10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

The Roles of Segment and Tone in Bi-dialectal Auditory Word recognition

Min Liu[1], *Yiya Chen*[2]

[1]College of Chinese Language and Culture, Jinan University, China
[2]Leiden University Centre for Linguistics, Leiden University, the Netherlands
nwliumin@gmail.com, yiya.chen@hum.leidenuniv.nl

Abstract

This study investigated if and how cross-dialect phonological similarity in segment and tone affects bi-dialectal listeners' lexical access during spoken word recognition. Balanced bi-dialectal speakers of Xi'an Mandarin and Standard Chinese took part in an auditory-auditory priming experiment with a generalized lexical decision task. The primes were monosyllabic homophones from either Xi'an Mandarin or Standard Chinese, while the targets were disyllabic Xi'an Mandarin or Standard Chinese words. Primes and the first syllable of the target words overlapped in either only segment or both segment and tone within the same dialect or across the two dialects. In addition, a control condition was included where primes and targets shared neither tone nor segment. The results showed that cross-dialect phonological similarity in segment alone does not affect lexical access in bi-dialectal auditory word recognition while cross-dialect phonological similarity in both segment and tone poses a threat to the recognition system of bi-dialectal listeners. Cross-dialect homophones have been processed much more slowly and less accurately. We conclude that tonal information plays a significant role in constraining word activation in bi-dialectal auditory word recognition.

Index Terms: segment, tone, bi-dialectal auditory word recognition, Standard Chinese, Xi'an Mandarin

1. Introduction

Little research has been conducted to investigate the effect of phonological similarity on bi-dialectal spoken word recognition in tonal languages, despite the mounting evidence on the role of phonological similarity in bilingual word recognition in non-tonal Indo-European languages.

Speaking of phonological similarity, an extreme case of it is homophony. Bilingual word recognition studies from Indo-European languages have consistently shown that interlingual homophones are harder for bilingual speakers to process than non-homophonous control words. And the effect is robust across experimental tasks and modalities, be it a lexical decision task [1-4], a gating task [5] or a word form priming task [6], and be stimuli presented in the visual [1, 2] or the auditory modality [3]. These studies suggest parallel activation of homophone candidates from both languages, which leads to an interference effect on word recognition.

For tonal languages, phonological similarity between languages can be more complicated as there could be phonological overlap in not only segments but also supra-segmental lexical tones. Little, however, do we know about how similarities in segment and/or tone affect spoken word recognition for bilingual tonal language speakers. Among the many tonal languages, Chinese is perhaps the most widely studied. One long-neglected fact in the existing literature is that the majority of speakers are proficient in at least two varieties: the national language Standard Chinese and their regional native dialect. Little attention has been paid to the possible co-activation of homophones in bi-dialectal word recognition. The question that arises is whether in bi-dialectal lexical processing, homophones co-activate and interfere, as in the bilingual situation. More specifically, for tonal language speakers, what role does lexical tone plays in the activation and processing of bi-dialectal lexical representations during spoken word recognition?

It is important to note that existing studies on the role of lexical tone in spoken word processing have mostly been conducted on Chinese varieties assuming a monolingual context. These studies [e.g., 7, 8] have showed controversial results as to whether tonal information constrains lexical activation in Standard Chinese. In [7], a facilitatory priming effect was found when primes and targets overlapped in both segment and tone. Segment-only overlap (minimal tone pair) did not produce any priming effect, comparable to the baseline condition where primes and targets overlapped in neither segment nor tone. The latter result was taken as evidence that tones are used to constrain lexical activation. However, in [8]. a facilitatory priming effect was found for segment-only overlapped primes and targets. To shed further light on this issue, this study takes stock of the fact that the majority of speakers are bi-dialectal speakers and taps into the roles of segment and tone in Standard Chinese in a bi-dialectal context.

The two dialects examined in our study are Standard Chinese (SC) and Xi'an Mandarin (XM). Both belong to the Mandarin family, which is one of the largest dialect families in Chinese [9, 10]. SC and XM share a common logographic writing system and overlap largely in segmental features. Both have four lexical tones, and there is one-to-one mapping between SC and XM tones in their F0 contours [11]. Relevant to this study is that XM_T1 tone (e.g., ma^{T1}, "mother") overlaps with SC_T4 tone (e.g., ma^{T4}, "to scold") in their F0 contour (both level) but not in tonal category (i.e. T1 vs. T4). And XM_T2 tone (e.g., ma^{T2}, "hemp") overlaps with SC_T2 (e.g., ma^{T2}, "hemp") in both tonal contour (both rising) and tonal category (both T2).

2. Method

2.1. Participants

One-hundred balanced XM_SC bi-dialectal speakers (41 males, 59 females) were selected and paid to participate in the experiment. To assess their language proficiency in the two dialects, we asked participants to read the story "the North Wind and the Sun" in both SC and XM. In addition, an adapted version of the LEAP-Q questionnaire [12] was used to thoroughly check their language background and language

 10.21437/SpeechProsody.2020-131

proficiency. All the selected participants were of high and comparable speaking proficiency (XM vs. SC: 7.8 vs. 8.1, $t(99) = -1.05$, $p = .30$) and spoken language comprehension skills (XM vs. SC: 8.3 vs. 8.6, $t(99) = -1.11$, $p = .27$) in the two dialects, on a scale from 0 (none) to 10 (perfect). They were born and raised in the urban areas of Xi'an and had no living experiences outside of Xi'an. All were undergraduate or graduate students at local universities, with an age range from 19 to 28 ($M \pm SD$: 21.7 ± 3.2). None of them reported any speech or hearing disorders. Informed consent was obtained from all the participants before the experiment.

2.2. Stimuli

Forty SC_T1 monosyllables and their corresponding interdialectal homophonous XM_T4 monosyllables with comparable high word frequency were selected as primes. Each monosyllabic prime was paired with five disyllabic targets that differed in their degrees of phonological similarity to the primes (see below for more details). In total, there were 400 prime-target trials (40 × 2 Prime types × 5 Target types = 400).

The first syllable of the disyllabic targets is our focus of interest. We therefore differentiated the five disyllabic target types according to the phonological property of their first syllable. The second syllable of the disyllabic targets always bears a T2, because T2 shows great resemblance in acoustic realization between the two dialects, i.e., SC_T2 maps onto XM_T2 both categorically and acoustically, and thus lends no ambiguity to the dialect membership of the disyllabic word by itself. Together with the first syllable, however, T2 syllables do cue the dialect membership information of the disyllabic word. For each prime type (e.g., SC_T1, level contour, "bang^{T1}/帮", "help"), the five disyllabic target types included a within-dialect segment and tone overlap target (e.g., SC_T1 target, level contour, "bang^{T1}mang^{T2}/帮忙", "help"; hereafter coded as Identical), a within-dialect segment-only overlap target (e.g., SC_T4 target, falling contour, "bang^{T4}qiu^{T2}/棒球", "baseball"; hereafter coded as D+Seg), a cross-dialect segment-only overlap target (e.g., XM_T1 target, low contour, "bang^{T1}mang^{T2}/帮忙", "help"; hereafter coded as D−Seg), a cross-dialect segment and tone overlap, i.e., an interlingual homophone, target (e.g., XM_T4 target, level contour, "bang^{T4}qiu^{T2}/棒球", "baseball"; hereafter coded as D−Homophone), and a within-dialect control target which had neither segment nor tone overlap with the prime and served as baseline (e.g., control target, rising contour, "wan^{T2}cheng^{T2}/完成", "finish"; hereafter coded as Baseline). All selected targets had comparable word frequencies, according to the SUBTLEX-CH frequency list [13].

In addition to the SC_T1 and XM_T4 monosyllabic primes and the disyllabic word targets, the same number of monosyllabic primes and disyllabic nonword targets were paired for other tone pairs and added in the experiment based on similar logic.

2.3. Stimuli recording

A balanced XM_SC bi-dialectal male speaker was recruited to produce the stimuli in two separate blocks for the two dialects. This speaker was born and raised in the urban area of Xi'an and had no living experience out of Xi'an. He learned XM and SC simultaneously when he was young and was of high and comparable proficiency in the two dialects. He was an undergraduate student at a local university and used the two dialects equally frequently in his daily life. All the stimuli were recorded by him in a soundproof room at 16-bit resolution and a sampling rate of 44.1 kHz on a laptop via an external digitizer (UA-G1). The recorded stimuli were trimmed of silence and normalized amplitude for perception using Praat [14].

2.4. Procedure

We adopted an auditory-auditory priming paradigm in the perception experiment, with a monosyllabic prime preceding a disyllabic target in each trial. All the trials were distributed in a Latin Square design, so that participants only heard the same stimulus (for both prime and target) once during the experiment. Consequently, all the stimuli were divided into five lists. Each list contained both prime types and all the five target types. For a given prime type, only one of the five target types occurred in every list. The prime-nonword target trials were constructed in the same way in the list. In sum, each list included 80 prime-word target trials (40 × 2 Prime types × 1 Target type) and 80 prime-nonword target trials (40 × 2 Prime types × 1 Target type) with the five types of targets equally distributed.

Participants were tested individually on one list only in a soundproof booth of the behavioral lab at Shaanxi Normal University in Xi'an. Across all participants, the five lists were presented equally often (20 participants/list). All the trials in each list were presented to the participants using the E-Prime 2.0 software through headphones at a comfortable listening level. Trials were pseudo-randomized with the restriction that the shortest distance between the two interdialectal homophone primes was 9 trials and the shortest distance between two targets of the same type was 3 trials.

The experiment included a practice block and two experimental blocks. The practice block contained 10 trials to familiarize the participants with the task. These trials were not used in the experimental blocks. Each experimental block contained 80 trials. Between each block, there was a self-paced break. Each trial started with a 100 ms warning beep, followed by a 300 ms pause. Participants then heard a pair of speech items separated by a 250 ms interval. The first item was a monosyllabic prime, and the second a disyllabic target. They were then given up to 3 seconds after target offset to perform a lexical decision task. They were asked to respond as accurately and quickly as possible. They had to press the button labelled "yes" on the keyboard if they think the target is a real word in either of the dialects. If the target is neither a word in SC nor in XM, they were asked to press the button labeled "no" on the keyboard. This task is also known as the generalized lexical task. Button-press latencies were measured from the target offset. Instructions were given both visually on screen in simplified Chinese characters and orally by the experimenter in mixed fashion of the two dialects (both SC and XM) before the experiment.

2.5. Data analysis

We restricted our analyses to the prime-word targets trials. The dependent variables included response accuracy and reaction time. Response accuracy was defined as the percentage of correct judgments of the word targets in the lexical decision task. Reaction time was defined as the response time relative to the offset of the word targets which were correctly responded to. To normalize the distribution, raw reaction times were transformed using the natural logarithm.

Statistical analyses were carried out with the package *lme4* [15] in R version 3.1.2 [16]. Analysis of response accuracy was performed using binomial logistic regression models, and analysis of reaction time was performed using linear mixed-effects regression models. The models included Prime type (SC_T1, XM_T4), Target type (Identical, D+Seg, D−Seg, D−Homophone, Baseline) and their interactions as fixed factors, and Subjects and Items as random factors. The fixed factors were added in a stepwise fashion and their effects on model fits were evaluated via model comparisons based on log-likelihood ratios. For Target type, all the conditions were first compared with the baseline condition. Post-hoc pairwise comparisons between different target conditions were conducted using *lsmeans* package [17] with single-step *p*-value adjustment. For models of reaction time, trials with absolute standardized deviations exceeding 2.5 from the mean were considered as outliers and removed from further analysis.

3. Results

3.1. Response accuracy

Figure 1 presents the response accuracy with 95% CI for different target types preceded by SC_T1 and XM_T4 primes. Results showed a significant main effect of Target type ($\chi^2(4) = 18.73$, $p < .001$) and a significant two-way interaction of Prime type × Target type ($\chi^2(4) = 29.28$, $p < .001$). No main effect of Prime type was found ($\chi^2(1) = 0.40$, $p = .52$).

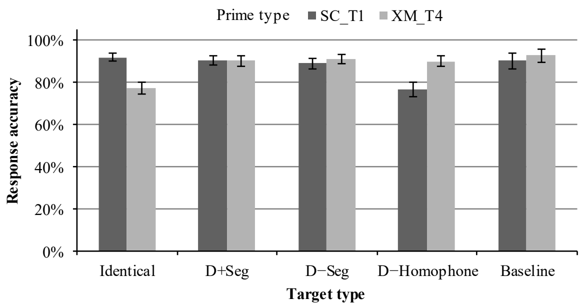

Figure 1: *Response accuracy with 95% CI for different target types.*

Separate models were constructed for subset data of different prime types. When the prime was SC_T1, there was a significant main effect of Target type ($\chi^2(4) = 20.19$, $p < .001$), indicating that the response accuracy differed significantly among different target types. Further multiple pairwise comparisons showed that the response accuracy for the interdialectal homophone target was significantly lower than that for the other four target types (all *p*s < .05). No difference was found for any other pair of target types (all *p*s > .05). Overall, a SC_T1 prime made the recognition of its interdialectal homophone target XM_T4 words more erroneous for the balanced XM_SC bi-dialectal listeners.

When the prime was XM_T4, there was also a significant main effect of Target type ($\chi^2(4) = 27.88$, $p < .001$). Surprisingly, multiple pairwise comparisons showed that the response accuracy for the identical target, rather than the interdialectal homophone target from SC, was significantly lower than that for the other four conditions (all *p*s < .05). No difference was found for any other pair of conditions (all *p*s > .05). In other words, when a XM_T4 monosyllabic prime preceded a XM_T4 disyllabic target, the recognition of the latter became a more error-prone process. Taken together, irrespective of whether the prime was the SC version or XM version of the homophone, the balanced XM_SC bi-dialectal listeners recognized the XM disyllabic target less accurately than the other target types.

3.2. Reaction time

2.3% of the data points were identified as outliers and removed from further analysis. Figure 2 presents the average reaction time with 95% CI for different target types preceded by SC_T1 and XM_T4 primes. The overall analyses showed a significant main effect of Target type ($\chi^2(4) = 15.22$, $p = .004$) and a significant two-way interaction of Prime type × Target type ($\chi^2(4) = 75.29$, $p < .001$). No main effect of Prime type was found ($\chi^2(1) = 0.09$, $p = .77$).

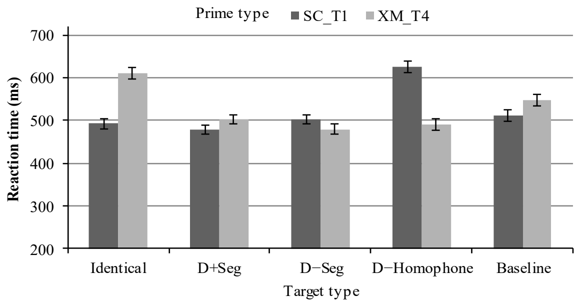

Figure 2: *Average reaction time with 95% CI for different target types.*

Separate models were constructed for subset data of different prime types. When the prime was SC_T1, there was a significant main effect of Target type ($\chi^2(4) = 46.05$, $p < .001$). Pairwise comparisons showed a non-significant facilitatory priming effect by the SC_T1 prime for the identical target compared to the baseline target ($\beta = -0.07$, $t = -1.31$, $p = .69$). Likewise, SC_T1 primes did not yield significant priming effect for the within-dialect segment-only overlap target (D+Seg) ($\beta = -0.07$, $t = -1.33$, $p = .67$), nor for the cross-dialect segment-only overlap target (D−Seg) ($\beta = 0.03$, $t = -0.61$, $p = .97$). However, they did yield a significant inhibitory priming effect for the cross-dialect homophone target (D−Homophone) ($\beta = 0.25$, $t = 4.80$, $p < .001$).

Similar analyses were conducted for the data of XM_T4 primes, where a reversed pattern of reaction time was found compared to the SC_T1 prime. There was a significant main effect of Target type ($\chi^2(4) = 44.22$, $p < .001$). Contrary to the facilitatory priming trend of the SC_T1 prime for the identical target, the XM_T4 prime showed a significant inhibitory priming effect for the identical target ($\beta = 0.15$, $t = 2.85$, $p = .04$) and a null effect for the within-dialect segment-only overlap target ($\beta = -0.07$, $t = -1.41$, $p = .62$). The cross-dialect segment-only overlap target ($\beta = -0.16$, $t = -3.14$, $p = .02$) and the inter-dialect homophone target ($\beta = -0.15$, $t = -2.98$, $p = .03$), on the other hand, showed a similar facilitatory priming effect by the XM_T4 prime.

The pattern for the XM_T4 prime data was counterintuitive. If the SC_T1 prime and XM_T4 prime were represented equally well in the mental lexicon of the XM_SC bi-dialectal speakers, we should have expected that the two prime types performed similarly on each target type. Yet the SC_T1 prime and XM_T4 prime behaved in a completely reversed fashion. What could be the possible reason for this?

One alternative way of viewing the reversed pattern of the XM_T4 prime data is revealed via a rearrangement of the reaction time data in Figure 3. This figure shows that the two prime types had an almost identical effect on the target types. Given that these two prime types are interdialectal homophones with basically no pronunciation difference, it is very likely that participants did not recognize the dialect membership upon hearing the XM_T4 primes and treated them as SC_T1 primes.

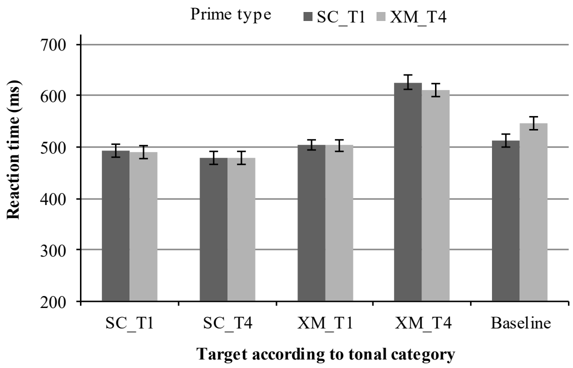

Figure 3: *Average reaction time with 95% CI for different targets arranged according to tonal category.*

We also ran statistical analyses for the rearranged data. Linear mixed-effects regression models were built for reaction time with the fixed factors Prime type (SC_T1, XM_T4) and Target tonal category (SC_T1, SC_T4, XM_T1, XM_T4, Baseline). We only found a significant main effect of Target tonal category ($\chi^2(4) = 88.16$, $p < .001$). Neither a main effect of Prime type ($\chi^2(1) = 0.06$, $p = .81$) nor a significant two-way interaction of Prime type × Target tonal category ($\chi^2(4) = 2.85$, $p = .58$) was found. The null effect of the Prime type suggests that the SC_T1 prime and XM_T4 prime were not treated differently. We are tempted to conclude that the XM_T4 monosyllabic primes were treated as their interdialectal homophonous SC_T1 equivalents in the current mixed dialect context.

4. Discussion

The present study investigated if and how cross-dialect phonological similarity in segment and tone affects lexical access in bi-dialectal auditory word recognition of balanced bi-dialectal tonal language listeners. In an auditory-auditory priming experiment with a generalized lexical decision task, we found that when the prime was in Standard Chinese, there was a non-significant within-dialect facilitatory priming trend for targets overlapping in both segment and tone with the prime, and also for targets overlapping only in segment with the prime. Both priming trends were of similar magnitude. The Standard Chinese prime did not produce any cross-dialect priming effect for the cross-dialect target with segment-only overlap. It, however, produced a significant inhibitory effect for the interdialectal homophone target relative to the unrelated Standard Chinese control target, as evidenced by the lower response accuracy and longer reaction time. The overall pattern was reversed when the prime was in Xi'an Mandarin, presumably because the Xi'an Mandarin prime was treated as its interdialectal homophonous Standard Chinese prime in the current mixed dialect setting. It seems that cross-dialect phonological similarity in segment alone does not affect lexical access in bi-dialectal auditory word recognition while a cross-dialect phonological similarity in both segment and tone (cross-dialect homophones) does pose a threat to the recognition system of the bi-dialectal tonal language listeners. Tonal information plays a significant role in constraining word activation in bi-dialectal auditory word recognition.

In this study, we found an evident inhibitory priming effect of inter-dialectal homophones, but only with a significant effect on Xi'an Mandarin targets. As inhibitory priming has generally been taken as evidence of competition between lexical candidates activated by the prime and the target [18-20], it seems safe to consider that there is competition among the lexical candidates activated by the prime and the cross-dialect homophone target. This possibly reflects that inter-dialectal homophones co-activate and interfere in current bi-dialectal lexical access process, which is in line with previous bilingual studies [3, 6]. The asymmetrical effect between Standard Chinese and Xi'an Mandarin is likely due to a language dominance effect of Standard Chinese, indicating that though the XM_SC bi-dialectal listeners we recruited are equally competent in both dialects, perhaps they are not balanced XM_SC bi-dialectal listeners after all. They are more likely to have overall language dominance in Standard Chinese rather than in Xi'an Mandarin.

In the monolingual Standard Chinese context, a significant facilitatory priming effect has been consistently found for the identical primes and targets which overlap in both segment and tone [7, 8]. In the current bi-dialectal context, we found that a complete overlap in segment and tone between the Standard Chinese primes and targets showed a non-significant facilitatory priming trend. Our results contrast with the results in the monolingual context. The identity priming effect of Standard Chinese primes and targets in the current bi-dialectal context was not as strong as that in the monolingual context reported in the literature. It shrunk in size and did not reach significance. Moreover, in line with [7] but not [8], we found no priming effect for both the within-dialect and cross-dialect targets with only segment overlap. This lack of significant facilitatory priming in the minimal tone pair within and across dialects suggests that the members of the minimal tone pair were not treated as homophones. We infer that tonal information is indeed used to constrain lexical activation in spoken word recognition.

5. Conclusions

To conclude, the present study showed that phonological similarity in segment alone does not affect lexical access in bi-dialectal auditory word recognition, whereas phonological similarity in both segment and tone poses a threat to the recognition system of the bi-dialectal tonal language listeners. Tonal information plays a significant role in constraining word activation in bi-dialectal auditory word recognition.

6. Acknowledgements

We thank the support from the Guangdong Planning Office of Philosophy and Social Science (Grant-GD19YYY06) to ML, the European Research Council (ERC Starting Grant-206198) and the Dutch Organization for Scientific Research (VI.C.181.040) to YC.

7. References

[1] T. Dijkstra, J. Grainger, and W. J. B. Van Heuven, "Recognition of cognates and interlingual homographs: The neglected role of

phonology," *Journal of Memory and Language,* vol. 41, no. 4, pp. 496-518, 1999.
[2] E. A. Doctor and D. Klein, "Phonological processing in bilingual word recognition." *Advances in Psychology,* vol. 83, pp. 237-252, 1992.
[3] E. Lagrou, R. J. Hartsuiker, and W. Duyck, "Knowledge of a second language influences auditory word recognition in the native language," *Journal of Experimental Psychology: Learning, Memory, and Cognition,* vol. 37, no. 4, pp. 952-965, 2011.
[4] G. Nas, "Visual word recognition in bilinguals: Evidence for a cooperation between visual and sound based codes during access to a common lexical store," *Journal of Verbal Learning and Verbal Behavior,* vol. 22, no. 5, pp. 526-534, 1983.
[5] F. Grosjean, "Exploring the recognition of guest words in bilingual speech," *Language and Cognitive Processes,* vol. 3, no. 3, pp. 233-274, 1988.
[6] B. Schulpen, T. Dijkstra, H. J. Schriefers, and M. Hasper, "Recognition of interlingual homophones in bilingual auditory word recognition," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 29, no. 6, pp. 1155-1178, 2003.
[7] C.-Y. Lee, "Does horse activate mother? Processing lexical tone in form priming," *Language & Speech,* vol. 50, no. 1, pp. 101-123. 2007.
[8] J. A. Sereno and H. Lee, "The contribution of segmental and tonal information in Mandarin spoken word processing," *Language & Speech,* vol. 58, no. 2, pp. 131-151, 2015.
[9] C. N. Li and S. A. Thompson, *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press, 1981.
[10] H. Chappell, *Sinitic Languages of China: Typological Descriptions*. Berlin, Boston: De Gruyter Mouton, 2001.
[11] M. Liu, *Tone and intonation processing*. The Netherlands: LOT, 2018.
[12] V. Marian, H. K. Blumenfeld, and M. Kaushanskaya, "The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals," *Journal of Speech, Language, and Hearing Research,* vol. 50, pp. 940-967, 2007.
[13] Q. Cai, & M. Brysbaert, "SUBTLEX-CH: Chinese word and character frequencies based on film subtitles," *PLoS ONE,* vol. 5, no. 6, pp. e10729, 2010.
[14] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer (Version 5.4.2) [Computer program]," 2015.
[15] D. Bates, M. Mächler, B. M. Bolker, and S. C. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software,* vol. 67, no.1, pp. 1-48, 2015.
[16] R Core Team (2015). R: A language and environment for statistical computing. In *R Foundation for Statistical Computing*. Vienna, Austria.
[17] R. V. Lenth, "Least-squares means: The R package lsmeans," *Journal of Statistical Software, vol. 69*. no. 1, pp. 33, 2016.
[18] S. Dufour and R. Peereman, "Lexical competition in phonological priming: Assessing the role of phonological match and mismatch lengths between primes and targets," *Memory & Cognition,* vol. 31, no. 8, pp. 1271-1283, 2003.
[19] M. Radeau, J. Morais, and A. Dewier, "Phonological priming in spoken word recognition: Task effects," *Memory & Cognition,* vol. 17, no. 5, pp. 525-535, 1989.
[20] L. M. Slowiaczek and M. Hamburger, "Prelexical facilitation and lexical interference in auditory word recognition," *Journal of Experimental Psychology: Learning, Memory, and Cognition,* vol. 18, no. 6, pp. 1239-1250, 1992.