological differences such as the type of stimuli or the predictability of switches. The data also indicate that the switch cost asymmetry in speech production is not a consequence of changing response criteria for bivalent stimuli as hypothesized by Finkbeiner et al. (2006). Instead the outcomes of Experiments 1 and 2 suggest that the switch cost asymmetry is observed in speech production because of how the languages are activated/inhibited. Converging evidence for this claim comes from two additional sources. First, stimulus valence interacted with language, which suggests that the two factors affect a common process, most likely one involved in language selection. Second, stimulus valence affected the magnitude of the switch cost asymmetry, whereby the switch cost asymmetry was smaller for the univalent stimuli than for the bivalent stimuli. This is consistent with language specific information contained in the stimulus reducing the impact of language strength on the selection process.

If the switch cost asymmetry arises from how languages are activated/inhibited then this suggests that languages are activated differently during comprehension and speech production tasks. In order to test this possibility, Experiment 3 examines how switch costs are affected by stimulus valence when the univalent and bivalent stimuli are randomly intermixed in the naming task. Previous research suggests that in the lexical decision task, switch costs are only reduced for univalent items when univalent and bivalent stimuli are presented in separate blocks, as in Experiments 1 and 2 (Orfanidou and Sumner, 2005). When univalent and bivalent items are randomly intermixed in a single block of trials, switch costs are unaffected by stimulus valence in lexical decision (Von Studnitz and Green, 1997; Thomas and Allport, 2000; Orfanidou and Sumner, 2005). If the switch cost asymmetry arises in speech production but not in language comprehension because there are fundamental differences in how languages are activated in language comprehension and speech production tasks, then this raises the possibility that stimulus valence will continue to affect switch costs in speech production when univalent and bivalent stimuli are randomly intermixed. In contrast, if stimulus valence does not affect the magnitude of the switch costs in mixed lists during speech production, then this will mirror the pattern previously obtained in lexical decision (Von Studnitz and Green, 1997; Thomas and Allport,

2000; Orfanidou and Sumner, 2005) and will be inconsistent with the claim that languages are activated differently in language comprehension and speech production tasks.

How stimulus valence affects switch costs when univalent and bivalent stimuli are randomly intermixed also has implications for theories of the switch cost asymmetry in speech production. The outcomes of Experiments 1 and 2 are consistent with accounts that attribute the switch cost asymmetry to competition between languages (Grainger and Beauvillain, 1987; Grainger and Dijkstra, 1992; Green, 1998; Meuter and Allport, 1999; Thomas and Allport, 2000; Orfanidou and Sumner, 2005; Peeters et al., 2014). One type of account attributes the switch cost asymmetry to competition between language schemas (e.g., Von Studnitz and Green, 1997; Green, 1998; Meuter and Allport, 1999; Thomas and Allport, 2000; Orfanidou and Sumner, 2005). According to these accounts, each language is associated with a language task schema. Successful speech production in one language requires the activation of the response language and the inhibition of the non-response language, which persists involuntarily when switching languages. Performance on a switch trial is slowed by having to overcome the inhibition required to respond in the appropriate language on the previous trial, yielding a switch cost. The cost of overcoming the prior inhibition of the currently relevant language is a function of the relative strength of the two languages. If the languages are unbalanced in strength, then naming an item in L2 requires strong(er) inhibition of L1, therefore switching to L1 will yield a large(er) cost because the time to overcome the strong inhibition of L1 from the previous language schema is longer. In contrast, naming an item in L1 only requires weak(er) inhibition of the L2 language, therefore the time to overcome L2 inhibition is shorter when switching to L2⁶.

A second type of account attributes the switch cost asymmetry to competition within the lexicon (Grainger and Beauvillain, 1987; Dijkstra and Van Heuven, 1998; Van Heuven et al., 1998; Peeters et al., 2014). According to these accounts, lexical representations are inhibited without the need for language task schemas and switch costs can be explained by mechanisms entirely within the language system. Here, greater inhibition of representations in L1 is required, in order to name an item in L2, yielding larger switch costs when switching back to L1 (Grainger and Dijkstra, 1992; Van Heuven et al., 1998; Grainger et al., 2010; Peeters et al., 2014).

Researchers investigating the locus of switch costs have repeatedly argued that if the cost of switching languages arises within the language system (as suggested by Grainger and colleagues), then switch costs should be reduced for stimuli with language specific orthography because they will differentially activate lexical representations in the two languages thereby reducing competition (Thomas and Allport, 2000; Orfanidou and Sumner, 2005; Peeters et al., 2014). The observation that switch costs are not reduced for univalent stimuli with language specific orthography in lexical decision when the univalent and bivalent

⁶Although inhibition is often invoked to explain asymmetric switch costs (especially in language switching), it may not be required. For instance task set activation can also account for the data (see Koch et al., 2010, for a more elaborate discussion).

stimuli are randomly intermixed has been used to support the claim that the control processes involved in language switching are outside the language system (Von Studnitz and Green, 1997; Thomas and Allport, 2000; Orfanidou and Sumner, 2005). Thus, the observation that stimulus valence affects switch costs, including the switch cost asymmetry when univalent and bivalent stimuli are intermixed (and therefore insensitive to context) in speech production would be inconsistent with the claim that the switch cost asymmetry arises from competition outside the language system, between language task schemas (Grainger and Beauvillain, 1987; Green, 1998; Meuter and Allport, 1999; Orfanidou and Sumner, 2005; Grainger et al., 2010; Peeters et al., 2014).

Method

Participants

Forty-two undergraduate students at Trent University participated in return for credit in an eligible psychology course. These participants met the same criteria as participants in Experiments 1B and 2B. All participants had normal or corrected to normal vision.

Stimuli

The stimuli were the same numerals (bivalent) and number words (univalent) as Experiments 1B and 2B.

Apparatus

The apparatus was the same as Experiments 1B and 2B.

Procedure

The procedure was the same as Experiment 2 (alternating runs), except that the value of a target stimulus and its form (numeral vs. number word) were determined randomly on each trial (mixed blocks).

Results

The present study did not include experiment as a factor. In all other ways the data were analyzed in the same way as Experiments 1 and 2. Mean response latencies and percentage error are displayed in Table 4.

TABLE 4 | Mean RT (ms) and percentage error as a function of stimulus type (numerals vs. words) and language (L1 vs. L2) and trial type (switch vs. non-switch) in Experiment 3.

	Numerals		Words	
	L1	L2	L1	L2
RESPONSE TIME				
Switch	637	644	583	575
Non-switch	553	602	534	544
Cost	84	42	49	31
PERCENTAGE ERROR				
Switch	8.1	10.0	2.5	1.5
Non-switch	3.0	5.6	0.8	0.5
Cost	5.1	4.4	1.7	1.0

RT

Prior to analyzing the RT data, trials with incorrect responses (3.97%) or voice-key errors (2.27%) were removed. RTs for correct responses were subjected to the same recursive trimming procedure used in Experiments 1 and 2 (Van Selst and Jolicoeur, 1994). This resulted in the removal of an additional 2.35% of the data.

There was a main effect of stimulus type where subjects took 50 ms longer to respond to numerals compared to number words, $F_{(1, 41)} = 79.3$, $p < 0.001$, $MSE = 2629$, $\eta_p^2 = 0.66$. The effect of stimulus type was smaller than when stimulus valence was manipulated between blocks in Experiment 1 [$F_{(1, 78)} = 126.76$, $p < 0.001$, $MSE = 4291$, $\eta_p^2 = 0.62$] and Experiment 2 [$F_{(1, 74)} = 50.23$, $p < 0.001$, $MSE = 7108$, $\eta_p^2 = 0.40$]⁵.

Consistent with our participants being unbalanced bilinguals, there was a main effect of language where responses in L1 were 14 ms faster than responses in L2 [$F_{(1, 41)} = 5.90$, $p < 0.05$, $MSE = 2911$, $\eta_p^2 = 0.13$]. The L1 advantage was modulated by Stimulus Type, being larger for numerals (28 ms) than for number words (1 ms) [$F_{(1, 41)} = 18.8$, $p < 0.001$, $MSE = 854$, $\eta_p^2 = 0.32$], as was observed for blocked list presentation.

There was a main effect of trial type where switch trials were 51 ms slower than non-switch trials, $F_{(1, 41)} = 229.6$, $p < 0.001$, $MSE = 965$, $\eta_p^2 = 0.85$. Inconsistent with lexical decision (Von Studnitz and Green, 1997; Thomas and Allport, 2000; Orfanidou and Sumner, 2005), the switch costs were larger for (bivalent) numerals (63 ms) than for (univalent) number words (40 ms), $F_{(1, 41)} = 68.7$, $p < 0.001$, $MSE = 159$, $\eta_p^2 = 0.63$. The magnitude of the effect did not differ from those reported in Experiments 1 ($F < 1$) and 2 ($F < 1$) where stimulus valence was blocked⁵.

As expected, there was an interaction between language and trial type where the cost of switching languages was larger for L1 (66 ms) than for L2 (36 ms), replicating the switch cost asymmetry [e.g., Meuter and Allport, 1999; $F_{(1, 41)} = 49.6$, $p < 0.001$, $MSE = 383$, $\eta_p^2 = 0.55$]. Finally, the three way interaction between Trial Type, Language and Stimulus Type was significant, indicating that the asymmetry was smaller for the number words (18 ms) compared to the numerals (42 ms), $F_{(1, 41)} = 8.04$, $p < 0.01$, $MSE = 356$, $\eta_p^2 = 0.16$.

Percent Error

There was nothing in the error data that compromised the interpretation of the RT data. There was a main effect of Stimulus Type where more errors were made for numerals (6.7%) compared to number words (1.3%), [$F_{(1, 41)} = 97.1$, $p < 0.001$, $MSE = 24.8$; $\eta_p^2 = 0.70$]. There was a main effect of language where more errors were made in L2 than L1 [$F_{(1, 41)} = 4.68$, $p < 0.05$, $MSE = 10.7$; $\eta_p^2 = 0.10$]. There was an interaction between Language and Stimulus Type where the effect of language was larger for numerals than for number words, $F_{(1, 41)} = 22.3$, $p < 0.001$, $MSE = 7.97$; $\eta_p^2 = 0.35$.

There was a main effect of Trial Type where more errors were made on switch (5.5%) than on non-switch (2.5%) trials, [$F_{(1, 41)} = 87.2$, $p < 0.001$, $MSE = 8.80$, $\eta_p^2 = 0.68$]. There was no interaction between Trial Type and Language, $F_{(1, 41)} = 2.28$, $MSE = 4.58$, $\eta_p^2 = 0.05$. However, there was an interaction

between Trial Type and Stimulus Type where the switch cost was larger for numerals (4.7%) than for words (1.3%), $F_{(1, 41)} = 48.2$, $p < 0.001$, $MSE = 4.96$, $\eta_p^2 = 0.54$. No other effects approached significance.

Discussion

Stimulus valence modulated overall switch costs and the switch cost asymmetry in the present experiment, despite the univalent and bivalent stimuli being presented in the same block of trials. The observation that stimulus valence influences switch costs irrespective of context during naming but not lexical decision is consistent with languages being activated differently and suggests that there are fundamental differences in how languages are controlled in comprehension and production tasks.

The present results are also inconsistent with accounts of language switching that place the control mechanisms entirely outside the language system such as the language task schema account of language switching, which predicts that stimulus valence should not affect switch costs under mixed list conditions (Thomas and Allport, 2000; Orfanidou and Sumner, 2005). The outcome of the present experiment is, however, consistent with within-language accounts of the switch cost asymmetry, which predict that the switch cost asymmetry should be affected by stimulus valence irrespective of context (Thomas and Allport, 2000; Orfanidou and Sumner, 2005; Finkbeiner et al., 2006; Peeters et al., 2014).

GENERAL DISCUSSION

The standard account of the switch cost asymmetry in unbalanced bilinguals is that it arises from differences in the relative strength of the two languages. This account predicts that a switch cost asymmetry should arise in any language task when the strength of the two languages is unbalanced. Inconsistent with this account, the switch cost asymmetry is observed in speech production tasks but not in comprehension tasks, such as lexical decision and semantic categorization (see **Table 1**). Experiments 1 and 2 ruled out the possibility that these differences were due to two methodological factors (stimulus valence and switch predictability) that are confounded with the type of task. Experiment 3 demonstrated that languages are activated or operate differently for language comprehension and speech production tasks. Therefore, the present data converge on the claim that there are important task-related differences in how languages are controlled. Indeed, there are now three indicators that switch costs differ in lexical decision and speech production. First, switch costs are symmetric in the lexical decision task (e.g., Von Studnitz and Green, 1997; Thomas and Allport, 2000; Orfanidou and Sumner, 2005) and asymmetric in the speech production task (e.g., Meuter and Allport, 1999; Costa and Santesteban, 2004; Costa et al., 2006; Macizo et al., 2012), even when univalent stimuli (Experiments 1 and 2) and predictable switches are used (Experiment 2) so as to match the conditions usually observed in comprehension tasks. Second, switch costs are not affected by stimulus valence in mixed list contexts in lexical decision (Von Studnitz and Green, 1997; Thomas and Allport, 2000; Orfanidou and Sumner, 2005), but are in speech

production (Experiment 3). Finally, stimulus valence has a larger impact for L1 than for L2 thereby reducing the switch cost asymmetry in speech production, but not in lexical decision (Thomas and Allport, 2000; Orfanidou and Sumner, 2005).

The present experiments are also consistent with at least some of the switch cost asymmetry arising from processing within the language system (Grainger and Beauvillain, 1987; Peeters et al., 2014). A switch cost asymmetry was observed for univalent number words in Experiments 1 and 2 when a single response selection criterion could be used, inconsistent with the switch cost asymmetry arising at the level of response selection as suggested by Finkbeiner et al. (2006; see also Peeters et al., 2014). Inconsistent with the switch cost asymmetry arising from competition between language schemas the switch cost asymmetry was affected by stimulus valence in Experiment 3 where the univalent and bivalent stimuli were randomly intermixed. According to language schema accounts switch costs should not be affected by factors that affect processing within the language system such as language specific orthography (Meuter and Allport, 1999; Thomas and Allport, 2000), especially under mixed list conditions (Orfanidou and Sumner, 2005). Both of these outcomes are predicted by within-language accounts of the switch cost asymmetry in which univalent stimuli either (1) directly activate a language node that specifies the appropriate language or (2) lead to stronger activation of lexical entries in the appropriate language by way of having a direct match in only one lexicon (this is not to say that do not activate lexical entries in the other language, e.g., Kroll et al., 2013).

Why Are Asymmetric Switch Costs Not Observed in Lexical Decision?

The present experiments suggest that there are important differences in how languages are controlled during speech production and language comprehension tasks, and that at least some of these differences arise in the lexicon. Here, we propose that language switch costs arise from (at least) three sources: (1) language task schema activation / competition that affects response selection and initiation (Von Studnitz and Green, 1997; Thomas and Allport, 2000; Orfanidou and Sumner, 2005), (2) early activation of the language task schema (Orfanidou and Sumner, 2005) or language nodes (Grainger et al., 2010; Peeters et al., 2014) using stimulus attributes, and (3) within language activation/competition. Language task schemas specify which language is to be used and the specific configuration of the language system that is used to perform the language task (i.e., lexical decision, semantic categorization, speech production, etc.). According to this account, lexical decision, semantic categorization, and speech production tasks differ in terms of the specific lexical/semantic systems (e.g., orthography, phonology, semantics, syntax, etc.) required for task performance (see Green, 1998; Von Studnitz and Green, 2002). This information is specified as part of the language task schema. For instance, semantic categorization and lexical decision are often conceptualized as comprehension tasks because of a greater reliance on orthographic and semantic systems (e.g., Van Heuven et al., 1998; Peressotti et al., 2003; Yap et al., 2011) whereas naming is conceived as a production

task because of its dependence on retrieving a representation for output (Meuter and Allport, 1999).

Here, we hypothesize that whether switch costs are symmetric or asymmetric in unbalanced bilinguals depends on both the relative strength of the language task schema and the specific levels of the language system specified as part of the language schema (e.g., semantics vs. phonology). Given the dependence of speech production tasks on phonological processing, we hypothesize that phonological processing is particularly susceptible to interference from activated entries in competing languages. Consistent with this possibility, generating a phonological code from print is known to require central attention, whereas orthographic processing does not (Reynolds and Besner, 2006).

Evidence that at least some of the switch cost asymmetry arises from competition within the language system does not preclude that some of the switch cost asymmetry arises outside the language system. Therefore, the observation that asymmetric switch costs persisted for univalent items could either be due to competition between phonological entries, despite the use of univalent number words, or it could be due to additional sources. Here, we consider two additional sources outside the lexicon that could give rise to the residual switch cost asymmetry, namely (1) response execution, and (2) language task schemas.

Unlike lexical decision, where the responses are likely equally novel for both L1 and L2, unbalanced bilinguals have more practice speaking in L1 than in L2. Therefore, it could be the case that there is competition between the processes involved in articulation or setting the parameters of a self-speech monitoring mechanism that checks one's own speech for errors or other problems. Although the present study does not rule out these possibilities, there is some evidence in the literature that suggests that the switch cost asymmetry does not arise from processes involved in response execution. For instance, if a switch cost asymmetry arises from competition between unbalanced response schemas, then a switch cost asymmetry should be observed in speech production whenever the languages are unbalanced in strength. Inconsistent with this prediction, the switch cost asymmetry is often absent for highly proficient bilinguals when switching between unbalanced languages (Costa and Santesteban, 2004; Costa et al., 2006; Martin et al., 2013). Furthermore, a switch cost asymmetry is not observed for unbalanced bilinguals in speech production when the languages switches are voluntary (Gollan and Ferreira, 2009). The switch cost asymmetry is not consistently observed for unbalanced bilinguals when the sequence of language switches is determined by patterns maintained in memory (Declerck et al., 2012). These latter approaches differ from cued studies in terms of how language selection takes place, but not how responses are executed. Taken together with the present findings, these studies suggest that very little, if any, of the switch cost asymmetry arises from response execution processes in unbalanced bilinguals.

A second source outside the lexicon that could give rise to the residual switch cost asymmetry is competition between language task schemas. If the relative strength of the language task schema depends on experience with the configuration of the language system required for task performance and stimulus

response mapping (at least in the case of lexical decision), then for unbalanced bilinguals this should result in more experience with the configuration of the language system required for speech production in L1 than L2. Thus, it is possible that the language task schemas will be unbalanced in speech production for unbalanced bilinguals. However, this should be less pronounced in a task like lexical decision, which, as noted by Thomas and Allport (2000), requires "the introduction of arbitrary, task-specific components to the use of the bilingual's languages" (p. 62). As such, the language task schemas will be balanced in lexical decision, and other tasks where the system configuration is novel, yielding symmetric switch costs.

Although we postulate that at least part of the switch cost asymmetry arises from competition between phonological representations, which may be more sensitive to competition from entries in other languages, another possibility is that sensitivity to competition is tied to a more general difference between the organization of input and output lexicons. Models of language processing often specify separate input and output orthographic and phonological lexicons (e.g., Coltheart et al., 2001; Coltheart, 2004). To the best of our knowledge, task type (comprehension vs. production) has been confounded with the type of internal representation required for task completion (e.g., orthographic vs. phonological). It is therefore unclear whether differences in performance are a consequence of how input and output systems are controlled, or whether there are differences in how orthographic and phonological systems are interconnected in bilingual speakers. For instance, Grainger et al.'s (2010) developmental version of the Bilingual Interactive Activation (BIA) model attributes switch costs to different mechanisms in comprehension (which have used words as stimuli) and production tasks (which have used bivalent stimuli such as numerals and pictures). In their view, univalent words exogenously activate the appropriate language node in comprehension tasks, which selectively enhances processing in one language relative to the other language. In speech production, the use of bivalent stimuli requires top-down control over the language node. According to this account the switch cost asymmetry arises in speech production because endogenous activation of the language node yields greater inhibition of the lexical representations for L1 than for L2. The observation that univalent stimuli reduce switch costs and the switch cost asymmetry can be explained by univalent stimuli exogenously activating the language nodes, thereby reducing the contribution of endogenous control processes that give rise to the asymmetry. One issue that this account has difficulty explaining is the persistence of a switch cost asymmetry in Experiments 2 and 3 where predictable switches between languages occurred. In these experiments, the average response-stimulus interval (RSI) was long (756 ms in Experiment 2A, 497 ms in Experiment 2B, and 706 ms in Experiment 3)⁷. A switch cost asymmetry has also been reported for unbalanced bilinguals using predictable switches and an RSI of 1500 ms by Jackson et al. (2001). This is problematic for the endogenous control account of the switch cost asymmetry

⁷The RSI consisted of how long it took the researcher to code a subject's vocal response and a 250 delay at the beginning of the trial next trial.

because evidence suggests that at RSIs beyond 500 ms, switch costs are primarily driven exogenously by the stimulus itself when switches are predictable (Rogers and Monsell, 1995). Therefore, the persistence of the switch cost asymmetry at long RSIs suggests that the switch cost asymmetry is not due to endogenous control. Converging evidence comes from studies examining the role of advance preparation (endogenous control) under conditions where the language switches are random (as in Experiment 1). In these studies, the role of endogenous processes in the switch cost asymmetry assessed by examining how it is affected by the cue-stimulus interval (CSI). These studies have reported inconsistent effects of CSI on the switch cost asymmetry (e.g., Philipp et al., 2007; Verhoef et al., 2009; Declerck et al., 2012; Fink and Goldrick, 2015). Consequently, there seems to be little evidence to support the claim that the switch cost asymmetry in speech production is driven purely by top-down endogenous control processes.

Why Are Language Comprehension and Speech Production Affected Differently by Stimulus Valence?

Accounts of language switching also need to explain why stimulus valence does not affect switch costs in the lexical decision task when the univalent and bivalent stimuli are randomly intermixed in a single block of trials. Here, we hypothesize that there was no reduction in the switch costs for univalent items in lexical decision because switch costs in this task have been largely due to processes outside the lexicon, such as competition between language task schemas (Von Studnitz and Green, 1997; Thomas and Allport, 2000). To date, the words used in lexical decision have been unique to one language (and therefore arguably univalent), yet stimulus valence was further defined according to the presence or absence of language specific orthographic cues (e.g., combinations of letters). If switch costs arising from within the language system are largely limited to competition between orthographic-lexical representations in lexical decision (as opposed to phonological-lexical representations in speech production), then this raises the possibility that switch costs arising within the language system were already minimized by the language specific nature of the words, and therefore were not reduced further by using words with language specific orthography. Support for this hypothesis comes from evidence that the majority of the switch cost in lexical decision is due to changing the response selection criteria (Von Studnitz and Green, 1997; Thomas and Allport, 2000).

Non-Task Associated Differences

Finally, it is always possible that there are other task-associated differences that become candidates for differences in language switching across tasks. For instance, lexical decision includes the use of non-words, which were not included in the present experiments. There is evidence from research on visual word recognition and reading aloud that the presence of non-words in a context can change how sublexical and lexical information affect one another. For instance, stimulus quality and word frequency yield additive effects in lexical decision (Yap and

Balota, 2007). In contrast, stimulus quality and word frequency interact in reading aloud (O'Malley et al., 2007) unless non-words are added to the list context, in which case their effects are additive (O'Malley and Besner, 2008). This suggests that the presence of non-words could change how lexical information is activated (see also Thomas and Allport, 2000). In this instance, the presence of non-words is unlikely to be the driving factor that determines whether switch costs are symmetric, because switch costs are symmetric in semantic categorization, where non-words are not part of the stimulus set. This is not to say, however, that there are no other differences. At present, however, we believe that there is sufficient evidence to justify further investigation into task related differences in bilingual language switching.

Implications for Highly Proficient Bilinguals

The switch cost asymmetry is not usually observed when highly proficient bilinguals (e.g., those that are balanced in L1 and L2), switch between established languages (L1, L2, or L3), but is observed when they switch between languages of low proficiency (L3, L4, or a new language; Costa and Santesteban, 2004; Costa et al., 2006; Martin et al., 2013). In order to explain this pattern, Costa and colleagues suggested that highly proficient bilinguals have available two mechanisms for selecting a language (1) a language-specific selection mechanism and (2) within-language inhibitory control. The language-specific selection mechanism operates when switching between languages with established lexicons by setting different criteria for lexical selection in each language. This mechanism does not change how the languages operate; instead it operates on the output of the language system. Switching between language-specific selection criteria is independent of language strength and therefore yields symmetric switch costs. Inhibitory control operates when one of the lexicons is not well formed so that a language specific selection criterion cannot be established. In this instance inhibitory mechanisms affect lexical representations in the dominant language (e.g., L1) proportional to language strength yielding asymmetric switch costs. The present findings are consistent with unbalanced bilinguals using an inhibitory mechanism that affects processing within a language system. Interestingly, this dual process account of language switching in highly proficient bilinguals could be tested by examining how switch costs are affected by stimulus valence in mixed list conditions as in Experiment 3. If stimulus valence affects switch costs arising from within the language system, then stimulus valence should interact with switch costs when highly proficient bilinguals switch between low proficiency languages (e.g., L3 and L4) because the within-language inhibitory mechanism will be operating. In contrast, stimulus valence should not affect switch costs when highly proficient bilinguals switch between established languages (e.g., L1 and L2) where only the language-specific selection mechanism is operating.

CONCLUSION

Experiments 1 and 2 demonstrated that in speech production the asymmetric switch costs are not dependent on the presence

of bivalent stimuli, nor on switch predictability. Experiment 3 demonstrated that the effects of stimulus valence affects switch costs and the asymmetric switch cost during speech production, despite numerous demonstrations that this is not the case in lexical decision (Thomas and Allport, 2000; Orfanidou and Sumner, 2005). Furthermore, the modulation of the switch cost asymmetry by stimulus valence and the persistence of the switch cost for univalent items is best accounted for by theories of language switching that posit a role for competition within the language system. In particular, we suggest that the switch cost asymmetry arises because a component of the language system required for speech production (namely

phonology) is particularly susceptible to interference from the competing language. The observation that speech production continues to reveal a different pattern of switch costs compared to comprehension tasks suggests that future research needs to continue to examine the similarities and differences in performance across tasks.

ACKNOWLEDGMENTS

The present research was supported by an NSERC Discovery grant (258603) to MR and by the National Science Foundation under Grant No. 1349042 to FP.

REFERENCES

- Abutalebi, J., and Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguist.* 20, 242–275. doi: 10.1016/j.jneuroling.2006.10.003
- Allport, A., Styles, E. A., and Hsieh, S. (1994). “Shifting intentional set: exploring the dynamic control of tasks,” in *Conscious and Nonconscious Information Processing: Attention and Performance XV*, eds C. Umiltà and M. Moscovitch (Cambridge, MA: MIT Press), 421–452.
- Altmann, E. M. (2007). Comparing switch costs: alternating runs and explicit cuing. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 475–483. doi: 10.1037/0278-7393.33.3.475
- Balota, D. A., and Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 340–357. doi: 10.1037/0096-1523.10.3.340
- Christoffels, I. K., Firk, C., and Schiller, N. O. (2007). Bilingual language control: an event-related potential study. *Brain Res.* 1147, 192–208. doi: 10.1016/j.brainres.2007.01.137
- Chumbley, J. I., and Balota, D. A. (1984). A word’s meaning affects the decision in lexical decision. *Mem. Cognit.* 12, 590–606. doi: 10.3758/BF03213348
- Coltheart, M. (2004). Are there lexicons? *Q. J. Exp. Psychol.* 57, 115–1171. doi: 10.1080/02724980443000007
- 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
- Costa, A., and Santesteban, M. (2004). Lexical access in bilingual speech production: evidence from language switching in highly proficient bilinguals and L2 learners. *J. Mem. Lang.* 50, 491–511. doi: 10.1016/j.jml.2004.02.002
- Costa, A., Santesteban, M., and Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *J. Exp. Psychol.* 32, 1057–1074. doi: 10.1037/0278-7393.32.5.1057
- Damian, M. F. (2004). Asymmetries in the processing of Arabic digits and number words. *Mem. Cognit.* 32, 164–171. doi: 10.3758/BF03195829
- Declerck, M., Koch, I., and Philipp, A. M. (2012). Digits vs. pictures: the influence of stimulus type on language switching. *Bilingualism* 15, 896–904. doi: 10.1017/S1366728912000193
- Dijkstra, A., and Van Heuven, W. J. B. (1998). “The BIA- model and bilingual word recognition,” in *Localist Connectionist Approaches to Human Cognition*, eds J. Grainger and A. Jacobs (Hillsdale NJ: Erlbaum), 189–225.
- Fink, A., and Goldrick, M. (2015). Pervasive benefits of preparation in language switching. *Psychon. Bull. Rev.* 22, 808–814. doi: 10.3758/s13423-014-0739-6
- Finkbeiner, M., Almeida, J., Janssen, N., and Caramazza, A. (2006). Lexical selection in bilingual speech production does not involve language suppression. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 1075–1089. doi: 10.1037/0278-7393.32.5.1075
- Gollan, T. H., and Ferreira, V. S. (2009). Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 640–665. doi: 10.1037/a0014981
- Grainger, J., and Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *Q. J. Exp. Psychol.* 39A, 295–319. doi: 10.1080/14640748708401788
- Grainger, J., and Dijkstra, T. (1992). “On the representation and use of language information in bilinguals,” in *Cognitive Processing in Bilinguals*, ed R. J. Harris (Amsterdam: Elsevier), 207–220.
- Grainger, J., Midgley, K., and Holcomb, P. J. (2010). “Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d),” in *Language Acquisition Across Linguistic and Cognitive Systems*, eds M. Kail and M. Hickmann (New York, NY: Benjamins), 267–284.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism* 1, 67–81. doi: 10.1017/S1366728998000133
- Humphreys, G. W., Price, C. J., and Riddoch, M. J. (1999). From objects to names: A cognitive neuroscience approach. *Psychol. Res.* 62, 118–130. doi: 10.1007/s004260050046
- Jackson, G. M., Swainson, R., Cunnington, R., and Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism* 4, 169–178. doi: 10.1017/s1366728901000268
- Jackson, G. M., Swainson, R., Mullin, A., Cunnington, R., and Jackson, S. R. (2004). ERP correlates of a receptive language-switching task. *Q. J. Exp. Psychol.* 57, 223–240. doi: 10.1080/02724980343000198
- Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., et al. (2010). Control and interference in task switching - A review. *Psychol. Bull.* 136, 849–874. doi: 10.1037/a0019842
- Koch, I., Gade, M., Schuch, S., and Philipp, A. M. (2010). The role of inhibition in task switching - A review. *Psychon. Bull. Rev.* 17, 1–14. doi: 10.3758/PBR.17.1.1
- Kroll, J. F., Gullifer, J. W., and Rossi, E. (2013). The multilingual lexicon: the cognitive and neural basis of lexical comprehension and production in two or more languages. *Annu. Rev. Appl. Linguist.* 33, 102–127. doi: 10.1017/S0267190513000111
- Macizo, P., Bajo, T., and Paolieri, D. (2012). Language switching and language competition. *Second Lang. Res.* 28, 131–149. doi: 10.1177/0267658311434893
- Martin, C. D., Strijkers, K., Santesteban, M., Escera, C., Hartsuiker, R. J., and Costa, A. (2013). The impact of early bilingualism on controlling a language learned late: an ERP study. *Front. Psychol.* 4:815. doi: 10.3389/fpsyg.2013.00815
- Meuter, R. F. I., and Allport, A. (1999). Bilingual language switching in naming: asymmetrical costs of language selection. *J. Mem. Lang.* 40, 25–40. doi: 10.1006/jmla.1998.2602
- Ministry of Education and Training (1998). *The Ontario Curriculum: Core French, Grades 4-8*. Available online at: <http://www.edu.gov.on.ca/eng/curriculum/elementary/fsl.html>
- Ministry of Education and Training (1999). *The Ontario Curriculum Grades 9 and 10: Core, Extended, and Immersion French*. Available online at: <http://www.edu.gov.on.ca/eng/curriculum/secondary/fsl.html>
- O’Malley, S., and Besner, D. (2008). Reading aloud: qualitative differences in the relation between stimulus quality and word frequency as a function of context. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 1400–1411. doi: 10.1037/a0013084
- O’Malley, S., Reynolds, M. G., and Besner, D. (2007). Qualitative differences between the joint effects of stimulus quality and word frequency in reading

- aloud and lexical decision: extensions to Yap and Balota. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 451–458. doi: 10.1037/0278-7393.33.2.451
- Orfanidou, E., and Sumner, P. (2005). Language switching and the effects of orthographic specificity and response repetition. *Mem. Cogn.* 33, 355–369. doi: 10.3758/BF03195323
- Peeters, D., Runnqvist, E., Bertrand, D., and Grainger, J. (2014). Asymmetrical switch costs in bilingual language production induced by reading words. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 284–292. doi: 10.1037/a0034060
- Peressotti, F., Cubelli, R., and Job, R. (2003). On recognizing proper names: the orthographic cue hypothesis. *Cogn. Psychol.* 47, 87–116. doi: 10.1016/S0010-0285(03)00004-5
- Perry, C., Ziegler, J. C., and Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychol. Rev.* 114, 273–315. doi: 10.1037/0033-295X.114.2.273
- Philipp, A. M., Gade, M., and Koch, I. (2007). Inhibitory processes in language switching: evidence from switching language-defined response sets. *Eur. J. Cogn. Psychol.* 19, 395–416. doi: 10.1080/09541440600758812
- Reynolds, M., and Besner, D. (2006). Reading aloud is not automatic: processing capacity is required to generate a phonological code from print. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 1303–1323. doi: 10.1037/0096-1523.32.6.1303
- Rogers, R. D., and Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *J. Exp. Psychol.* 124, 207–231. doi: 10.1037/0096-3445.124.2.207
- Thomas, M. S. C., and Allport, A. (2000). Language switching costs in bilingual visual word recognition. *J. Mem. Lang.* 43, 44–66. doi: 10.1006/jmla.1999.2700
- Treasury Board of Canada Secretariat (2012). *Qualification Standards in Relation to Official Languages*. Available online at: <http://www.tbs-sct.gc.ca/gui/squn03-eng.asp>
- Vandierendonck, A., Liefoghe, B., and Verbruggen, F. (2010). Task switching: interplay of reconfiguration and interference control. *Psychol. Bull.* 136, 601–626. doi: 10.1037/a0019791
- Van Heuven, W. J. B., Dijkstra, A., and Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *J. Mem. Lang.* 39, 458–483. doi: 10.1006/jmla.1998.2584
- Van Selst, M., and Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Q. J. Exp. Psychol.* 47A, 631–650.
- Verhoef, K., Roelofs, A., and Chwilla, D. J. (2009). Role of inhibition in language switching: evidence from event-related brain potentials in overt picture naming. *Cognition* 110, 84–99. doi: 10.1016/j.cognition.2008.10.013
- Von Studnitz, R. E., and Green, D. W. (1997). Lexical decision and language switching. *Int. J. Bilingualism*, 1, 3–24. doi: 10.1177/136700699700100102
- Von Studnitz, R. E., and Green, D. W. (2002). The cost of switching language in a semantic categorization task. *Bilingualism* 5, 241–251. doi: 10.1017/S1366728902003036
- Yap, M. J., and Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 274–296. doi: 10.1037/0278-7393.33.2.274
- Yap, M. J., Tan, S. E., Pexman, P. M., and Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness, on lexical decision, speeded naming and semantic classification. *Psychon. Bull. Rev.* 18, 742–750. doi: 10.3758/s13423-011-0092-y

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

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



The Emergence of a Phoneme-Sized Unit in L2 Speech Production: Evidence from Japanese–English Bilinguals

Mariko Nakayama¹, Sachiko Kinoshita^{2,3} and Rinus G. Verdonschot^{4*}

¹ Faculty of Letters, Arts and Sciences, Waseda University, Tokyo, Japan, ² ARC Centre of Excellence in Cognition and its Disorders, Sydney, NSW, Australia, ³ Department of Psychology, Macquarie University, Sydney, NSW, Australia, ⁴ Waseda Institute for Advanced Study, Tokyo, Japan

OPEN ACCESS

Edited by:

Jay Rueckl,
University of Connecticut, USA

Reviewed by:

Claudio Mulatti,
Università degli Studi di Padova, Italy
Kenneth Forster,
University of Arizona, USA
Padraig G. O'Seaghdha,
Lehigh University, USA

*Correspondence:

Rinus G. Verdonschot
rinus@aoni.waseda.jp

Specialty section:

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

Received: 03 August 2015

Accepted: 29 January 2016

Published: 23 February 2016

Citation:

Nakayama M, Kinoshita S
and Verdonschot RG (2016)
The Emergence of a Phoneme-Sized
Unit in L2 Speech Production:
Evidence from Japanese–English
Bilinguals. *Front. Psychol.* 7:175.
doi: 10.3389/fpsyg.2016.00175

Recent research has revealed that the way phonology is constructed during word production differs across languages. Dutch and English native speakers are suggested to incrementally insert *phonemes* into a metrical frame, whereas Mandarin Chinese speakers use syllables and Japanese speakers use a unit called the mora (often a CV cluster such as “ka” or “ki”). The present study is concerned with the question how *bilinguals* construct phonology in their L2 when the phonological unit size differs from the unit in their L1. Japanese–English bilinguals of varying proficiency read aloud English words preceded by masked primes that overlapped in just the onset (e.g., bark-BENCH) or the onset plus vowel corresponding to the mora-sized unit (e.g., bell-BENCH). Low-proficient Japanese–English bilinguals showed CV priming but did *not* show onset priming, indicating that they use their L1 phonological unit when reading L2 English words. In contrast, high-proficient Japanese–English bilinguals showed significant onset priming. The size of the onset priming effect was correlated with the length of time spent in English-speaking countries, which suggests that extensive exposure to L2 phonology may play a key role in the emergence of a language-specific phonological unit in L2 word production.

Keywords: language production, masked priming, phonological unit, proximate unit, Japanese, bilingualism

INTRODUCTION

Speaking a word naturally requires assembling its phonology. According to the influential language production model by Levelt et al. (1999), this takes place through a process called *prosodification*. This entails first accessing a word's *phonological segments* (e.g., phonemes in English/Dutch), which are then incrementally inserted into a metrical frame (a structure specifying the number of syllables and the stress position). That is, producing a word such as “table” in English will first require access to its phonemes (i.e., /t/ /e/ /l/ /b/ /ə/ /l/) and metrical structure (i.e., bi-syllabic with stress on first syllable) which are then merged together to form the phonological word (i.e., /teɪ’bəl/). Constructing phonology on-line is essential for languages such as English and Dutch (on which the Levelt et al. (1999) model is mainly based) as these languages often need re-syllabification depending on the local context. For instance, the sentence “He’ll escort us.” is normally pronounced as /hɪl-ə-skər-təs/. As the cliticized form (/ə-skər-təs/) would not be stored in the lexicon, whether

the syllable /-skɔr/ or /-skɔrt/ will be created depends on the utterance context (Levelt et al., 1999, p. 23). The evidence that this process initially occurs in phoneme-sized units comes from results obtained in Dutch using the implicit priming (also called the form preparation) paradigm (Meyer, 1990, 1991). In this paradigm, participants learn prompt-response pairs (e.g., say “DANS” [dance] when presented with the prompt “feest” [party]). The prompted words are grouped in such a way that they either all overlapped in their initial segment(s) or not. Response words were produced significantly faster when there was overlap (e.g., DANS [dance], DOP [cap], DEUGD [virtue]) compared to when there was no overlap (e.g., DANS [dance], HEKS [witch], STOEP [sidewalk]). This significant facilitation is referred to as *the preparation effect*. In contrast, rime related overlap (e.g., BOEK [book], DOEK [canvas], SNOEK [pike]) did not produce facilitation, attesting to the incremental left-to-right (i.e., beginning to end) nature of the segment-to-frame association process.

Research on reading aloud has also revealed a similar left-to-right incremental segment-to-frame association process. Masked priming research using English (e.g., Forster and Davis, 1991; Kinoshita, 2000) and Dutch (e.g., Schiller, 2004) has also shown that when a prime is briefly presented (e.g., 50 ms) before a to-be-read-aloud target, naming latencies are significantly faster when the onset phoneme is shared (e.g., pole-PEAR) than when it is not (e.g., take-PEAR). Similar to findings observed with the implicit priming paradigm, no facilitation was observed in masked priming when only the last segments were shared (e.g., Kinoshita, 2000; Schiller, 2008). While the masked onset priming effect was originally interpreted in terms of a serial *letter-to-phoneme mapping process* (e.g., Forster and Davis, 1991), the emerging consensus is that this left-to-right incremental nature of reading reflects a speech production process (see e.g., Grainger and Ferrand, 1996; Roelofs, 2004; Malouf and Kinoshita, 2007).

The evidence for the left-to-right phoneme-to-frame association process mentioned above has come from European languages, mainly Dutch and English. However, languages differ in many respects, and recently it has been suggested that the unit used to fill the metrical frame may not always be the phoneme, but other languages may employ different unit sizes (see O’Seaghdha, 2015; Roelofs, 2015). For instance, in Mandarin Chinese (hereafter “Chinese”), Chen et al. (2002) and O’Seaghdha et al. (2010) employing the implicit priming paradigm, found reliable preparation effects only when a group of response words overlapped in the first (atonal) *syllable*; no facilitation was observed when a group of response words overlapped in the onset phoneme. The initial unit involved to build phonology (termed the “proximate unit” by O’Seaghdha et al., 2010) in Chinese, therefore, seems to be the syllable, and not the phoneme (see also You et al., 2012, for related results).

The Proximate Unit of Japanese Word Production

Japanese is known to have a mora-based timing (Warner and Arai, 2001; Kureta et al., 2006). The Japanese mora is a supra

phonemic unit that usually involves a CV or V (e.g., /ka/ or /a/), nasal coda (/N/), or a geminate (/Q/) combination, but *never* a single consonant (e.g., /k/). The mora as a proximate unit has accounted for many Japanese psycholinguistic findings ranging from speech segmentation (e.g., Cutler and Otake, 1994), speech errors (e.g., Kubozono, 1989), and children’s word games (e.g., “shiritori,” in which the players take turns in generating a word that starts with the final mora of the word the other player has produced: e.g., “kobuta” (piglet) – “tanuki” (badger) – “kitsune” (fox) – “neko” (cat) and so on, see e.g., Katada, 1990). Phonological awareness tests typically assess skills of mora level manipulation, not phonemes (e.g., Sasanuma et al., 1996). The central importance of the mora as *the* phonological unit in Japanese is further evidenced in the phenomenon of “vowel epenthesis,” a form of phonological restoration: When presented with a non-word containing an illegal consonant cluster like “ebzo,” Japanese listeners hear an illusory vowel, reporting they heard “ebuzo” (Dupoux et al., 2001). Moreover, Japanese listeners show no mismatch negativity in evoked potentials to a change from “ebzo” to “ebuzo,” whereas French listeners do (Dehaene-Lambertz et al., 2000). Additionally, when producing English words, Japanese people typically insert vowels when a word contains phoneme clusters (Broselow and Park, 1995; Broselow and Kang, 2013).

Previous studies on word production also indicate the critical role of the mora during Japanese phonological encoding. Kureta et al. (2006) using the implicit priming paradigm found significant preparation effects in Japanese *only* when a group of response words overlapped in the initial mora, but *not* when they merely overlapped in the onset phoneme. Using a masked priming read-aloud paradigm, Verdonschot et al. (2011) reported that Japanese words were read aloud significantly faster when a target was preceded by a prime overlapping in the initial mora (e.g., teki-TENS^{HI}) relative to unrelated primes (e.g., heki-TENS^{HI}). Critically, reading of the Japanese words *never* benefited from a prime overlapping in the onset phoneme (e.g., tomi-TENS^{HI}) relative to a control prime (e.g., gomi-TENS^{HI}).

One important point in interpreting the masked onset priming effect is the role of script. Indo-European languages like English and Dutch use the alphabetic writing system, in which a letter (or letter cluster e.g., “sh”) maps onto a phoneme. Chinese is written using a logography in which a character maps onto a (morpho-) syllable. Japanese is written both in “kanji” (literally “Chinese characters”), the logography borrowed from the Chinese, and in “kana” (hiragana/katakana), two inventories, consisting of 48 characters each, mapping onto a mora (e.g., ニ [ni]・ホ [ho]・ン [N] and に [ni]・ほ [ho]・ん [N], for katakana and hiragana, respectively).¹ In the masked priming read-aloud experiments mentioned earlier, all words were presented in their native script, i.e., alphabetic letters in English (e.g., Kinoshita, 2000) and Dutch (e.g., Schiller, 2004) and kana in Japanese (Verdonschot et al., 2011). As noted earlier, an alternative interpretation of the masked onset priming effect in reading-aloud is that it might

¹Hiragana and katakana are allographs, somewhat akin to the uppercase/lowercase distinction in the Roman alphabet. Katakana are used for foreign loan words and hiragana is used for words of Japanese origin and grammatical morphemes.

originate in the mapping of letters to phonemes (Forster and Davis, 1991). From this perspective, the absence of masked onset priming effect in studies that used non-alphabetic script like kana may be interpreted as reflecting the size of the unit involved in the mapping of written script to phonology, rather than the size of the phonological unit involved in speech production. To test this possibility, Verdonschot et al. (2011) conducted two experiments in which Japanese target words were presented in “romaji” (alphabetic transcriptions). However, no significant onset priming effect was found in either experiment, suggesting that for the Japanese speakers, the effect depended on the size of the phonological unit used in speech production rather than print-to-speech conversion.

Phonological Units in L2 Word Production

Given the increasing evidence concerning the difference between languages in the proximate unit (the primary unit used in the phonological encoding process), the logical next step is to investigate how *bilinguals* process words, and how L2 proficiency modulates this process. Not surprisingly, the earlier L2 is acquired, the more native-like the bilingual speakers' pronunciation of L2 becomes (see Piske et al., 2001, for a review). As noted by Alario et al. (2010) however, most of research on this issue has focused on the acoustic properties of bilinguals' speech, and studies focusing on the cognitive mechanisms involved in the spoken production of L2 are very scarce. In particular, it is currently unknown what phonological unit is used in L2 production when bilinguals' two languages have different unit sizes.

To our knowledge, only one study to date has investigated this matter (Verdonschot et al., 2013). That study involved highly proficient Chinese-English bilinguals to read aloud English targets primed by English words. Naming latency was significantly faster when a target was primed by an onset-related English word (e.g., bark – BENCH) than by an unrelated prime (e.g., dark-BENCH). As noted earlier, the phonological unit of monolingual Chinese speakers is known to be a syllable (e.g., CVC). Therefore, the significant onset priming observed for Chinese-English bilinguals suggests that highly proficient bilinguals used a phonological unit suited to produce L2 words (i.e., phonemes), one that is different in size from the phonological unit normally used in their L1 production (i.e., syllable).

A possible limitation concerning Verdonschot et al. (2013) is that the Chinese-English bilinguals were all highly proficient. It is therefore unknown whether the ability to prepare phonology in the unit of L2 develops with proficiency in L2. Also, Verdonschot et al. (2013) did not include a group of native English speakers. Therefore, it would be important to show that the high-proficient bilinguals behave more like native English speakers than the low-proficient bilinguals in producing a significant onset priming effect with the same set of stimuli.

Present Study

The present study investigated the proximate unit used by Japanese-English bilinguals of varying proficiency in reading aloud L2 (English) words. Specifically, we were interested in

whether the L1 Japanese speaker constructs L2 English phonology by placing moras (CVs) or phonemes (specifically consonants, given that a vowel is also a mora) in the metrical frame, and whether the size of the phonological unit is modulated by L2 proficiency. To assess this, low-proficient bilinguals, high-proficient bilinguals and native English monolingual speakers (Experiments 1–3, respectively) read aloud English target words that were preceded either by English prime words that shared the initial onset phoneme (bark-BENCH) or by words that shared the initial CV (i.e., mora; bell-BENCH), with priming effects measured against their respective unrelated primes (dark-BENCH and cell-BENCH). Assuming that the low-proficient Japanese-English bilinguals would use the phonological unit of their first language (the mora), they should show CV (*mora*) priming effects (bell-BENCH < cell-BENCH), but not onset priming (phoneme) effects (bark-BENCH = dark-BENCH). Alternatively, if a significant onset effect is observed for low-proficient bilinguals, this would then suggest that the proximate unit of L2 English (phoneme) can be adopted relatively early in the course of L2 acquisition. In contrast, high-proficient bilinguals are more likely to show onset effects, based on the finding by Verdonschot et al. (2013) with high-proficiency Chinese-English bilinguals. If so, this would extrapolate previous findings (L1-Chinese vs. L2-English) to a group of bilinguals whose two proximate units also diverge in their two languages (L1-Japanese vs. L2-English). Finally, we expect the group of native English speakers to show significant onset priming effects, in line with previous studies (Forster and Davis, 1991; Kinoshita, 2000; Schiller, 2004).

EXPERIMENT 1: LOW-PROFICIENT JAPANESE-ENGLISH BILINGUALS

Methods

Participants

Forty-five low proficient Japanese-English bilingual students from Waseda University (Tokyo, Japan) participated in the experiment in return for payment of 1000 Yen (~US\$8). Their mean TOEIC (Test of English for International Communication) score was 715 (range = 600–790).² This study was carried out in accordance with the recommendations of ‘the Ethics Guidelines for Scientific Research with Human Subjects, Ethics Review Committee on Research of Waseda University’ and ‘the Human Research Ethics Committee of Macquarie University.’ Prior to the experiments, all subjects gave written informed consent in accordance with the Declaration of Helsinki.

²The TOEIC test is a paper-and-pencil, multiple-choice assessment developed and administered by Educational Testing Service (ETS). There are two separately timed sections of 100 questions each. It assesses a broad range of English skills (particularly reading and listening), especially in business settings. The test scores range from 10 to 990, with higher values indicating greater English proficiency. Many university students in Japan voluntarily take the test to quantify their English ability because many Japanese companies request to report TOEIC scores on their job applications.

Stimuli

The critical stimuli were 42 English medium frequency words ($M = 50.3$ occurrences per million, Kućera and Francis, 1967). The mean letter length and syllable size of the targets were 5.1 ($SD = 0.9$) and 1.5 syllables. The syllable length was equally distributed between one ($n = 21$) or two syllables ($n = 21$). For each target, four types of monosyllabic English word primes were selected: (1) C prime: a word that had the same onset phoneme with the target (e.g., bark-BENCH), (2) C-control prime: a word that shared all the letters with the onset prime except for the initial letter (e.g., dark-BENCH), (3) CV prime: a word that had the same CV with the target (e.g., bell-BENCH) and (4) CV-control prime: a word that shared all the letters with the CV prime except for the initial letter (e.g., cell-BENCH).³ This ensured that CV prime-target pairs do not have an additional letter (and phoneme) overlap compared with C prime-target pairs (e.g., Verdonschot et al., 2011, 2013). In addition, the bodies of C/CV primes and their controls always had the same pronunciation (e.g., -ark in “bark – dark” or –ell in “bell – cell”). The mean word frequencies (per million) of the four types of primes (C, C-control, CV, CV-control) were comparable: 52.8, 59.4, 58.2, and 50.9, respectively. The word lengths (in letters) of the four types of primes were also comparable (3.6, 3.6, 3.8, and 3.8). For the C and CV conditions, there were two counterbalanced lists; within each condition, half of the targets were primed by the critical primes in one list, and the same targets were primed by their control primes in the other lists, and vice versa. The list of prime and target stimuli used can be found in the Supplementary Materials.

To check the possibility that an absence of masked priming might be due to the lack of familiarity with the alphabetic letters, an identity priming condition was included (and also in subsequent two experiments). The masked identity priming effect is known to be unaffected by word frequency (Forster and Davis, 1984) and it is generally interpreted to reflect a “head-start” in orthographic processing (Gomez et al., 2013). The presence of a typical identity priming effect (e.g., the sizes of priming effects being ± 10 ms of the prime duration, see Forster et al., 2003) would indicate participants’ ability to process masked primes in alphabetic script.

For the identity priming condition, a different set of 42 medium frequency targets were selected ($M = 83.2$ per million). The mean length of these targets was 4.4 letters (60% consisted of one syllable, 40% consisted of two syllables). Each target (e.g., SOFT) was primed either by the target itself (i.e., soft) or by a control prime that did not share any letters with the target at the same position (e.g., page). The mean word frequency and the mean length of the control primes were 74.3 per million and 4.4 letters. None of the words in the identity priming condition were used in the C/CV conditions. For the identity priming condition, there also were two counterbalanced lists in order to present the same targets to all

participants but each participant saw only one of the prime-target pairings.

Apparatus and Procedure

Participants were tested individually using the DMDX software package (Forster and Forster, 2003). Each trial began with the presentation of a forward mask (#####) for 500 ms followed by a 50 ms presentation of a lower case prime. Immediately following the prime, a target was presented in upper case. The target remained on the display until the participant made a response. Participants were instructed to read aloud the target as quickly and accurately as possible. The stimuli were presented at the center of the screen in 12-pt Courier New font. The presence of primes was not mentioned to any participant. Participants completed 16 practice trials to familiarize themselves with the task.

For the C and CV priming conditions, the same set of 42 targets was presented twice, once in the C condition and once in the CV condition. The identity priming condition was always presented in between the C and CV conditions. Half of the participants were presented with the C condition in the first block, and the CV condition in the third block; the other half were presented with the CV condition in the first block, and the C condition in the third block. Targets primed by critical primes (either C or CV) in the first block were primed by control primes in the third block, and vice versa. Therefore, for the C/CV conditions, although there were two counterbalancing lists with regard to prime-target relationships (i.e., related vs. control), there were four presentation orders differing in whether the target was paired with an C prime or a CV prime first, crossed with the two lists.

Results

Raw naming reaction times (RTs) were checked using CheckVocal Software (Protopapas, 2007). We used a linear mixed-effect (LME) model (lme4; Baayen, 2008; Bates et al., 2008) implemented in R (R Development Core Team, 2008) to analyze RT for correct trials and error rates. *lmerTest* package in R was used to calculate the *p*-values using Satterthwaite’s approximation for the degrees of freedom (Kuznetsova et al., 2014). In order to meet the distributional assumptions of LME, we applied the inverse transformation to the RTs ($-1000/RT$) to better approximate normality in the RT distribution (see Box and Cox, 1964). Correct data points that were 3.5 SD away from the individual’s mean per condition were removed as outliers (both 0.3% of the data in the C/CV conditions and Identity condition, respectively). In the identity priming condition, three items (DENY, TINY, RIFLE) were removed due to high error rates ($> 55\%$).

For the C and CV conditions, the initial model included Overlap (CV vs. C), Prime Type (related vs. control) and Order (first vs. third block) and their interaction as fixed factors, and by-subject intercept and slope and by-item intercept and slope of Overlap, Prime Type, and their interaction as random factors. Note that Block 2 is not considered in the Order variable as it always contained the identity primes. Each of the categorical variables was contrast coded

³ As most moras found in Japanese consist of an onset and a nucleus (i.e., CV), the CV primes used in our experiments are functionally equivalent to a mora (especially for low proficient bilinguals).

TABLE 1 | Mean naming latencies (ms) and percentage errors (%) for English targets primed by C primes, C control primes, CV primes, CV control primes, identity primes and identity control primes in Experiment 1, for low-proficient bilinguals.

	C condition	CV condition	Identity condition
Related prime	635 (5.2%)	627 (3.4%)	660 (6.1%)
Control prime	633 (5.4%)	652 (4.2%)	701 (8.6%)
Priming effect	−2 (0.2%)	25 (0.8%)	41 (2.5%)

by 0.5/−0.5. We also entered the following target lexical characteristics as fixed factors: Log sublex frequency (Brysbaert and New, 2009), Orthographic neighborhood size (Ortho-N), and Length. These continuous variables were centered around their respective means. In addition, because Length and Ortho-N were moderately correlated ($r = 0.53$), Ortho-N was regressed against Length and their residuals were used as a predictor for Ortho_N (i.e., res_Ortho-N). Thus the model used in the analyses was $[\text{invRT} \sim \text{Overlap} * \text{PrimeType} * \text{Order} + \text{Log sublex frequency} + \text{res_Ortho-N} + \text{Length} + (1 + \text{Overlap} * \text{Prime Type} | \text{subject}) + (1 + \text{Overlap} * \text{Prime Type} | \text{target})]$.

For the identity priming condition, the model used in the analyses was the same as above except that Order was not included as a factor. For the C, CV and Identity conditions, errors were analyzed using a mixed-effects logistic model (Jaeger, 2008) using the same fixed factors used for RT analyses.⁴ However, the error rates were small and there were no significant priming effects in any conditions except in the identity condition ($p = 0.02$), therefore we will only report the results of response latencies analyses. Table 1 shows the mean RT and error rates for the three conditions. Table 1 shows the mean RT and error rates for the three conditions.

Onset (C) and CV Priming Effects

Order did not significantly affect the patterns of priming as indicated by the lack of three-way interaction between Order, Overlap, Prime Type ($t < 1$) and also by the lack of two-way interaction between Order and Prime Type ($t < 1$). The main effect of Order was statistically significant ($t = 5.10, p < 0.001$); naming latencies were significantly faster in the third block than in the first block. The main effect of Prime Type was significant ($t = 4.13, p < 0.001$). The main effect of Overlap was not significant ($t < 1$). Importantly, there was a significant interaction between Overlap and Prime Type ($t = 2.57, p = 0.014$). Follow-up analyses of this interaction revealed that there was *no* C (onset) priming ($t = 1.02, p > 0.10$, a -2 ms difference; in contrast, there was a significant CV priming effect ($t = 4.46, p < 0.001$, a 25 ms effect). As for the effects of target lexical characteristics, there was a significant effect of Log sublex frequency ($t = -7.47, p < 0.001$), Ortho-N ($t = -3.76, p < 0.001$), and Length ($t = 4.23, p < 0.001$), that is, faster naming latencies were associated with targets with higher frequency, more orthographic neighbors, and shorter lengths.

⁴In the analyses of errors, the initial model often failed to converge because of the complex specifications of random factors. In that case, we successively removed a random factor term until a model successfully converged.

Identity Priming Effects

The effect of Prime type was significant ($t = 8.05, p < 0.001$); targets were named 41 ms faster when they were primed by identity words than by control words. This confirmed that low proficient bilinguals are able to process masked English primes sufficiently. The model also revealed a significant effect of Log sublex frequency ($t = -2.92, p < 0.01$), and Length ($t = 3.26, p < 0.01$); higher frequency and shorter targets were associated with faster naming latency. The effect of Ortho_N was not significant ($t < 1$).

Discussion

The critical result of Experiment 1 was that low-proficient Japanese–English bilinguals did not show an onset priming effect for L2-English targets (e.g., bark-BENCH = dark-BENCH). This finding differs from the significant onset priming effects typically found in reading aloud with native speakers of European languages (e.g., Forster and Davis, 1991; Kinoshita, 2000; Schiller, 2004) or the result obtained in Verdonshot et al. (2013) with proficient Chinese–English bilinguals. The low-proficient bilinguals, nevertheless, showed significant CV (mora) priming (bell-BENCH < cell-BENCH). In fact, the absence of onset priming together with the presence of CV (mora) priming parallel those reported by Verdonshot et al. (2011) with Japanese native speakers reading aloud Japanese kana and romaji-transcribed words. These data taken together suggest that the low-proficient bilinguals carried over their L1 unit to L2 word production.

In Experiment 2, high-proficient Japanese–English bilinguals were tested. Based on the results of Verdonshot et al. (2013) who found significant onset priming with proficient Chinese–English bilinguals, we expect to replicate that finding.

EXPERIMENT 2: HIGH-PROFICIENT JAPANESE–ENGLISH BILINGUALS

Methods

Participants

Forty-four highly proficient Japanese–English bilingual students from Waseda University (Tokyo, Japan) participated in the experiment for 1000 Yen (US\$8). Their mean TOEIC score was 876 (range = 800–990) and they started studying English on average at the age of 9.9.

Stimuli

The stimuli were same as Experiment 1.

Apparatus and Procedure

These were identical to Experiment 1.

Results

The data were analyzed identically to Experiment 1. For response latency analyses, the same outlier removal resulted in the removal of 0.3% of the data in the C/CV conditions, and 0.4% of the data in the Identity condition. In the identity priming condition, one item (DENY) was removed due to high error rates (>55%). Errors were analyzed identically to Experiment 1. However,

TABLE 2 | Mean naming latencies (ms) and percentage errors (%) for English targets primed by C primes, C control primes, CV primes, CV control primes, identity primes and identity control primes in Experiment 2, for high-proficient bilinguals.

	C condition	CV condition	Identity condition
Related prime	586 (3.9%)	563 (3.6%)	589 (6.1%)
Control prime	603 (2.8%)	584 (4.2%)	630 (7.3%)
Priming effect	17 (−1.1%)	21 (0.6%)	41 (1.2%)

again, error rates were generally very small, and there was no significant priming effect in any conditions, therefore, we only report the results of the response latency analyses. **Table 2** shows the mean RT and error rates for the three conditions.

Onset (C) and CV Priming Effects

As was the case in Experiment 1, Order did not significantly modulate the patterns of priming effects ($t_s < 1$). As expected, the main effect of Order was significant ($t = 7.95, p < 0.001$), with targets being named significantly faster in the third than in the first block (note: Block 2 always contained identity primes, and therefore was not analyzed). The main effect of Prime Type was significant ($t = 7.70, p < 0.001$). There also was a significant effect of Overlap ($t = -3.56, p < 0.001$); across Prime Type, targets in the CV condition were named significantly faster than targets in the C condition. The two-way interaction between Overlap and Prime Type was marginally significant ($t = 1.91, p = 0.064$). Follow-up analyses of this marginal interaction revealed that high-proficient Japanese–English bilinguals showed a significant C (onset priming) effect ($t = 3.71, p < 0.001$, a 17 ms effect) as well as a significant CV priming effect ($t = 6.63, p < 0.001$, a 21 ms effect). The significant onset priming effect was consistent with the result of Verdonchot et al. (2013) with high-proficient Chinese–English bilinguals. As for the lexical characteristics of the targets, shorter naming latencies were associated with higher target frequency, ($t = -7.73, p < 0.001$), more orthographic neighbors ($t = -2.60, p < 0.05$) and shorter target length ($t = 3.53, p < 0.001$).

Identity Priming Effects

For response latency, as expected, the effect of Prime type was highly significant ($t = 9.58, p < 0.001$). Targets were named 41 ms faster when they were primed by identity words than by control words, again displaying the ability of bilinguals to efficiently process masked English primes. Among the target lexical characteristics, there was a significant effect of length ($t = 3.08, p < 0.01$) and a marginally significant effect of frequency ($t = -1.94, p = 0.061$). Shorter response latency was associated with higher target frequency and shorter target length. The effect of orthographic neighborhood size was not significant ($t < 1$).

Discussion

Consistent with our prediction, highly proficient Japanese–English bilinguals showed a significant onset priming effect (17 ms) when reading aloud English words. This result suggested that high-proficient Japanese–English bilinguals employed a

phoneme-sized proximate unit when producing L2 English words, although their L1 proximate unit is the mora (CV). The fact that the present results mirror those reported earlier with Chinese–English bilinguals (Verdonchot et al., 2013) strengthens the view that high-proficient bilinguals are able to use the proximate unit of the L2 language being spoken.

Both the low-proficient bilinguals (Experiment 1) and the high-proficient bilinguals (Experiment 2) showed a significant CV priming effect (bell-BENCH < cell-BENCH). This effect is not critical to our hypothesis (which concerns primarily the onset priming effect) and it could reflect mora priming (the basis on which we have expected low-proficiency bilinguals to show priming), or alternatively, it may reflect priming due to an overlap of two phonemic segments (initial C and V). It is not possible to determine *a priori* whether the CV priming effect observed with the high-proficient bilinguals reflects the usage of mora or phonemes. Nevertheless, there is one particular clue pointing toward the latter possibility, which is the fact that unlike the low-proficient bilinguals, the high-proficient bilinguals did not show statistically significantly greater priming due to an overlap in CV than C (onset) alone. This is consistent with the pattern that has been observed with monolingual speakers of English: an additional overlap in the vowel segment beyond the consonantal onset overlap leads to only a small increment in priming. For example, Kinoshita (2000) used 3-letter CVC non-word targets and reported that the onset priming effect (e.g., suf-SIB vs. muf-SIB) was substantial but an extra vowel overlap (sif-SIB) added only a statistically non-significant 3 ms increment; similarly, Mousikou et al. (2010) reported a small 4 ms (though statistically significant) increment.

In Experiment 3, we tested monolingual native speakers of English (i.e., a non-moraic language) using the same set of stimuli used in the preceding experiments. A successful demonstration of significant onset priming with native English speakers will further support the interpretation that high-proficient bilinguals (Experiment 2) used a phoneme-sized unit in producing the English words. Further, if the native English speakers show similar C and CV priming patterns as the high-proficient bilinguals, then such results will suggest that the high-proficient bilinguals' CV priming effect was likely due to phonemic segmental overlap rather than mora-level overlap.

EXPERIMENT 3: MONOLINGUAL NATIVE ENGLISH SPEAKERS

Methods

Participants

Forty-four monolingual native English speakers from Macquarie University (Sydney, Australia) participated in the experiment in return for course credit.

Stimuli

The same stimuli used in previous experiments were used.

Apparatus and Procedure

These were identical to the previous experiments.

TABLE 3 | Mean naming latencies (ms) and percentage errors (%) for English targets primed by C primes, C control primes, CV primes, CV control primes, identity primes and identity control primes in Experiment 3, for monolingual English speakers.

	C condition	CV condition	Identity condition
Related prime	439 (2.0%)	432 (1.7%)	419 (3.0%)
Control prime	459 (3.7%)	459 (4.8%)	468 (5.4%)
Priming effect	20 (1.7%)	27 (3.1%)	49 (2.4%)

Results

The data were analyzed identically to Experiments 1 and 2. For response latency analyses, 0.1% of the data was removed as outliers in the C/CV conditions and also in the Identity condition. Errors were also analyzed identically to Experiment 1 and 2. However, again, error rates were generally very small, therefore, in what follows, we only report the results of response latencies. **Table 3** shows the mean RT and error rates for the three conditions.

Onset (C) and CV Priming Effects

Again, the Order did not significantly affect the patterns of priming effects ($t_s < 1.68$, $p_s > 0.10$). The main effect of Order was significant ($t = 2.05$, $p < 0.05$) with faster naming latency in the third than in the first block. The main effect of Prime Type was significant ($t = 11.19$, $p < 0.001$). The effect of Overlap was not significant ($t < 1$). Similar to Experiment 2, there was a marginally significant interaction between Overlap and Prime Type ($t = 1.94$, $p = 0.061$). As expected, the follow-up interaction of the marginal interaction confirmed that there was a significant C priming effect ($t = 8.67$, $p < 0.001$, a 20 ms effect) as well as a significant CV priming effect ($t = 8.52$, $p < 0.001$ a 27 ms effect). Higher frequency targets were significantly associated with faster responding ($t = -3.91$, $p < 0.001$). Effects of target length or orthographic neighborhood size were not significant, both $t_s < 1$.

Identity Priming Effects

There was a significant identity priming effect ($t = 12.89$, $p < 0.001$); targets primed by identity primes were named 49 ms faster than the same targets primed by unrelated primes. There was a significant effect of target frequency ($t = -3.58$, $p < 0.001$). There were no effects of ortho_N or Length (both $t_s < 1.1$).

Discussion

Monolingual native English speakers showed a significant onset priming effect, and a CV priming effect that did not differ in size (statistically) from the onset priming effect. This pattern mirrors that observed with the high-proficient bilinguals and contrasts with the low-proficient bilinguals (who showed a significant CV priming effect but not an onset priming effect). We take the results of Experiment 3 to suggest that the CV priming effect observed with the high-proficient bilinguals likely reflected an effect of phonemic segmental overlap.

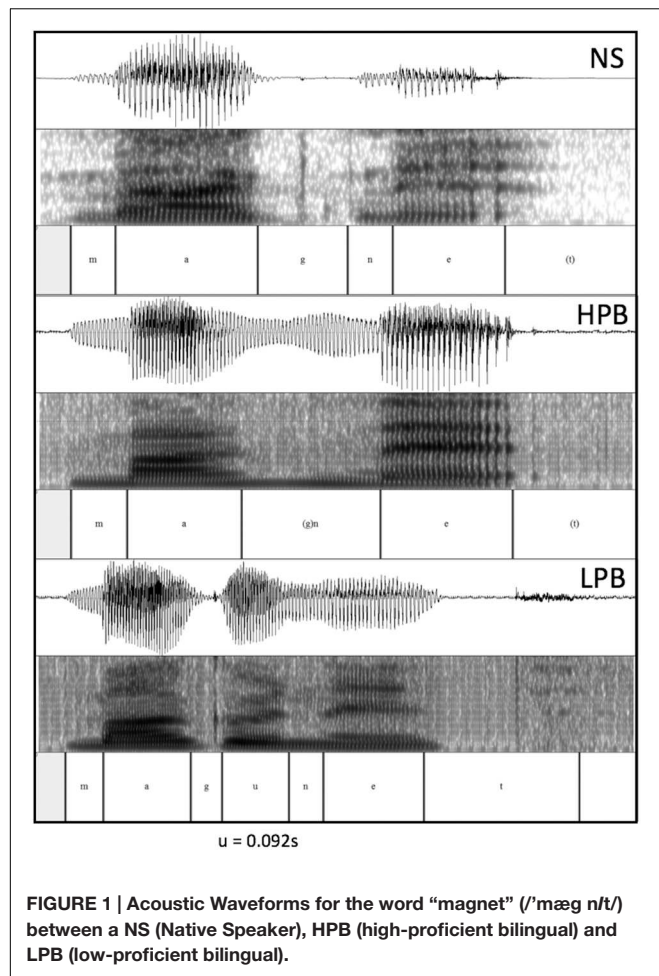
GENERAL DISCUSSION

Previous studies have shown that the phonological unit used in word production differs across languages: for L1 English and Dutch speakers, the unit is suggested to be the phoneme (Levelt et al., 1999; Roelofs, 2015), for Chinese, the syllable (Chen et al., 2002; O'Seaghdha et al., 2010), and for Japanese, the mora (e.g., Kureta et al., 2006; Verdonschot et al., 2011; Verdonschot and Tamaoka, 2015). The current paper examined the phonological unit size used in L2 word production when bilinguals' L1 and L2 languages employ different phonological unit sizes. The second, most essential, goal of this study was to investigate whether L2 English proficiency plays a role in the emergence of a phoneme-sized unit in English word production. To answer these two questions, we tested high- and low- proficient Japanese-English bilinguals in a masked priming read-aloud task. The results were clear: high-proficient bilinguals showed significant onset priming, but low-proficient bilinguals did not. The two groups of bilinguals, nevertheless, produced virtually identical identity and CV priming.

The results obtained with the low-proficient bilinguals – the absence of onset priming in the bilinguals whose first language is Japanese reading aloud English words – is important in establishing that the onset priming effect is not driven solely by the type of script (alphabetic letters). As noted, the original interpretation of masked onset priming effect was in terms of a serial letter-to-phoneme mapping process (Forster and Davis, 1991). The fact that low-proficient Japanese-English bilinguals do not show the onset priming effect indicates that reading aloud involves more than the mapping of letters to phonemes, and a full explanation of priming effects in reading aloud needs to take into account the processes involved in speech production.

Consistent with the assumption that the low-proficient bilinguals would use the phonological unit of their first language, the mora (CV), they showed no onset priming effect. In contrast, the high-proficient bilinguals showed an onset priming effect. These results suggest that highly proficient bilinguals seem to construct L2 English phonology similarly to native English speakers by incrementally inserting phonemes into the metrical frame.

Additional support for the claim that low-proficient bilinguals used their L1 proximate unit (mora) to read aloud L2 words can be seen in the evidence of vowel insertions into a consonant cluster. **Figure 1** shows the acoustic waveforms for the word “magnet” produced by a native English speaker, a high-proficiency bilingual, and a low-proficiency bilingual speaker (all female). It can be seen that compared to native-speakers and high-proficient bilinguals (who do not insert vowels at g) this particular low-proficient bilingual is inserting an extra vowel in the word-medial consonant cluster, thereby changing the word structure from a disyllable to three (or possibly four) morae. Considering that the duration of a “real” vowel of this particular participant (“a” in “mAgnet”) is about 0.12 s, it seems reasonable to suggest that the “u” (~0.092 s) is an epenthetic vowel with full insertion. We should point out that not all of our stimuli contained a consonant cluster, and also the likelihood of vowel insertion varies between consonant clusters (it is most evident



for consonant clusters containing voiced stops). A more formal analysis of this phenomenon will therefore remain a topic for the future.

What Aspects of L2 Proficiency are Responsible for the Use of an L2 Phonological Unit?

An obvious question that arises from the present study is *what aspect of proficiency* in L2 (English) is responsible for the shift in the proximate unit size used in L2 production. In the present experiments, the mean TOEIC score for highly proficient bilinguals was significantly higher than for low proficient bilinguals [mean for the highly proficient group = 876, low = 715, $t(87) = 15.33$, $p < 0.001$]. Our language history questionnaire, however, indicated that the two groups of bilinguals also differed in two other potentially important variables: (1) L2 AoA (the age at which the participant started learning English) [mean for the highly proficient group = 9.88, low = 11.53, $t(87) = -2.94$, $p = 0.004$]; and (2) the number of months spent in an English-speaking country [mean for the highly proficient group = 21.20, low = 1.74, $t(86) = 4.00$, $p < 0.001$]. That is, our high proficient bilinguals started learning

English significantly earlier and spent much longer in English speaking countries than low-proficient bilinguals did.

In order to find out which of the three factors: TOEIC (range 615–990), L2 AoA (range: 2–13), and Time spent in an English speaking country (range: 0–120 months) mostly contributed to the use of phoneme-size unit in speaking English words we analyzed the data from all bilinguals with a LME model, using the factors as continuous variables. Our initial analyses indicated that across all bilinguals, the three variables were correlated with each other: (1) TOEIC and the L2 AoA ($r = -0.399$, $p < 0.001$); (2) TOEIC and the Time spent ($r = 0.511$, $p < 0.001$); and (3) L2 AoA and Time spent, ($r = -0.493$, $p < 0.001$). In order to assess the unique predictive ability of each variable, the three factors were entered simultaneously in the model along with their respective interaction term with Prime Type, with the inverse RT as a dependent variable. All of the continuous variables were centered around their respective means.

The analysis revealed that the time spent in an English-speaking country significantly modulated onset priming ($t = 2.23$, $p = 0.026$), suggesting that the more time the participant spent in an English-speaking country, the greater the onset priming effect. Somewhat surprisingly, neither the TOEIC score nor the L2 AoA themselves uniquely explained the size of onset priming (both $t < 1$). This was also the case even when the effect of each variable was assessed individually ($t = 2.30$, $p = 0.021$ for the time spent in an English-speaking country, both $t_s < 1$ for the TOEIC and L2 AoA).

Our analyses, therefore, showed that it was not the TOEIC score or L2 AoA, but it was the time spent in English speaking countries that contributed to the development of the phoneme-sized unit in L2 English production. Naturally, immersion in the L2 environment also leads to higher English proficiency as indicated by the significant relationship between TOEIC scores and the time spent in an English speaking country. The fact that the TOEIC score did not predict the onset priming effect is perhaps not too unexpected, given the test places greater emphasis on reading and listening comprehension rather than speech production. Thus to our question “what aspects of L2 proficiency are responsible for adopting an L2 proximate unit?”, a viable answer would be the extensive exposure to the L2 language environment (which is also associated with higher proficiency in L2). As this conclusion is based on a *post hoc* analysis, it needs to be corroborated in future studies using other indices that assess speech production ability more directly. However, from a practical point of view, the finding that the acquisition of the phoneme-sized phonological unit did not depend on L2 AoA is rather encouraging, as it suggests that the L2-specific proximate unit can be adopted by typical L2 learners of English residing in Japan who usually start learning English around the age of 10–13. Although acquisition of many aspects of phonology in a non-native language (e.g., accents) are suggested to be restricted by L2 AoA (e.g., Flege, 1988; Alario et al., 2010), the phonological encoding processes seem to be able to adapt their internal workings well after the L1 phonological unit size has been fully developed.

CONCLUSION

O'Seaghdha et al. (2010) recently put forward the “proximate unit” principle, which suggests that the *initial phonological unit* used in the word-form encoding process differs across languages. Here we showed that when phonologically encoding English words, while low-proficient Japanese–English bilinguals use the phonological unit of their first language, namely the mora (CV), high-proficient bilinguals are able to use the phonological unit of the target language, namely the phoneme. Our data further showed that neither the L2 AoA or proficiency measured by standard tests of proficiency in English as a second language, but extensive exposure to its phonology seems to play a key role in the emergence of a phonological unit used in the construction of speech sounds in the second language.

A term frequently found in the psycholinguistic literature is the “Masked Onset Priming Effect” or MOPE (e.g., Schiller, 2008; Mousikou et al., 2010) which refers to the finding in Indo-European languages (such as English, Dutch) that faster speech onset latencies occur when reading aloud target words that are preceded by a prime sharing its onset with the target. However, it might be more reasonable to use the term “Masked Initial Segment Priming Effect” or MISPE instead as it has been shown that the effect may depend on the language at hand (e.g., the *onset* in Dutch/English, the *mora* in Japanese and the *syllable* in Chinese) as well as an individual's proficiency level.

An issue that should be investigated in future studies is whether the present findings will be generalized to other tasks

that are known to tap similar underlying phonological encoding processes (such as the form preparation paradigm). It will be also important to systematically examine how the development of phoneme-size units will affect various aspects of word processing in the L2 language (e.g., the ability to articulate a cluster of consonants without the vowel insertion, ability to manipulate phonemes, and so on).

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

This research was supported by a Grant-in-Aid for Young Scientists (B) and Grant-in-Aid for Scientific Research (C) to MN. RV is supported by a Grant-in-Aid for Research Activity Start up (15H06687). We would like to thank Hinako Masuda for carrying out the acoustic analysis presented at **Figure 1**.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00175>

REFERENCES

- Alario, F.-X., Goslin, J., Michel, V., and Laganaro, M. (2010). The functional origin of foreign accent: evidence from the syllable frequency effect in bilingual speakers. *Psychol. Sci.* 21, 15–20. doi: 10.1177/0956797609354725
- Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge, MA: Cambridge University Press.
- Bates, D., Maechler, M., and Dai, B. (2008). *lme4: Linear Mixed-Effects Models Using Eigen and S4* [Computer Software Manual]. Available at: <http://lme4.r-forge.r-project.org/>
- Box, G. E. P., and Cox, D. R. (1964). An analysis of transformations (with discussion). *J. R. Stat. Soc. B* 26, 211–252.
- Broselow, E., and Kang, Y. (2013). “Second language phonology and speech,” in *The Cambridge Handbook of Second Language Acquisition*, eds J. Herschensohn and M. Young-Scholten (Cambridge, MA: Cambridge University Press), 529–554.
- Broselow, E., and Park, H. B. (1995). “Mora conservation in second language prosody,” in *Phonological Acquisition and Phonological Theory*, ed. J. Archibald (Hillsdale, MI: Lawrence Erlbaum), 151–168.
- Brysbaert, M., and New, B. (2009). Moving beyond Kuæera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990. doi: 10.3758/BRM.41.4.977
- Chen, J.-Y., Chen, T.-M., and Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *J. Mem. Lang.* 46, 751–781. doi: 10.1006/jmla.2001.2825
- Cutler, A., and Otake, T. (1994). Mora or phonemes? Further evidence for language-specific listening. *J. Mem. Lang.* 33, 824–844. doi: 10.1006/jmla.1994.1039
- Dehaene-Lambertz, G., Dupoux, E., and Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *J. Cogn. Neurosci.* 12, 635–647. doi: 10.1162/089892900562390
- Dupoux, E., Pallier, C., Kakehi, K., and Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Lang. Cogn. Process.* 16, 491–505. doi: 10.1080/01690960143000191
- Flæge, J. M. (1988). Factors affecting degree of perceived foreign accent in English sentences. *J. Acoust. Soc. Am.* 84, 70–79. doi: 10.1121/1.396876
- Forster, K. I., and Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 680–698.
- Forster, K. I., and Davis, C. (1991). The density constraint on form-priming in the naming task: interference effects from a masked prime. *J. Mem. Lang.* 30, 1–25. doi: 10.1016/0749-596X(91)90008-8
- Forster, K. I., and Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35, 116–124. doi: 10.3758/BF03195503
- Forster, K. I., Mohan, K., and Hector, J. (2003). “The mechanics of masked priming,” in *Masked Priming: The State of the Art*, eds S. Kinoshita and S. J. Lupker (New York, NY: Psychology Press), 3–37.
- Gomez, P., Perea, M., and Ratcliff, R. (2013). A diffusion model account of masked versus unmasked priming: are they qualitatively different? *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 1731–1740. doi: 10.1037/a0032333
- Grainger, J., and Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: cross-task comparisons. *J. Mem. Lang.* 35, 623–647. doi: 10.1006/jmla.1996.0033
- Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007
- Katada, F. (1990). On the representation of moras: evidence from a language game. *Linguist. Inq.* 21, 641–646.
- Kinoshita, S. (2000). The left-to-right nature of the masked onset effect in naming. *Psychon. Bull. Rev.* 7, 133–141. doi: 10.3758/BF03210732

- Kubozono, H. (1989). The mora and syllable structure in Japanese: evidence from speech errors. *Lang. Speech* 32, 249–278.
- Kučera, H., and Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.
- Kureta, Y., Fushimi, T., and Tatsumi, I. F. (2006). The functional unit of phonological encoding: evidence for moraic representation in native Japanese speakers. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 1102–1119.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2014). *LmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models (lmer Objects of lme4 Package)*. R Package Version 2.0-6. Available at: <http://CRAN.R-project.org/package=lmerTest>
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–75. doi: 10.1017/S0140525X99001776
- Malouf, T., and Kinoshita, S. (2007). Masked onset priming effect for high-frequency words: further support for the speech planning account. *Q. J. Exp. Psychol.* 60, 1156–1168.
- Meyer, A. S. (1990). The time course of phonological encoding in language production: the encoding of successive syllables of a word. *J. Mem. Lang.* 29, 524–545. doi: 10.1016/0749-596X(90)90050-A
- Meyer, A. S. (1991). The time course of phonological encoding in language production: phonological encoding inside a syllable. *J. Mem. Lang.* 30, 69–89. doi: 10.1016/0749-596X(91)90011-8
- Mousikou, B., Coltheart, M., Finkbeiner, M., and Saunders, S. (2010). Can the dual-route cascaded computational model of reading offer a valid account of the masked onset priming effect? *Q. J. Exp. Psychol.* 63, 984–1002. doi: 10.1080/17470210903156586
- O'Seaghdha, P. G. (2015). Across the great divide: proximate units at the lexical-phonological interface. *Jpn. Psychol. Res.* 57, 4–21. doi: 10.1111/jpr.12074
- O'Seaghdha, P. G., Chen, J.-Y., and Chen, T.-M. (2010). Proximate units in word production: phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition* 115, 282–302. doi: 10.1016/j.cognition.2010.01.001
- Piske, T., MacKay, I. R. A., and Flege, J. E. (2001). Factors affecting the degree of foreign accent in an L2: a review. *J. Phon.* 29, 191–215. doi: 10.1006/jpho.2001.0134
- Protopapas, A. (2007). Check vocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behav. Res. Methods* 39, 859–862. doi: 10.3758/BF03192979
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Mem. Cogn.* 32, 212–222. doi: 10.3758/BF03196853
- Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: from germanic languages to mandarin Chinese and Japanese. *Jpn. Psychol. Res.* 57, 22–37. doi: 10.1111/jpr.12050
- Sasanuma, S., Ito, H., Patterson, K., and Ito, T. (1996). Phonological alexia in Japanese: a case study. *Cogn. Neuropsychol.* 13, 823–848. doi: 10.1080/026432996381827
- Schiller, N. O. (2004). The onset effect in word naming. *J. Mem. Lang.* 50, 477–490. doi: 10.1016/j.jml.2004.02.004
- Schiller, N. O. (2008). The masked onset priming effect in picture naming. *Cognition* 106, 952–962. doi: 10.1016/j.cognition.2007.03.007
- Verdonschot, R. G., Kiyama, S., Tamaoka, K., Kinoshita, S., La Heij, W. L., and Schiller, N. O. (2011). The functional unit of Japanese word naming: evidence from masked priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1458–1473. doi: 10.1037/a0024491
- Verdonschot, R. G., Nakayama, M., Zhang, Q.-F., Tamaoka, K., and Schiller, N. O. (2013). The proximate phonological unit of chinese-english bilinguals: proficiency matters. *PLoS ONE* 8:e6154. doi: 10.1371/journal.pone.0061454
- Verdonschot, R. G., and Tamaoka, K. (2015). The production of speech sounds across languages (editorial). *Jpn. Psychol. Res.* 57, 1–3. doi: 10.1111/jpr.12073
- Warner, N., and Arai, T. (2001). Japanese mora-timing: a review. *Phonetica* 58, 1–25. doi: 10.1159/000028486
- You, W.-P., Zhang, Q.-F., and Verdonschot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PLoS ONE* 8:e46595. doi: 10.1371/journal.pone.0046595

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

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



A Principled Relation between Reading and Naming in Acquired and Developmental Anomia: Surface Dyslexia Following Impairment in the Phonological Output Lexicon

Aviah Gvion^{1,2,3} and Naama Friedmann^{1*}

¹ Language and Brain Lab, Tel Aviv University, Tel Aviv, Israel, ² Reuth Medical and Rehabilitation Center, Tel Aviv, Israel,

³ Communication Sciences and Disorders Department, Ono Academic College, Kiryat Ono, Israel

OPEN ACCESS

Edited by:

Simone Sulpizio,
University of Trento, Italy

Reviewed by:

Claudio Mulatti,
Università degli Studi di Padova, Italy
Maximiliano A. Wilson,
Université Laval, Canada

*Correspondence:

Naama Friedmann
naamafr@post.tau.ac.il

Specialty section:

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

Received: 12 November 2015

Accepted: 23 February 2016

Published: 30 March 2016

Citation:

Gvion A and Friedmann N (2016) A
Principled Relation between Reading
and Naming in Acquired and
Developmental Anomia: Surface
Dyslexia Following Impairment in the
Phonological Output Lexicon.
Front. Psychol. 7:340.
doi: 10.3389/fpsyg.2016.00340

Lexical retrieval and reading aloud are often viewed as two separate processes. However, they are not completely separate—they share components. This study assessed the effect of an impairment in a shared component, the phonological output lexicon, on lexical retrieval and on reading aloud. Because the phonological output lexicon is part of the lexical route for reading, individuals with an impairment in this lexicon may be forced to read aloud via the sublexical route and therefore show a reading pattern that is typical of surface dyslexia. To examine the effect of phonological output lexicon deficit on reading, we tested the reading of 16 Hebrew-speaking individuals with phonological output lexicon anomia, eight with acquired anomia following brain damage and eight with developmental anomia. We established that they had a phonological output lexicon deficit according to the types of errors and the effects on their naming in a picture naming task, and excluded other deficit loci in the lexical retrieval process according to a line of tests assessing their picture and word comprehension, word and non-word repetition, and phonological working memory. After we have established that the participants have a phonological output lexicon deficit, we tested their reading. To assess their reading and type of reading impairment, we tested their reading aloud, lexical decision, and written word comprehension. We found that all of the participants with phonological output lexicon impairment showed, in addition to anomia, also the typical surface dyslexia errors in reading aloud of irregular words, words with ambiguous conversion to phonemes, and potentiophones (words like “now” that, when read via the sublexical route, can be sounded out as another word, “know”). Importantly, the participants performed normally on pseudohomophone lexical decision and on homophone/potentiophone reading comprehension, indicating spared orthographic input lexicon and spared access to it and from it to lexical semantics. This pattern was shown both by the adults with acquired anomia and by the participants with developmental anomia. These results thus suggest a principled relation between anomia and dyslexia, and point to a distinct type of surface dyslexia. They further show the possibility of good comprehension of written words when the phonological output stages are impaired.

Keywords: aphasia, dyslexia, surface dyslexia, Hebrew, phonological output lexicon, naming

INTRODUCTION

Lexical retrieval and reading aloud are often viewed as two separate processes. We draw different models for them and refer to individuals with a lexical retrieval deficit as “anomic” and to those with a deficit in reading aloud as “dyslexic”. However, these processes are not completely separate—they share components. In this study we assessed the effect of an impairment in a shared component, the phonological output lexicon, on lexical retrieval and on reading aloud.

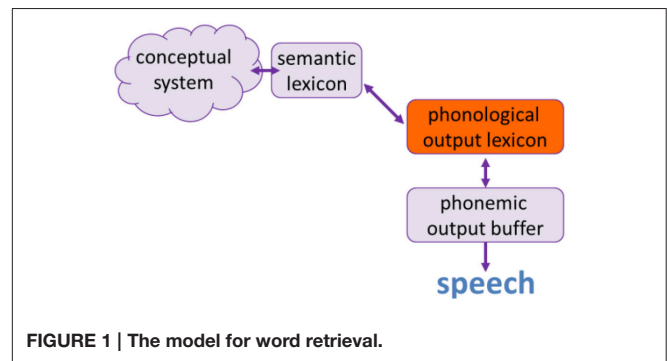
The Lexical Retrieval Process and Types of Anomia

The Lexical Retrieval Process

Lexical retrieval is a multi-component process, where each of the components and the connections between them can be selectively impaired and give rise to a different anomia (see **Figure 1**, which is a composite model based on Butterworth, 1989, 1992; Levelt, 1989, 1992; Nickels, 1997, 2002; Nickels and Howard, 2000; Friedmann et al., 2013). The first stage of the lexical retrieval process is the formation of a conceptual representation in the *conceptual system*, an a-modal representation that is still not formulated in words, which contains what the person knows about a concept, probably including its semantic properties, visual image, its function, and so on. Such concept can be created or activated from an idea someone has, or after identifying an object or event through the senses—in the case of neuropsychological assessments, usually identification of an object in a picture.

This non-lexical concept then activates a lexical-semantic representation in the *semantic lexicon* (Butterworth, 1989; Nickels, 2002; Friedmann and Biran, 2003; Biran and Friedmann, 2005, 2012)¹. The semantic lexicon is organized semantically and contains words and information about the meaning of words. Highly imageable (concrete) words are easier to access in the semantic lexicon than low-imageability (abstract) words (Nickels and Howard, 1994; Nickels, 1995; Howard and Gatehouse, 2006). Some approaches suggest that it does not contain words as other lexicons do, but is rather a “hub” that connects the conceptual system and the various lexicons (phonological and orthographic input and output lexicons).

The selected semantic lexical entry activates the lexical-phonological representation in the *phonological output lexicon*,



the protagonist of the current study². The representations in the phonological output lexicon contain information about the spoken form of the word, which includes its metrical information (number of syllables and stress pattern) and its segmental information (its phonemes—consonants and vowels, and their relative positions, Butterworth, 1992; Levelt, 1992). The phonological output lexicon is organized by word frequency, and as a result high-frequency words are accessed more readily than low-frequency ones. As for the representation of morphologically complex words (at least those with regular inflections), it seems that this lexicon only includes the stems of the word, namely, it includes “orange” but not “oranges”, “smile” but not “smiled”.

The activation is in turn transferred from the phonological output lexicon to the *phonological output buffer*, a post-lexical, sub-lexical short-term memory stage. The phonological output buffer is a phonological short-term store, which holds the phonological representation that arrives from the phonological lexicon or from a sublexical route (see Section The Word Reading Process below) until the word is produced (e.g., Garrett, 1976, 1992; Kempen and Huijbers, 1983; Patterson and Shewell, 1987; Dell, 1988; Butterworth, 1989, 1992; Levelt, 1989, 1992; Nickels, 1997). This buffer holds units of various sizes: phonemes as well as pre-assembled morphemes, number words, and possibly also function words (Dotan and Friedmann, 2015). The phonological output buffer is responsible for assembling words by inserting the phonemes into the metrical frame (e.g., Meyer, 1992; Shattuck-Hufnagel, 1992; Biran and Friedmann, 2005). According to some recent studies, it is also responsible for composing morphologically complex words from their morphemes, multi-digit number names from number words, and for incorporating function words within sentences (Kohn and Melvold, 2000; Dotan and Friedmann, 2015). Given that the phonological output buffer is a short term memory component, it is affected by the length of the phonemic string it holds (namely, the number of phonemes in a word, or the number of words in a multi-digit number)—longer strings that include more units are harder to maintain and produce, and strings that include more units than the buffer can hold are impossible to maintain and produce in full.

¹As surveyed by Nickels (2002), lexical retrieval models differ with respect to whether or not they assume a semantic lexicon that is separate from a semantic-conceptual system [as in models (e) and (f) in Nickels, 2002, p. 6]. Such separate semantic lexicon has been suggested by Butterworth (1989), and a similar idea can also be seen in Levelt (1989), where a distinction was suggested between a pre-verbal conceptual stage and a “mental lexicon”. We adopt such distinction on the basis of patients who show good conceptual abilities, as indicated by good comprehension of non-verbal concepts from pictures and gestures (for example, in picture association and odd one out picture tasks), who are impaired in the comprehension and production of the parallel spoken and written words (for example in word association and odd one out word tasks). Such patients were reported for example by Friedmann and Biran (2003) and Biran and Friedmann (2012).

²The entry in the semantic lexicon also activates the relevant entry in a syntactic lexicon (Biran and Friedmann, 2012), which we will not discuss in the current study.

Anomias: Impairments in the Lexical Retrieval Process

Anomia is a deficit in lexical retrieval, which can be acquired, i.e., occur following brain damage, or developmental—exist from birth. There exist several types of anomia, each resulting from a deficit in a different component of the lexical retrieval process (or from impaired connections between the components; Kay and Ellis, 1987; Butterworth, 1992; Nickels and Howard, 1994; Nickels, 1995, 1997, 2002; Miceli et al., 1996; Howard and Gatehouse, 2006).

A deficit in the conceptual system gives rise to an inability to name objects, but it has much wider repercussions: it also affects the ability to understand spoken words, written words, and even recognize and use objects, so it is quite clear that it should not be termed “anomia”. A deficit in the semantic lexicon, however, is a deficit that has to do with words. Because the semantic lexicon participates both in word comprehension and in word production, an anomia due to a deficit in the semantic lexicon affects both the comprehension and the production of words. Because the semantic lexicon participates in the semantics of written and spoken words, a semantic lexicon anomia affects the comprehension of both spoken and written words. Errors in naming in this type of anomia involve semantically-related word errors, as well as circumlocutions and definitions. Lexical retrieval in this anomia is affected by word imageability. Because it is a deficit in verbal processing, non-verbal material is not impaired, so pictures are understood correctly even when they are not named well; because it is a lexical deficit, the processing of non-words is unimpaired, so both reading and repetition of non-words are normal.

Phonological-lexicon anomia, an anomia that results from a deficit in the phonological output lexicon, affects the production of words, keeping the comprehension of words intact. Individuals with phonological-lexicon anomia make phonological- as well as semantic errors in production. When they produce semantic errors, they often comment that this is not exactly the word they were looking for. Because the phonological output lexicon is organized by frequency, these individuals show a frequency effect on naming. Given that the deficit is lexical, their non-word processing is normal.

Finally, a deficit in the phonological output buffer causes difficulties in word and non-word production. Errors in words and non-words are phonological; in morphologically complex words, phonological errors occur in the stems of the words, whereas the inflectional and derivational morphemes exhibit whole-morpheme substitutions and omissions. Number words in multi-digit numbers are omitted or substituted with other number words. Because the phonological output buffer is a short-term memory component, it is affected by length: stimuli with more phonemes induce more errors than shorter stimuli. Because the deficit is post the semantic stages, comprehension is intact, and no semantic errors occur. Non-words in phonological output buffer deficit are affected more gravely than words of the same length, because lexical feedback from the phonological output lexicon may support the activation of phonemes in real words but not of phonemes of non-words.

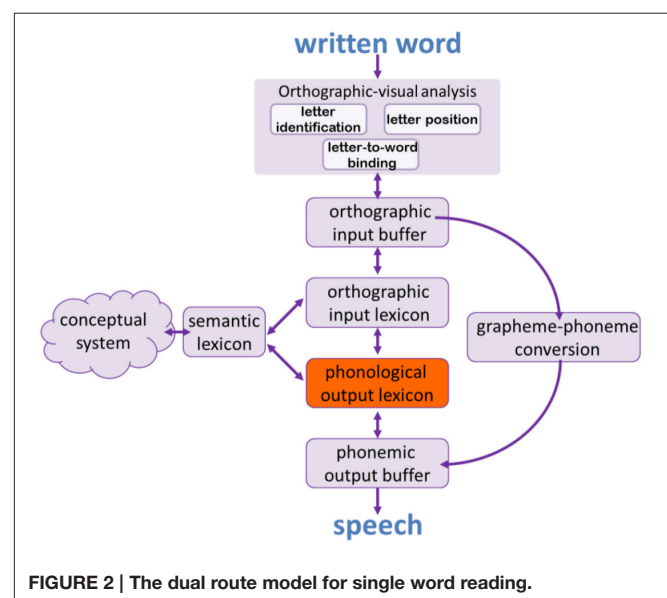
The Word Reading Process

The word reading process, like the naming process, is also a multi-staged process, in which each of the stages and components may be affected, giving rise to a different type of dyslexia. Figure 2 presents the dual route model for single word reading (cf. Ellis and Young, 1996; Coltheart et al., 2001; Friedmann and Coltheart, 2016). The first stage of word reading, orthographic-visual analysis, is responsible for three processes: abstract letter identification, encoding of relative positions of letters within words, and binding of letters to the words they appear in (Coltheart, 1981; Ellis et al., 1987; Humphreys et al., 1990; Ellis, 1993; Peressotti and Grainger, 1995; Ellis and Young, 1996; Friedmann and Coltheart, 2016). The information from the orthographic-visual analyzer is then held in an orthographic input buffer until it flows in two routes: the lexical route, and the sublexical one.

The lexical route, which includes the orthographic input lexicon and the phonological output lexicon, allows for the accurate reading of words that the reader already knows. The orthographic input lexicon holds the orthographic information about the written form of words we know, and the phonological output lexicon, as we described above, holds the phonological information about the sounds of the spoken words we know. The direct connection between these two lexicons in the lexical route allows for a rapid and accurate conversion of a written word to its phonological form. The lexical route has another branch, which connects the orthographic input lexicon and the semantic lexicon and allows for the comprehension of written words.

The sublexical route allows for reading of new words and nonwords via the conversion of graphemes (letters or groups of letters) into phonemes. This route is typically slower and less efficient than the lexical route, and is less accurate in reading words that do not follow the grapheme-to-phoneme conversion rules.

Importantly, a look at the dual route model depicted in Figure 2 shows that the lexical retrieval process that we have



described in the previous section (and in **Figure 1**) is actually part of the reading route: it includes all the stages between the conceptual system and the phonological output buffer (Friedmann et al., 2013). One of these shared components, the phonological output lexicon, are the topic of the current study. The phonological output lexicon is a part of the lexical retrieval process and of the lexical route for reading. As a result we expect that when the phonological output lexicon is impaired, not only lexical retrieval would be affected, but also reading via the lexical route.

Surface Dyslexia

A deficit in the lexical route is called “surface dyslexia” (Marshall and Newcombe, 1973; Coltheart et al., 1983; Newcombe and Marshall, 1985; Coltheart and Funnell, 1987; Howard and Franklin, 1987; Coltheart and Byng, 1989; Castles and Coltheart, 1993, 1996; Temple, 1997; Ellis et al., 2000; Masterson, 2000; Judica et al., 2002; Ferreres et al., 2005; Castles et al., 2006; Friedmann and Lukov, 2008, 2011). Because readers with surface dyslexia cannot read via the lexical route, they are forced to read words via the sublexical route, as if these were new words. Reading words via the sublexical route is slower than reading via the direct lexical route, and, importantly, such reading also affects reading accuracy. Some words do not obey the grapheme-to-phoneme conversion rules (e.g., words with silent letters such as *talk*, *walk*, *often*, or words in which the accurate conversion to a phoneme is the less common one, as in *door*, *have*). Other words include letters and letter sequences that have several options for conversion to phoneme strings (e.g., the *ea* in *head*, the *g* in *general*, Schmalz et al., 2015). These words, when read via the sublexical route, may be read incorrectly due to conversion that obeys the grapheme-to-phoneme conversion rules but is not appropriate for the target word. These errors are called “regularization errors”. A special group of irregular and ambiguous-conversion words are the *potentiophones*. These are words that, when read via grapheme-to-phoneme conversion, yield other existing words (Friedmann and Lukov, 2008). Examples for potentiophones in English are *move*, which can be sounded out via the sublexical route as “mauve”, *none*, which can be read sounding like “known”, and *phase*, which may be read like “face”. These words are especially challenging for individuals with surface dyslexia because they are not ruled out as non-words and even get feedback from the phonological output lexicon.

Regular words, in which grapheme-to-phoneme conversion rules create the correct reading, are read accurately in surface dyslexia, because they can be read correctly via the sublexical route. In addition, non-words are read correctly, because they do not need the lexical route.

Types of Surface Dyslexia

Surface dyslexia is thus defined as a deficit in the lexical route and there exist different types of surface dyslexia, depending on the component of the lexical route that is impaired (Coltheart and Funnell, 1987; Friedmann and Lukov, 2008). Deficits in each of the components of the lexical route result in reading aloud via the sublexical route and hence in inaccurate

and slow reading aloud. The difference between the different types of surface dyslexia relates to the different patterns with respect to lexical decision and written word comprehension (Friedmann and Lukov, 2008). A deficit in the orthographic input lexicon affects not only reading aloud but also lexical decision (of pseudohomophones like *klood*, *cranbery*, and *phun*) and comprehension of homophones (*aloud*, *bear*, *which*, *cite*). When the deficit in the lexical route spares the orthographic lexicon, lexical decision would be intact. When the orthographic input lexicon, the semantic lexicon, and the connection between them are intact, comprehension of homophones should also be intact.

Surface dyslexia as a result of a deficit in the orthographic input lexicon was reported in cases of acquired dyslexia (Coltheart and Funnell, 1987; Howard and Franklin, 1987; Coltheart and Byng, 1989; Weekes and Coltheart, 1996), and developmental dyslexia (Friedmann and Lukov, 2008). Additional cases of surface dyslexia that can be ascribed to the input orthographic lexicon are JC and MS, reported by Marshall and Newcombe (Marshall and Newcombe, 1973; Newcombe and Marshall, 1981, 1984, 1985). Friedmann and Lukov (2008) reported on three cases of developmental surface dyslexia as a result of an impairment to the connections of the orthographic input lexicon to the phonological output lexicon and to the semantic lexicon, and six cases with developmental impairment between the orthographic input lexicon and the phonological output lexicon.

Two interesting cases are described of people who showed surface dyslexia that can be ascribed to an impairment at the phonological output lexicon³. EE, the patient described by Coltheart (1982) and Howard and Franklin (1987), showed impaired naming alongside good semantic abilities (good comprehension of pictures and relatively good comprehension of auditorily presented words) and good phonological output buffer abilities (good non-word repetition and no length effect), suggesting a deficit in the phonological output lexicon or in the connection between the semantic lexicon and the phonological output lexicon. In reading, EE showed surface dyslexia in reading aloud. The fact that he also performed poorly in input tasks involving written pseudohomophones and homophones, indicated that his surface dyslexia resulted (also) from a deficit in the input stages of reading.

EST, the patient described by Kay and Patterson in the seminal Surface Dyslexia book (Kay and Patterson, 1985) and by Kay and Ellis (1987), showed a naming pattern characterized by error types and effects on naming that are typical to impaired activation of the phonological output lexicon, alongside surface dyslexia in reading aloud: better reading of regular than irregular words, and regularization errors. His orthographic lexical judgment of pseudohomophones and his comprehension of irregular words

³Another patient that was reported as showing “common mechanisms in dysnomia and post-semantic surface dyslexia” is RF, reported by Margolin et al. (1985). However, RF also was not a clear case of phonological output lexicon anomia and surface dyslexia: her error pattern in naming included only definitions but no phonological errors, so her naming deficit may have not been in the phonological output lexicon, and only 38% of her reading errors were surface errors, and she demonstrated no significant difference in reading regular and irregular words.

were better than his oral reading, yet not normal. However, he also made phonological errors in non-word repetition, especially for longer non-words, and his comprehension of abstract words was impaired, suggesting that his deficit, too, was not purely at the phonological output lexicon, and may have involved the phonological output buffer and the lexical-semantic system or the access to it as well.

Kay and Ellis (1987) noticed a very interesting pattern in EST's reading: in his first reading of a word, he initially tried to read the words via the lexical route, and made phonological errors, and then moved to the sublexical route, with the result of regularization errors. It might be that his initial phonological errors resulted from a further phonological output buffer deficit.

In the current study we further explore the effect of a deficit in the phonological output lexicon on reading, with individuals with acquired or developmental anomia whose phonological output lexicon deficit was selective, and for many of whom the input reading stages and the phonological output buffer were not impaired.

The Hebrew Orthography and Its Interaction with Surface Dyslexia

Hebrew, the language tested in the current study, is a Semitic language that is read from right to left. Words in Hebrew are often morphologically complex, derived from a three consonant root inserted in a derivational template and inflected for inflectional morphology. Several properties of Hebrew orthography make surface dyslexia especially noticeable and easy to detect. Vowels in the middle of the word are not consistently represented. In addition, each of the vowel letters can also be read as a consonant. Four consonant letters have ambiguous conversion to sound, and may be converted to either of two consonants. Additionally, nine phonemes can be converted to one of two or three different letters, and the stress position is lexical and not represented orthographically. These characteristics of the Hebrew orthography cause reading via the sublexical route to be error-prone, and surface dyslexia very easy to detect. In fact, there is no regular word in Hebrew, i.e., there is no word that can be unambiguously converted to a phonological string. In addition, Hebrew has many potentiophones, which makes Hebrew reading even more sensitive to surface dyslexia—for many Hebrew words, reading via the sublexical route results in another existing word so the reader cannot rule out the erroneous response based on lexicality considerations (Friedmann and Lukov, 2008, 2011). (A script that includes diacritics, *nikud*, which disambiguates most of the ambiguities in letter conversion exists but it is only used by young children in the beginning stages of reading acquisition, and in prayers and poetry).

PARTICIPANTS

The participants were 16 adults and adolescents who were included in the study on the basis of their phonological output lexicon anomia. Eight of them had acquired anomia following brain damage and eight had developmental anomia. The acquired anomia group included two women and six men,

and the developmental anomia group included two men and six adolescents (three girls and three boys). Background information about each of the participants is summarized in **Table 1**.

Four of the participants with developmental anomia were of the same family: TAF was the father of AFI, MAD, and ARO (an additional daughter had only a mild anomia and was therefore not included in the study). All participants, including the four who immigrated to Israel after the beginning of elementary school, reported that Hebrew was the main language they used for reading and writing.

The participants were selected from a pool of individuals with acquired or developmental language deficits who were complaining of naming difficulties and were referred to a rehabilitation center in central Israel or to our Language and Brain Lab at Tel Aviv University.

GENERAL PROCEDURE

Each participant was tested individually in a quiet room. During the testing sessions, the experimenter wrote down every response that differed from the target. All the sessions were audio-recorded and two judges listened to the recordings after the sessions, and checked the transcriptions from the session against the recordings, completing and correcting them when necessary. The pictures and the written stimuli of the various tests were presented to each participant over the desk, printed on a white page. In the oral reading tasks, the participant was requested to read aloud as accurately as possible; in the lexical decision and comprehension tasks (see Section Reading Tests), the participant was requested to perform the task without reading aloud. According to the availability for testing of each of the participants, some of them were tested with more tests from the battery, and some with fewer tests—the results of each test for each participant appear in the tables below. No time limit was imposed during testing, and no response-contingent feedback was given by the experimenter, only general encouragement. The participants were told that whenever they needed a break they could stop the session or take a break.

Data Analysis

To compare the performance of each experimental participant to her/his age-matched control group, we used Crawford and Howell's (1998) *t*-test, and reported that an individual performed significantly below the control group when $p < 0.05$ in this test⁴. The children were compared to control groups not only by age but also by grade level, and all the adult participants in all the control groups had 12 years education and above, as did the anomic participants. The effects on word retrieval (length,

⁴The *t*-test suggested by Crawford and Howell (1998, see also Crawford and Garthwaite, 2002, 2012) allows neuropsychological studies to test the difference between a single case and a control group, and determine whether the performance of a person is significantly poorer than that of a sample of matched healthy control participants. The Crawford and Howell *t*-test takes care of the risk for inflated Type I errors (leading researchers to incorrectly conclude that a patient shows abnormal performance), by treating the mean and SD of the control group as statistics, namely, as belonging to a control sample rather than as known population means and SD.

TABLE 1 | Background description of the participants.

Participant	Age (years)	Gender	Hand	Hebrew	Education (years)	Etiology	Lesion localization	Time post onset (months)
ACQUIRED ANOMIA								
DAN	42	M	Right	Native	12	Anoxic brain damage	Small hypodense focus in the left caudate head	2
YOS	74	M	Right	63 years	15	Hydrocephalus		2
ZAB	77	M	Right	64 years	12	Stroke	Basal ganglia infarct	4
BAR	57	F	Right	39 years	15	Stroke	–	2
LER	78	M	Right	native	12	4th stroke	Left parieto-occipital stroke	3
ARI	67	M	Right	Native	12	Stroke	Left fronto-temporal-occipital infarct	48
NAV	72	F	Right	Native	15	Stroke	Infarct in the thalamus and left internal capsule	1
DOR	55	M	Right	Native	17	Encephalitis	Left temporal lesion	17
DEVELOPMENTAL ANOMIA								
Participant	Age (years; months)	Gender	Hand	Hebrew	Education (current grade/years)			
LEO	28;8	M		24 years	16 (MA student; a teacher)			
TAF	49;3	M		20 years	17 (MA)			
AFI	16;6	F	left	Native	9th grader			
ARO	10;4	M	left	Native	4th grader			
MAD	12;6	M		Native	7th grader			
SAN	13;1	F	left	Native	7th grader			
NIV	12;8	M	left	Native	7th grader			
SHL	12;9	F	right	Native	7th grader			

frequency) were calculated as the point biserial correlation between the word property and the success in producing the target word. An alpha level of 0.05 was used in all analyses.

ASSESSMENT OF LEXICAL RETRIEVAL: TESTING TO ESTABLISH PHONOLOGICAL OUTPUT LEXICON DEFICIT FOR INCLUSION IN THE STUDY

Tests Establishing a Phonological Output Lexicon Deficit

In order to establish that a participant had a phonological output lexicon deficit and could be included in our study, we started by testing picture naming for all participants. Those who showed impaired naming that could result from a phonological output lexicon deficit received additional tests to assess the exact locus of their impairment in the lexical retrieval process. These tests included conceptual tests; repetition of words, pseudo-words, and morphologically complex words; comprehension of heard and written words; reading aloud of Arabic numbers; and phonological short-term memory tests.

Naming Task

Picture naming was assessed using the *SHEMESH* test (Biran and Friedmann, 2005), which includes 100 color pictures of objects of various semantic categories. The target nouns were feminine and masculine nouns, morphologically simple and complex, with regular and irregular gender morphology, 1–4 syllable long, 3–10 phonemes, with ultimate and penultimate stress and with various first phonemes.

The frequency of the target words, judged by 75 Hebrew-speaking participants with no language deficits, ranged from 2.39 to 6.84 on a scale of 1–7 ($M = 4.90$, $SD = 1.09$). The performance of Hebrew speakers without a language deficit in this test is very high (average 95.6% correct, $SD = 4.2\%$, for 67 control subjects aged 50–80; average 98.7% correct, $SD = 1.7\%$, for 87 control subjects aged 20–40; and 94.1% correct, $SD = 2.3\%$, for 35 control subjects aged 12–14, Biran and Friedmann, 2004, 2005).

Additional Tasks to Establish a Phonological Output Lexicon Deficit and Exclude Impairments at Other Levels

A naming deficit that results from a deficit in the phonological output lexicon should not affect semantic and conceptual abilities, nor should it impair non-word processing. We thus tested these abilities, using several additional tasks.

The *conceptual system* was tested using a **picture association test** (MA KASHUR pictures, Biran and Friedmann, 2007). This task includes 35 triads of pictures, a target object presented at the top (e.g., *cow*) and two pictures at the bottom, one semantically related to the target picture (e.g., *milk*) and one unrelated but from the same category or associated with the other picture on the bottom (e.g., *Coca-Cola*). The participants are requested to choose the picture that is semantically related to the top picture.

The *semantic lexicon* and the access to it from written words were tested using the verbal counterpart of the picture association

task, the **written word association test** (MA KASHUR words, Biran and Friedmann, 2007)⁵. This task includes 35 triads of written words. Of these, 25 are identical to 25 of the pictorial triads, and 10 triads include abstract terms (e.g., honesty–truth/lie).

An additional task that we used to examine the semantic lexicon was a **spoken word-to-picture matching task** from the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA 47, Kay et al., 1992; Hebrew version Gil and Edelstein, 2001). This test consists of 40 groups of five pictures including a target word (e.g., *a dog*) and four close and distant semantic distracters (e.g., a cat, a giraffe, a rocking horse, and a kite, respectively). The participants are requested to select the picture that matches the word they heard.

The *phonological output buffer* was assessed using a **non-word repetition test** (BLIP, Friedmann, 2003). The participants were requested to repeat 48 non-words that the experimenter said. The test includes 24 easy non-words of 2, 3, and 4 CV syllables (8 of each length), and 24 phonologically complex non-words (of 2, 3, and 4 syllables) with clusters in various word position or with phonological feature similarity.

A phonological output buffer impairment also affects the production of morphologically complex words and multi-digit numbers (Dotan and Friedmann, 2015). Therefore, as another tool to assess a phonological output buffer impairment we administered a test of **repetition of morphologically complex words** (the MURKAMOT test, from the Buffy battery, Friedmann, 2006). This test consists of 36 morphologically complex words, 24 of the words included a stem/root and inflectional or derivational morphemes (half with 1 morpheme and half with 2), and 12 were long morphologically-simple words. (LER and ZAB were tested using a short version of the non-word and morphological complex repetition tests that included only 10 items each).

Multi-digit number processing was tested using a task of **oral reading of multi-digit Arabic numbers**, which included 60 numbers pf 2–5 digits, 15 numbers of each length.

Additional tests for the phonological output buffer included phonological STM tasks from the *FriGvi* battery (Friedmann and Gvion, 2002; Gvion and Friedmann, 2012). These included a **basic word recall span** test that tests the recall of sequences of 2–7 phonologically different two-syllable words (five word sequences in each length); a **long word recall span** test, with sequences of 2–7, phonologically different four-syllable words (five word sequences in each length); and a **non-word recall span** testing sequences of 2–7, two-syllable non-words, constructed by changing a single consonant in real words (five non-word sequences in each length). To measure the participants' input span in a task that does not involve speech output, and allow for the comparison between span tasks with and without overt speech in order to evaluate the input and output buffers separately, we administered to some of the participants a recognition STM task, the **matching word order span**. In this

⁵Good performance in the homophone-potentiophone written comprehension test (described in Section Reading Tasks without Oral Output below) is also indicative of spared lexical semantic and conceptual abilities.

task, the participants heard, in each item, two sequences of 2–7 words containing the same words (2-syllable phonologically dissimilar words, 10 items per length) either in the same or in a different order, and were asked to judge whether the order of the items in the two lists was the same. On the non-identical pairs, the two lists differed in the order of two adjacent words. The span level is defined as the maximal level at which the participant performed correctly on at least 7 of 10 items.

Results: Lexical Retrieval Performance and Locus of Deficit

Acquired Anomia

The performance of the individuals with acquired anomia on the picture-naming test is summarized in **Table 2**. As demonstrated in the table, the performance of each of them was significantly below that of their age-matched control groups at a level of $p < 0.0001$. The participants named correctly 21%–81% of the pictures, with an average of 53.8% ($SD = 22\%$) correct.

To examine the locus of the deficit in the lexical retrieval process of each of the participants, and to establish whether they have a phonological output lexicon impairment, we used three criteria: Error types, effects on naming, and performance on the other, semantic and phonological tasks.

Error types

As shown on **Table 2**, the error pattern of each of the participants was the one typical of phonological output lexical impairment:

hesitations and long response latencies ($M = 29.4\%$, $SD = 17\%$), relevant paraphrases ($M = 20.6\%$, $SD = 18\%$), phonological approximations ($M = 15.3\%$, $SD = 17\%$), phonologically-related words and non-words ($M = 7.5\%$, $SD = 12\%$), and semantically-related words, usually followed by self-correction attempts ($M = 3.1\%$, $SD = 5\%$). Other types of errors were relatively few.

Effects on naming

As shown on **Table 2**, the typical **frequency effect** on naming, with higher frequency words named better than lower frequency ones, was significant for five of the participants (DAN, ZAB, BAR, ARI, DOR), and marginally significant for the other three (YOS, LER, NAV). LER also showed a significant length effect, and DAN and ARI showed a marginally significant one.

Performance in other tasks

The performance of each of the individuals with acquired anomia in the additional semantic and phonological tests is summarized on **Table 3**. The tests that assessed their semantic-conceptual abilities, which included picture-picture, word-picture, and word-word matching tasks, indicated that the lexical-semantic and conceptual levels of each of the participants are preserved. Seven of the eight participants were tested using the word-word association test, which assessed the lexical-semantic level (as well as the conceptual system), where they all performed at least 95% correct. One patient was tested only using the picture association task, on which he also performed 95% correct.

TABLE 2 | Picture naming: %correct, error types, and effects—acquired anomia.

	DAN	YOS	ZAB	BAR	LER	ARI	NAV	DOR
% Correct naming	68%***	81%***	56%***	73%***	41%***	21%***	62%***	28%***
Error types								
Phonologically related non-word	27%		2%	3%	43%	10%		
Phonologically related word	2%				18%	10%	5%	
Phonological approximations	18%		7%		3%	45%	15%	34%
Hesitations, long latency	27%	35%	57%	35%	11%	1%	42%	28%
Paraphrases, definitions	9%	61%	12%	24%	3%	8%	25%	23%
Semantically related word	13%		5%		5%	1%		1%
Morphological error					1%		2%	2%
No response/don't know	2%		7%			7%		1%
Naming in another language				29%				
Related gesture			10%	6%	4%			9%
Superordinate category							2%	2%
Unrelated syllables						14%		
Neologism	2%				1%	1%	2%	
Perseveration		4%		3%	11%	3%	7%	
Frequency effect	$r = 0.27$, $p = 0.003$	$r = 0.15$, $p = 0.07$	$r = 0.20$, $p = 0.02$	$r = 0.28$, $p = 0.002$	$r = 0.15$, $p = 0.08$	$r = 0.23$, $p = 0.01$	$r = 0.15$, $p = 0.07$	$r = 0.20$, $p = 0.02$
Length effect	$r = -0.17$, $p = 0.05$	NO	NO	NO	$r = -0.19$, $p = 0.03$	$r = -0.15$, $p = 0.07$	NO	NO

The percentages in the error type cells represent percentage of the total number of errors the participant made.

***Comparison of percentage correct of naming of each participant to his/her matched control group, $p < 0.001$.

TABLE 3 | %Correct performance of the individuals with acquired anomia in tasks that involve conceptual, lexical semantics, and phonemic output buffer.

Tasks	DAN	YOS	ZAB	BAR	LER	ARI	NAV	DOR
CONCEPTUAL AND SEMANTIC TASKS								
Picture association	95%		100%			97%		
Written word association ^a		100%	100%	95%*	100%	100%	100%	98%
Word to picture matching				95%	88%			
PHONOLOGICAL BUFFER TASKS								
Non-word repetition				100%	30%*	29%*	100%	100%
Morph word repetition			90%		90%		100%	
Word repetition	66%*						100%	
Arabic number reading			97%		81%*	85%*		
Basic word span	2*	4				2*		4
Non-word span						1.5*		2*
Matching word order span	4*		5			7		

^a The scores for ZAB and ARI refer to their performance in the written word association test (Biran and Friedmann, 2007). For the other participants the scores refer to the homophone-potentiophone written comprehension test.

*Significantly below the control group, $p < 0.05$.

In the tests that assessed their phonological (input and) output buffer, five of participants with acquired anomia (YOS, ZAB, BAR, NAV, and DOR) showed good performance, and three (DAN, LER, and ARI) showed indications of an additional impairment in the input and/or output phonological buffers. On the basis of their performance in the span tasks, ARI probably had a deficit in the phonological output buffer, as his recognition span was within matched controls range, whereas DAN's limited input span suggests that he also has a phonological input buffer deficit, which may have contributed to his difficulty in word repetition. His phonological output buffer was also impaired, as indicated by the length effect he demonstrated in naming (Table 2). (DOR's non-word span was 0.5 words below the normal range, but given his 100% correct repetition in the difficult non-word repetition task, and his normal word span, we considered his phonological buffers unimpaired).

Thus, the error pattern, which is typical of a deficit at the phonological output lexicon, as well as the frequency effect and the performance on semantic and phonological tasks, indicate that the anomia of all eight participants with acquired anomia resulted from a deficit in the phonological output lexicon. Three of them (DAN, LER, and ARI) probably had an additional deficit in the phonological output buffer. (We conclude that they had a phonological output buffer deficit in addition to a phonological output lexicon and not only a phonological output buffer deficit on the basis of the frequency effect they showed in naming, as well as on the basis of the semantic errors that they made in picture naming, which cannot be explained by a phonological output buffer, and cannot be ascribed to impaired semantic-conceptual system, because they all demonstrated good lexical-semantic and conceptual abilities).

Developmental Anomia

The results of the naming test of the individuals with developmental impairments, including the rate of correct responses, types of errors, and effects on naming, are summarized in Table 4. Each of the participants performed significantly below

her/his age-matched group in the naming test, $p < 0.001$. They named between 68% and 85% of the pictures correctly.

Error types

Similarly to the participants with acquired anomia, the types of errors that the participants with developmental anomia made were the typical errors evinced in phonological output lexicon anomia: hesitations and long response latencies ($M = 39.4\%$, $SD = 14.3\%$), no responses or "don't know" responses ($M = 16.2\%$, $SD = 19.8\%$), semantically related words, usually followed by self-correction attempts ($M = 12.7\%$, $SD = 16.5\%$), naming in another language ($M = 11.4\%$, $SD = 15.6\%$), relevant definitions and circumlocutions ($M = 5\%$, $SD = 4.7\%$), related gestures ($M = 4.5\%$, $SD = 10.7\%$), and phonologically-related words and non-words ($M = 3\%$, $SD = 5\%$). Other types of errors were relatively few.

Effects on naming

Six of the participants (TAF, AFI, ARO, MAD, SAN, and NIV) manifested the typical frequency effect ($p \leq 0.04$). NIV also showed a length effect ($p = 0.04$). Two other participants (LEO, SHL) were not affected by either frequency or length effects.

Performance in other tasks

The performance of each of the individuals with developmental anomia in the semantic and phonological tests is summarized on Table 5. The good performance of the developmental anomic participants on the conceptual and lexical-semantic tests indicates that their semantic lexicon and conceptual system are preserved.

All but two of the developmental anomic participants performed well on the non-word repetition task, indicating well-functioning input and output phonological buffers. Two girls (SAN and SHL), however, performed poorly on repetition of non-words. SHL showed impaired performance on the input span task, and did not show a length effect in naming, so her poor non-word repetition may be attributed to a limited phonological input buffer, in addition to phonological output

TABLE 4 | Picture naming test: Correct performance, error types, and effects—developmental anomia.

	LEO	TAF	AFI	ARO	MAD	SAN	NIV	SHL
% Correct	80%***	76%***	85%***	71%***	68%***	85%***	80%***	84%***
Error types								
Phonologically related non-words		5%			3%	18%	6%	
Phonologically related words					3%	12%		8%
Phonological approximations					5%	12%		
Hesitations, long latency	44%	45%	33%	23%	57%	40%	17%	54%
Paraphrases, definitions	13%		6%	8%		6%		8%
Semantically related word	4%		17%		5%	12%	50%	15%
Morphological error		14%			3%			
No response/don't know		36%		50%			27%	15%
Naming in another language	4%		38%	19%	24%			
Related gesture	31%		6%					
Superordinate category	4%							
Frequency effect	NO	$r = 0.34$, $p = 0.0003$	$r = 0.4$, $p = 0.001$	$r = 0.31$, $p < 0.0001$	$r = 0.3$, $p = 0.003$	$r = 0.2$, $p = 0.04$	$r = 0.25$, $p = 0.006$	NO
Length effect	NO	NO	NO	NO	NO	NO	$r = -0.17$, $p = 0.04$	NO

The percentages in the error type cells represent percentage of the total number of errors the participant made.

***Comparison of percentage correct of naming of each participant to his/her matched control group, $p < 0.001$.

TABLE 5 | %Correct performance in tasks of conceptual, lexical semantics and phonemic output buffer—developmental anomia.

Task	LEO	TAF	AFI	ARO	MAD	SAN	NIV	SHL
CONCEPTUAL AND SEMANTIC TASKS								
Picture association			100%	100%	100%			
Written word association ^a	98%	100%	100%	100%	100%	93%	95%	100%
Word-picture matching				100%	100%			
PHONOLOGICAL BUFFER TASKS								
Non-word repetition	100%	92%	98%	96%	91%	58%*	90%	73%*
Morph word repetition				100%			94%	100%
Basic word span	6.5		5	4.5		3*	3*	2.5*
Long word span			4	3.5	4			
Non-word span			3			0*		2*
Matching order word span			7			4		3*

*Significantly worse than age-matched control group ($p < 0.05$).

^aThe scores for ARO and MAD refer to their performance in the written word association test (Biran and Friedmann, 2007). For the other participants the scores refer to the homophone-potentiophone written comprehension test.

lexicon deficit. SAN's poor non-word repetition is a bit more difficult to interpret, as her input span was within the normal range for her age, indicating intact phonological input buffer, but she also did not show length effect in naming, which casts doubt on a deficit in the phonological output buffer.

The naming of one participant (NIV) was affected not only by frequency effect but also by length effect. Length effect could indicate a phonological output buffer impairment, but given

that his repetition of non-words and morphologically complex words was relatively spared, it seems that he does not have a phonological output buffer deficit on top of his phonological output lexicon impairment.

Thus, based on typical errors, effects on naming, and the performance in semantic and phonological tasks, like the participants with acquired anomia, all eight participants with developmental anomia have a deficit in the phonological output lexicon. Although two of the developmental anomic participants did not manifest the expected frequency effect in naming, their performance in other tasks and the types of errors they made in naming imply that their deficit is at the phonological output lexicon. Two of the participants may have also had a phonological input or output buffer deficit, in addition to their phonological output lexicon deficit.

HOW DOES A DEFICIT AT THE PHONOLOGICAL OUTPUT LEXICON AFFECT READING?

In order to test our main research question for this study, the way a deficit in the phonological output lexicon affects reading, we assessed the participants' oral reading, as well as their performance in reading tasks that do not involve speech output and hence, do not involve the phonological output lexicon. For assessing oral reading, we administered a word reading aloud test that includes single words of various types, including irregular and potentiophonic words, and an additional test of oral reading of potentiophones, which are particularly sensitive to surface dyslexia. To evaluate the earlier, input stages of reading through the lexical route—the orthographic input lexicon and

its connection to the semantic lexicon, we used an orthographic lexical decision task and a task that assessed the comprehension of written homophones and potentiophones.

Reading Tests

Oral Reading Tasks

The TILTAN oral reading screening test (Friedmann and Gvion, 2003)

The screening test served two purposes: to examine whether the participants had surface dyslexia, by assessing their oral reading of irregular and potentiophonic words, and to test whether they had any other types of dyslexia, apart from surface dyslexia.

The screening test includes 136 single Hebrew words that were constructed so that they are sensitive to the various types of dyslexia: 65 migratable words, to detect letter position dyslexia; All the words in the test are sensitive to left neglect dyslexia at the word level, as all the words in the list are such that neglect errors on their left side yield other words; 104 of the words are sensitive to right neglect, as neglect errors on their right side create other existing words; 89 abstract words, for identifying deep dyslexia; function words and morphologically complex words, for identifying deep dyslexia and phonological output buffer dyslexia; words with many orthographic neighbors for identifying visual dyslexia; and words for which migrations, substitutions, omissions, or additions of a vowel letter create other existing words for identifying vowel letter dyslexia (Khentov-Kraus and Friedmann, 2011).

Most importantly for our study, the test includes words for identifying surface dyslexia. In Hebrew, as we explained above (Section The Hebrew Orthography and Its Interaction with Surface Dyslexia), there are no words that can be read unambiguously and correctly through grapheme-to-phoneme conversion. Therefore, essentially all words in the screening test are sensitive to surface dyslexia⁶. The test also included 35 potentiophones, which are most sensitive to surface dyslexia, and 33 irregular words that are parallel to irregular words in English—words with silent letters or with ambiguous letters that are converted to the less common rendition of the letter.

Potentiophone reading test

Potentiophone reading test (also from the TILTAN battery, Friedmann and Gvion, 2003).

To assess directly the participants' ability to read via the lexical route, we tested their reading of the stimuli that are most sensitive to sublexical reading: potentiophones. The potentiophone test includes 78 potentiophonic words, 2–6 letters long ($M = 3.7$ letters, $SD = 0.8$).

Reading Tasks without Oral Output

We used two tasks to examine the way the participants process pseudohomophones, homophones, and potentiophones when they were requested to avoid oral production and hence. This allowed us to examine how they read when their impaired phonological output lexicon is not involved.

Written lexical decision (Friedmann and Lukov, 2008)

To assess the orthographic input lexicon and the access to it, we tested the participants' ability to decide whether a pseudohomophone is a word or not, using a visual-word recognition task of lexical decision, which proved sensitive to surface dyslexia with orthographic input lexicon deficit (Friedmann and Lukov, 2008). The test consisted of 68 pairs, each pair includes a correctly spelled word (*shoe*) and its pseudohomophone (*shoo*). Twelve of the words were irregular (including a silent letter or a letter that is the less frequent orthographic representation of the phoneme), and the parallel pseudohomophone was the regular spelling of the word (e.g., *school-scool*). The other pairs included words in which at least one phoneme can be ambiguously converted to a letter (e.g., *city-sity*). The participants were requested to circle the correctly spelled word. The control groups for this test included 148 adult participants aged 20–72, with 12 years education and above—like the anomic participants, and 201 children and adolescents in 4th–9th grade (see Table 7).

Written homophone-potentiophone comprehension (Friedmann and Lukov, 2008)

To examine the participants' access from the orthographic input lexicon to the semantic lexicon (as well as the status of the orthographic and semantic lexicons themselves), we tested the comprehension of homophones or potentiophones. The test consisted of 40 triads, each triad includes a target word (e.g., *pay*), and two additional words. One word was semantically related to the target word (*buy*), the other word was a homophone or a potentiophone of the related word (*bye*). Twenty of the target words were abstract (the target word was of low imageability), and 20 target words were concrete nouns or verbs. Each participant was requested to find the word that is semantically associated with the target word, and draw a line between them. This test, too, was used in previous surface dyslexia studies and proved sensitive to surface dyslexia with input deficit (Friedmann and Lukov, 2008). The control groups for this test included 141 adult participants aged 20–70, and 169 children and adolescents in 4th–9th grade (see Table 7).

Results: Oral Reading Tasks

Table 6 summarizes the performance of each of the participants in the oral reading tests. The results indicate that all the 16 anomic participants with a phonological output lexicon deficit had surface dyslexia in reading aloud—namely, their oral reading indicated that they were using the sublexical, rather than the lexical route for reading aloud.

All the participants, those with acquired anomia and those with developmental anomia, performed significantly below the age-matched control readers in reading aloud of the single words in the screening test and of the potentiophone word list, and each of them made significantly more surface errors than their age-matched peers (one participant, AFI, had significantly more surface errors than the control group only in the potentiophone list). Their surface errors were the errors we typically see in the reading aloud of Hebrew-readers with surface dyslexia: reading the target word in a way that is a plausible reading

⁶If the reader finds herself/himself wondering why we have not compared the reading of irregular and regular words, this is the reason: this is a luxury that only languages that do have regular words can afford.

TABLE 6 | Oral reading tasks: %correct and number of surface errors.

Participant	Single words TILTAN screening		Potentiophones	
	%Correct	Surface errors	%Correct	Surface errors
ACQUIRED ANOMIA				
DAN	63%***	16***	58%***	32***
YOS	92%***	8***	81%***	12***
ZAB	85%***	6***	72%***	17***
BAR	93%***	4**	82%***	14***
LER	55%***	12***	51%***	25***
ARI	74%***	25***	60%***	31***
NAV	94%***	8***	87%***	10**
DOR	88%***	12***	83%***	15***
DEVELOPMENTAL ANOMIA				
LEO	93%***	7***	78%***	17***
TAF	74%***	27***	51%***	37***
AFI	96%	2	79%*	16*
ARO	38%***	64***		
MAD	85%***	17***	53%***	36***
SAN	81%***	9***		
NIV	82%***	12***		
SHL	81%***	12***		
Control groups M (SD)				
Adults, 12 years education and above N = 372	98.4% (1.4)	1.1 (1.2)	94.7% (3.6)	3.7 (2.7)
7th grade control N = 26	96.0% (1.4)	3.2 (1.7)	90.4% (4.6)	7.1 (3.7)
5th grade control N = 14	96.4% (2.5)	2.7 (1.9)	88.8% (6.7)	8.4 (5.0)
4th grade control N = 20	90.6% (5.4)	7.2 (5.1)		

Significantly more errors than age-matched control group, * $p \leq 0.02$, ** $p \leq 0.01$, *** $p \leq 0.001$.

according to grapheme-to-phoneme conversion rules, including errors of stress position, as the stress in Hebrew is not marked lexically, errors of the choice of vowels that are not marked orthographically, reading silent letters, and converting of a grapheme that has several possible conversions to a phoneme that is a possible conversion but not the right one for the target word.

Importantly, most of their errors in reading aloud were surface errors, namely, errors that were phonologically acceptable conversions of the target words, but which indicated that the words were not read via the lexical route. The number of surface errors in the screening task and in the potentiophone task is presented in **Table 6**⁷.

The oral reading of YOS, ZAB, NAV, DOR, LEO, TAF, AFI, and MAD was selectively impaired, and the pattern was that of a pure surface dyslexia. Some other participants (mainly

DAN, LER, ARI, ARO, SAN, NIV, and SHL) showed a clear surface dyslexia but also made additional types of errors in oral reading, including letter migrations, substitutions, omissions, and additions, which resulted from their letter position dyslexia or attentional dyslexia (ARO, SAN, NIV, and SHL) or from a phonological output buffer deficit (DAN, LER, and ARI). See Appendix A in Supplementary Material for a detailed presentation of all errors types each of the participants made in each of the reading aloud tests, and for further assessment of the additional dyslexias of these seven patients.

Results: Input Reading Tasks

We have seen that, when asked to read aloud, all the participants with phonological lexicon impairment read via the sublexical route, and therefore show a pattern that is characteristic of surface dyslexia. Does it indeed result, as we have suggested, from their phonological output lexicon deficit, or do they have a deficit in the orthographic input lexicon, which causes their sublexical reading? We examined this question by assessing their reading in tests that did not involve oral production.

The results of these tests, summarized in **Table 7**, indicated that most of the participants performed very well and not differently from age- (and grade-) matched controls when the reading task did not involve output. All but the two youngest children performed at a level of 93% correct and above in both tasks⁸.

Thus, the performance of the participants with impaired phonological output lexicon on the reading tasks indicates that they show a reading pattern of surface dyslexia in oral reading, but not in orthographic lexical decision and written comprehension tasks that do not involve oral reading. This indicates that these individuals rely on the sublexical route when they need to read words aloud, and that this does not result from a deficit in the orthographic input lexicon.

This pattern, of sublexical reading aloud with preserved orthographic input lexicon and access from it to the semantic system, applied both to the participants with acquired phonological output lexicon impairment and to those with a developmental phonological output lexicon impairment.

Recall that we selected the participants to this study solely on the basis of their naming deficit, which results from a phonological output lexicon impairment. We only then tested their reading patterns. Given that all these participants showed surface dyslexia in reading aloud, we can conclude that the phonological output lexicon deficit causes surface dyslexia in reading aloud, and that it can occur alongside good performance in tasks that do not involve speech output. Some of the participants, and particularly the youngest developmental anomic MAD and ARO, may have also had an orthographic input lexicon impairment or at least have not yet established a rich enough set of lexical entries in this lexicon, on top of their phonological output lexicon impairment. Importantly,

⁷Since, the completion of this paper, we tested two additional aphasic patients with phonological lexicon anomia, a man and a woman. Both showed the same pattern as the other 16: very impaired naming with mainly phonological paraphasias and long hesitations, alongside good semantic abilities and good non-word and morphologically complex word repetition. Both showed poor reading aloud of irregular words and potentiophones, with unimpaired lexical decision of pseudohomophones and homophone comprehension.

⁸The good performance of all the participants in the lexical decision and homophone comprehension tasks also bears on the reading of the four participants who acquired Hebrew reading in their teens. It indicates that the homophones and irregular words were represented correctly in their lexicons and their surface errors in reading aloud did not result from Hebrew being their second language.

TABLE 7 | Lexical decision and comprehension tasks: %correct.

Participant	Lexical decision	Homophone-potentiophone comprehension
ACQUIRED ANOMIA		
DAN	96%	95%
YOS	100%	100%
ZAB	93%**	93%*
BAR	100%	95%
LER	100%	100%
ARI	99%	93%*
NAV	100%	100%
DOR	98%	98%
DEVELOPMENTAL ANOMIA		
LEO	100%	98%
TAF	100%	100%
AFI	100%	100%
ARO	75.7%*	74%
MAD	89%**	90%*
SAN	96%	93%
NIV	100%	95%
SHL	96%	100%
CONTROL GROUPS M (SD)		
Adults, at least 12 years education	98.0% (1.5) N = 148	98.3% (2.4) N = 141
6th–9th grade control groups	9th grade	7–9th grade
	99.3% (1.1) N = 59	97.4% (2.7) N = 74
	7–8th grade	
	98.9% (2.7) N = 24	
6th grade		
	97.4% (2.3) N = 10	
5th grade control group	98.6% (1.8) N = 28	94.2% (4.7) N = 18
4th grade control group	93.8% (10.1) N = 80	87.9% (11.6) N = 77

Significantly more errors than age-matched control group, * $p < 0.05$, ** $p = 0.001$.

however, the fact that there were 12 participants who showed completely normal performance in the input reading tasks indicates that phonological output lexicon impairment can cause a very selective surface dyslexia, which only affects reading aloud.

DISCUSSION

Phonological Output Lexicon Deficit Causes Surface Dyslexia

Lexical retrieval and reading are often depicted using different models, and studied by different researchers. However, the current study demonstrated that they are tightly linked. We focused on a component that is part of both lexical retrieval and reading aloud: the phonological output lexicon. Our main finding is that individuals with acquired or developmental anomia that results from a deficit in the phonological output lexicon also show a very clear and consistent deficit in reading: when they read aloud they make regularization errors in

irregular words, indicating reading via the sublexical route, but when their silent reading is tested, in tests of lexical decision and written words comprehension, which do not involve phonological output, they perform normally. A look at the models of reading and lexical retrieval explains exactly why this is so: the phonological output lexicon is part of the lexical route for reading aloud, and its impairment results in reading aloud via the sublexical route. However, because the deficit is only located in a late, output stage of reading, their input, including the orthographic input lexicon, is not impaired, and this is what allows them to judge correctly pseudo-homophones as non-words, and to understand written words well, including homophones and potentiophones. This pattern held for individuals with various sources of phonological lexicon anomia: acquired and developmental, for individuals in different ages and levels of education. This indicates that this strong relation between phonological output lexicon anomia and surface dyslexia occurs independently of specific source of impairment. The fact that the individuals with developmental anomia showed the same pattern as the individuals with acquired anomia also suggests an interesting insight about reading acquisition. It suggests that entrees in the orthographic input lexicon and their connection to the semantic lexicon can be established even when the phonological output lexicon is impaired. Namely, the orthographic input lexicon can be established even without well-functioning reading aloud.

The Road Not Taken

Given that the phonological output lexicon is part of the lexical route, two possibilities are imaginable for the way a deficit in the phonological output lexicon may affect reading: one is that reading via the lexical route is blocked and hence reading has to proceed via the sublexical route, giving rise to surface dyslexia. The other is that the reader with phonological output lexicon would still use the impaired lexical route for reading aloud, and this would result in phonological errors in reading aloud that are similar to the errors made in speech production. Our results from the participants who had a selective deficit in the phonological output lexicon indicate that they only read via the sublexical route, and the other theoretically possible option is not attested: they do not read via the impaired lexical route, and do not make phonological errors in reading aloud. There were five participants in the current study who did make errors in reading beyond surface dyslexia errors that could be phonological errors. Importantly, such errors occurred only in the five participants who had, in addition to their phonological output lexicon impairment, also an impairment in the phonological output buffer. Their phonological errors in reading, thus, can be ascribed to the later, phonological output buffer deficit and not to reading via the lexical route. This may also explain the pattern of errors reported for Friedman and Kohn's (1990) patient HR: HR had impaired phonological output in naming, and in reading aloud he made phonological errors. On the basis of his impaired processing of non-words and the length effect he showed in all production tasks, one may conclude that his deficit did not lie at the phonological output lexicon but rather in the phonological output buffer, and this was the source of his phonological errors in reading aloud.

Our results suggest another type of surface dyslexia, which occurs both in acquired dyslexia and in developmental dyslexia: surface dyslexia that results from a deficit in the phonological output lexicon (see also EST and EE in Coltheart, 1982; Kay and Patterson, 1985; Howard and Franklin, 1987; Kay and Ellis, 1987, for earlier cases of acquired phonological anomia and surface dyslexia, albeit with a less selective pattern). This type of surface dyslexia joins other types of surface dyslexia that have been reported: a selective deficit in the orthographic input lexicon, a deficit in the output of the orthographic input lexicon (to the phonological output lexicon and to the semantic lexicon), and an inter-lexical deficit between the orthographic input lexicon and the phonological output lexicon (Coltheart and Funnell, 1987; Friedmann and Lukov, 2008).

Given the consistent effect the deficit in the phonological output lexicon had on the reading of the participants, regularization errors in reading aloud may be taken in the future as another tool for the functional localization of the source of anomia in the lexical retrieval process. It is often difficult, for example, to distinguish between a deficit in the connection between the semantic lexicon and the phonological output lexicon and a deficit in the phonological output lexicon itself. Our findings suggest a way to distinguish between the two, as a deficit in the connection between the semantic lexicon and the phonological output lexicon should not cause surface dyslexia (given that a direct connection between the orthographic input lexicon and the phonological output lexicon is still available for reading aloud)⁹, but phonological output lexicon deficit should.

⁹We assume that such a direct route exists on the basis of reports of patients who had intact orthographic and phonological lexica and intact semantic route (good comprehension of written words including homophones, and good naming) with surface errors in reading aloud (Friedmann and Lukov, 2008; Khentov-Kraus and Friedmann, 2011), and of patients who show the opposite dissociation, with impaired semantics and good reading aloud of irregular words (Wilson and Martinez-Cuitiño, 2012).

REFERENCES

- Biran, M., and Friedmann, N. (2004). *SHEMESH: Naming a Hundred Objects*. Tel Aviv: Tel Aviv University.
- Biran, M., and Friedmann, N. (2005). From phonological paraphasias to the structure of the phonological output lexicon. *Lang. Cogn. Process.* 20, 589–616. doi: 10.1080/01690960400005813
- Biran, M., and Friedmann, N. (2007). *Ma Kashur: Word Association Test*. Tel Aviv: Tel Aviv University.
- Biran, M., and Friedmann, N. (2012). The representation of lexical-syntactic information: Evidence from syntactic and lexical retrieval impairments in aphasia. *Cortex* 48, 11031–11127. doi: 10.1016/j.cortex.2011.05.024
- Butterworth, B. (1989). "Lexical access in speech production," in *Lexical Representation and Process*, ed W. Marslen-Wilson (Cambridge, MA: MIT Press), 108–135.
- Butterworth, B. (1992). Disorders of phonological encoding. *Cognition* 42, 261–286. doi: 10.1016/0010-0277(92)90045-J
- Castles, A., Bates, T., and Coltheart, M. (2006). John Marshall and the developmental dyslexias. *Aphasiology* 20, 871–892. doi: 10.1080/02687030600738952
- Castles, A., and Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition* 47, 149–180. doi: 10.1016/0010-0277(93)90003-E

The identification of this shared destiny between lexical retrieval and reading impairments also has clinical implications: when a person has a deficit in the phonological output lexicon, either due to brain damage or from birth, one may expect this person to have difficulties in oral reading as well. Treatment of the lexical retrieval difficulty is thus expected to also reduce errors in reading aloud. Importantly, these difficulties in reading only affect reading aloud. These people can still understand and recognize written words very well. Therefore, the clinician can provide a very straightforward recommendation to these individuals: do not read aloud.

AUTHOR CONTRIBUTIONS

The authors conceived the research question together, collected the data together, analyzed the data together, thought together about the results and their implications, and wrote the paper together.

ACKNOWLEDGMENTS

We are grateful to Maya Yachini and Dana Rusou for their valuable comments, and to Maya Yachini and Adi Kesselman for recruiting and testing SAN, NIV, and SHL. This research was supported by the Israel Science Foundation (grant no. 1066/14, Friedmann), by the Lieselotte Adler Laboratory for Research on Child Development, and by the Australian Research Council Centre of Excellence for Cognition and its Disorders (CE110001021) <http://www.ccd.edu.au>.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00340>

- Castles, A., and Coltheart, M. (1996). Cognitive correlates of developmental surface dyslexia: a single case study. *Cogn. Neuropsychol.* 13, 25–50. doi: 10.1080/026432996382051
- Coltheart, M. (1981). Disorders of reading and their implications for models of normal reading. *Visible Lang.* 15, 245–286.
- Coltheart, M. (1982). The psycholinguistic analysis of acquired dyslexias: some illustrations. *Philos. Trans. R. Soc. B Biol. Sci.* 298, 151–164. doi: 10.1098/rstb.1982.0078
- Coltheart, M., and Byng, S. (1989). "A treatment for surface dyslexia," in *Cognitive Approaches in Neuropsychological Rehabilitation*, ed X. Seron (London: Erlbaum), 159–174.
- Coltheart, M., and Funnell, E. (1987). "Reading and writing: one lexicon or two" in *Language Perception and Production*, eds A. Allport, D. Mackay, W. Prinz, and E. Sheerer (London: Academic Press), 313–339.
- Coltheart, M., Masterson, J., Byng, S., Prior, M., and Riddoch, J. (1983). Surface dyslexia. *Q. J. Exp. Psychol.* 35, 469–495. doi: 10.1080/14640748308402483
- 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
- Crawford, J. R., and Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia* 40, 1196–1208. doi: 10.1016/S0028-3932(01)00224-X

- Crawford, J. R., and Garthwaite, P. H. (2012). Single-case research in neuropsychology: a comparison of five forms of t-test for comparing a case to controls. *Cortex* 48, 1009–1016. doi: 10.1016/j.cortex.2011.06.021
- Crawford, J. R., and Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *Clin. Neuropsychol.* 12, 482–486. doi: 10.1076/clin.12.4.482.7241
- Dell, G. S. (1988). The retrieval of phonological forms in production: tests of predictions from a connectionist model. *J. Mem. Lang.* 27, 124–142. doi: 10.1016/0749-596X(88)90070-8
- Dotan, D., and Friedmann, N. (2015). Steps towards understanding the phonological output buffer and its role in the production of numbers, morphemes, and function words. *Cortex* 63, 317–351. doi: 10.1016/j.cortex.2014.08.014
- Ellis, A. W. (1993). *Reading, Writing, and Dyslexia: A Cognitive Analysis*, 2nd Edn. London: Erlbaum.
- Ellis, A. W., Flude, B. M., and Young, A. W. (1987). "Neglect dyslexia" and the early visual processing of letters in words and nonwords. *Cogn. Neuropsychol.* 4, 439–463. doi: 10.1080/02643298708252047
- Ellis, A. W., Lambon Ralph, M. A., Morris, J., and Hunter, A. (2000). "Surface dyslexia: description, treatment, and interpretation," in *Case Studies in the Neuropsychology of Reading*, ed E. Funnell (Hove: Psychology Press), 85–119.
- Ellis, A. W., and Young, A. W. (1996). *Human Cognitive Neuropsychology*. Hove: Erlbaum.
- Ferreres, A. R., Cuitiño, M. M., and Olmedo, A. (2005). Acquired surface alexia in Spanish: a case report. *Behav. Neurol.* 16, 71–84. doi: 10.1155/2005/473407
- Friedman, R. B., and Kohn, S. E. (1990). Impaired activation of the phonological lexicon: effects upon oral reading. *Brain Lang.* 38, 278–297. doi: 10.1016/0093-934X(90)90115-W
- Friedmann, N. (2003). *BLIP: Battery for Assessment of Phonological Abilities*. Tel Aviv: Tel Aviv University.
- Friedmann, N. (2006). *BUFFY: Battery for Assessment of the Phonological Output Buffer*. Tel Aviv: Tel Aviv University.
- Friedmann, N., and Biran, M. (2003). When is gender accessed? A study of paraphasias in Hebrew anomia. *Cortex* 39, 441–463. doi: 10.1016/S0010-9452(08)70258-2
- Friedmann, N., Biran, M., and Dotan, D. (2013). "Lexical retrieval and breakdown in aphasia and developmental language impairment," in *The Cambridge Handbook of Bilingualism*, eds C. Boeckx and K. K. Grohmann (Cambridge, UK: Cambridge University Press), 350–374.
- Friedmann, N., and Coltheart, M. (2016). "Types of developmental dyslexia," in *Handbook of Communication Disorders: Theoretical, Empirical, and Applied Linguistics Perspectives*, eds A. Bar-On and D. Ravid (Berlin, Boston, MA: De Gruyter Mouton).
- Friedmann, N., and Gvion, A. (2001). Letter position dyslexia. *Cogn. Neuropsychol.* 18, 673–696. doi: 10.1080/02643290143000051
- Friedmann, N., and Gvion, A. (2002). *FriGvi: Friedmann Gvion Battery for Assessment of Phonological Working Memory*. Tel Aviv: Tel Aviv University.
- Friedmann, N., and Gvion, A. (2003). *TILTAN: Battery for the Diagnosis of Dyslexias in Hebrew*. Tel Aviv: Tel Aviv University.
- Friedmann, N., Kerbel, N., and Shvimer, L. (2010). Developmental attentional dyslexia. *Cortex* 46, 1216–1237. doi: 10.1016/j.cortex.2010.06.012
- Friedmann, N., and Lukov, L. (2008). Developmental surface dyslexias. *Cortex* 44, 1146–1160. doi: 10.1016/j.cortex.2007.09.005
- Friedmann, N., and Lukov, L. (2011). "Subtypes of developmental surface dyslexia and its manifestations in Hebrew," in *Literacy and Language: Interactions, Bilingualism, and Difficulties*, eds O. Korat and D. Aram (Jerusalem: Magnes), 414–442.
- Friedmann, N., and Rahamim, E. (2007). Developmental letter position dyslexia. *J. Neuropsychol.* 1, 201–236. doi: 10.1348/174866407X204227
- Garrett, M. F. (1976). "Syntactic processes in sentence production," in *New Approaches to Language Mechanisms*, eds R. J. Wales and E. Walker (Amsterdam: North Holland Publishing Co.), 231–256.
- Garrett, M. F. (1992). Disorders of lexical selection. *Cognition* 42, 143–180. doi: 10.1016/0010-0277(92)90042-G
- Gil, M., and Edelstein, C. (2001). *Hebrew Version of the PALPA*. Ra'anana: Loewenstein Hospital Rehabilitation Center.
- Gvion, A., and Friedmann, N. (2012). Phonological short-term memory in conduction aphasia. *Aphasiology* 26, 579–614. doi: 10.1080/02687038.2011.643759
- Howard, D., and Franklin, S. (1987). "Three ways for understanding written words, and their use in two contrasting cases of surface dyslexia," in *Language Perception and Production: Relationships between Listening, Speaking, Reading and Writing*, eds A. Allport, D. MacKay, W. Prinz, and E. Scheerer (London: Academic Press), 340–366.
- Howard, D., and Gatehouse, C. (2006). Distinguishing semantic and lexical word retrieval deficits in people with aphasia. *Aphasiology* 20, 921–950. doi: 10.1080/02687030600782679
- Humphreys, G. W., Evett, L. J., and Quinlan, P. T. (1990). Orthographic processing in visual word identification. *Cogn. Psychol.* 22, 517–560. doi: 10.1016/0010-0285(90)90012-S
- Judica, A., De Luca, M., Spinelli, D., and Zoccolotti, P. (2002). Training of developmental surface dyslexia improves reading performance and shortens eye fixation duration in reading. *Neuropsychol. Rehabil.* 12, 177–197. doi: 10.1080/09602010244000002
- Kay, J., and Ellis, A. W. (1987). A cognitive neuropsychological case study of anomia: implications for psychological models of word retrieval. *Brain* 110, 613–629. doi: 10.1093/brain/110.3.613
- Kay, J., Lesser, R., and Coltheart, M. (1992). *PALPA: Psycholinguistic Assessments of Language Processing in Aphasia*. Hove: Erlbaum.
- Kay, J., and Patterson, K. E. (1985). "Routes to meaning in surface dyslexia," in *Surface Dyslexia: Neuropsychological and Cognitive Studies of Phonological Reading*, eds K. E. Patterson, J. C. Marshall, and M. Coltheart (London: Erlbaum), 79–103.
- Kempen, G., and Huijbers, P. (1983). The lexicalization process in sentence production and naming: indirect election of words. *Cognition* 14, 185–209. doi: 10.1016/0010-0277(83)90029-X
- Khentov-Kraus, L., and Friedmann, N. (2011). Dyslexia in vowel letters. *Lang. Brain* 10, 65–106.
- Kohn, S., and Melvold, J. (2000). Effects of morphological complexity on phonological output deficits in fluent and nonfluent aphasia. *Brain Lang.* 73, 323–346. doi: 10.1006/brln.2000.2179
- Kohnen, S., Nickels, L., Castles, A., Friedmann, N., and McArthur, G. (2012). When "slime" becomes "smile": developmental letter position dyslexia in English. *Neuropsychologia* 50, 3681–3692. doi: 10.1016/j.neuropsychologia.2012.07.016
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1992). Accessing words in speech production: stages, processes and representations. *Cognition* 42, 1–22. doi: 10.1016/0010-0277(92)90038-J
- Margolin, D. I., Marcel, A. J., and Carlson, N. R. (1985). "Common mechanisms in dysnomia and post-semantic surface dyslexia: processing deficits and selective attention," in *Surface Dyslexia: Neuropsychological and Cognitive Studies of Phonological Reading*, eds K. E. Patterson, J. C. Marshall, and M. Coltheart (London: Lawrence Erlbaum), 139–173.
- Marshall, J. C., and Newcombe, F. (1973). Patterns of paralexia: a psycholinguistic approach. *J. Psycholinguist. Res.* 2, 175–199. doi: 10.1007/BF01067101
- Masterson, J. (2000). "Developmental surface dyslexia," in *Neuropsychology of Reading*, eds E. Funnell (Gloucester: Psychology Press), 123–148.
- Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: achievements, limitations, and alternatives. *Cognition* 42, 181–211. doi: 10.1016/0010-0277(92)90043-H
- Miceli, G., Amitrano, A., Cappaso, R., and Caramazza, A. (1996). The treatment of anomia resulting from output lexicon damage: analysis of two cases. *Brain Lang.* 52, 150–174. doi: 10.1006/brln.1996.0008
- Newcombe, F., and Marshall, J. C. (1981). On psycholinguistic classifications of the acquired dyslexias. *Bull. Orton Soc.* 31, 29–46. doi: 10.1007/BF02658599
- Newcombe, F., and Marshall, J. C. (1984). Varieties of acquired dyslexia: a linguistic approach. *Semin. Neurol.* 4, 181–195. doi: 10.1055/s-2008-1041548
- Newcombe, F., and Marshall, J. C. (1985). "Reading and writing by letter sounds," in *Surface Dyslexia: Cognitive and Neuropsychological Studies of Phonological Reading*, eds K. E. Patterson, J. C. Marshall, and M. Coltheart (London: Erlbaum), 15–34.
- Nickels, L. (1997). *Spoken Word Production and Its Breakdown in Aphasia*. Hove: Psychology Press.

- Nickels, L. (2002). Theoretical and methodological issues in the cognitive neuropsychology of spoken word production. *Aphasiology* 16, 3–19. doi: 10.1080/02687040143000645
- Nickels, L. A. (1995). Getting it right – Using aphasic naming errors to evaluate theoretical models of spoken word recognition. *Lang. Cogn. Process.* 10, 13–45. doi: 10.1080/01690969508407086
- Nickels, L. A., and Howard, D. (2000). “When the words won’t come: relating impairments and models of spoken word production,” in *Aspects of Language Production*, ed L. R. Wheeldon (Hove: Psychology Press), 115–142.
- Nickels, L., and Howard, D. (1994). A frequent occurrence – factors affecting the production of semantic errors in aphasic naming. *Cogn. Neuropsychol.* 11, 289–320. doi: 10.1080/02643299408251977
- Patterson, K., and Shewell, C. (1987). “Speak and spell: dissociation and word-class effects,” in *The Cognitive Neuropsychology of Language*, eds M. Coltheart, R. Job, and G. Sartori (Hove: Erlbaum), 273–294.
- Peressotti, F., and Grainger, J. (1995). Letter position coding in random consonant arrays. *Percept. Psychophys.* 57, 875–890. doi: 10.3758/BF03206802
- Schmalz, X., Marinus, E., Coltheart, M., and Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychon. Bull. Rev.* 22, 1614–1629. doi: 10.3758/s13423-015-0835-2
- Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition* 42, 213–259. doi: 10.1016/0010-0277(92)90044-I
- Temple, C. M. (1997). *Developmental Cognitive Neuropsychology*. Hove: Psychology Press.
- Weekes, B., and Coltheart, M. (1996). Surface dyslexia and surface dysgraphia: treatment studies and their theoretical implications. *Cogn. Neuropsychol.* 13, 277–315. doi: 10.1080/026432996382033
- Wilson, M. A., and Martínez-Cuitiño, M. (2012). Semantic dementia without surface dyslexia in Spanish: unimpaired reading with impaired semantics. *Behav. Neurol.* 25, 273–284. doi: 10.3233/BEN-2012-119009

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

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

