INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Phonetic Detail Encoding in Explaining the Size of Speech Planning Window

Shan Luo

Yangzhou University, China
shanluo@yzu.edu.cn

Abstract

With the ultimate goal of understanding the production planning scope, this study manipulates phonetic information (place of articulation and voicing) and measures three acoustic cues to analyze consonant clusters across words produced by English (L1) and Mandarin (L2) speakers. We continue to explore a) how phonetic detail interacts with prosodic boundary in modulating surface realization, and b) the roles of phonetic information in speech planning motor control. The results show that L2 speakers exhibited different acoustic deviations varying with their proficiency level. The group with lower L2 proficiency significantly deviated from the L1 group in release likelihood and closure shortening, while the higher-proficiency group exhibited less nativelike performance in terms of closure durations. The results also discover that all speakers are subject to language-independent articulatory constraint at word boundaries, while language-specific phonetic detail accounts for more nonnative deviations. The core findings highlight a long-distance speech planning scope in native speech, with cross-word phonetic information interacting with prosodic encoding. It is argued that phonology applies blindly across words and is independent of lexical cognitive load.

Index Terms: speech planning, consonant clusters, phonetic detail

1. Introduction

Phonological variability across word boundaries has been vigorously explored as its implementation manifests phonetics-phonology and syntax-phonology interface [1,2,3,4]. Previous research has assumed that phonetic encoding is automatic while there is very little agreement about how messages are incrementally structured when planning upcoming words [5,6,7,8,9]. Following previous research [5,6], the core assumption in this study is that the minimal unit speakers prefer to plan should be a phrasal chunk. This argument would entail that phonology applies blindly across word boundaries if speakers treat two words as a chunk. If so, the low-level phonetic detail of C2 in C1#C2 across words would affect the surface coproduction patterns of the whole cluster when planning ahead.

Taking the core assumption one step further, if segment conditions are held constant and phonotactically legal according to the native grammar, the blind application of phonology is also expected, even in less familiar words. That is, differing degrees of cognitive load (e.g., real vs. nonsense phrases) may affect the degree of planning, but it should not affect the size of the planning window.

This study focuses on English stop-stop cross-word clusters as extensive literature on cluster coproduction already exists [3,10,11,12,13]. The most consistent result involves the place of articulation (POA) of C1 and C2 in stop-stop sequences. Previous research has discovered that in most stop-stop sequences, either within words or across words, the closure of C1 is not released until the closure of C2 is formed. More importantly, existing analyses have found a place order effect in stop clusters; that is, more C1 releases and more gestural overlap in front-back than in back-front clusters [3,10,12,13]. We then can compare our results at word boundaries to those reported previously, which would answer the question of whether speech planning has executed across words.

The current study extends an earlier study [14] on consonantal gesture coarticulation in English (L1) and Mandarin (L2) speakers in a number of directions. Firstly, the current study includes a new L2 group with lower English proficiency, which would allow us to determine whether inter-group similarity comes from higher L2 proficiency or is grounded on language-independent articulatory constraints. Secondly, this study examines three phonetic cues to better understand which acoustic feature(s) could systematically differentiate L1 and L2 speech, while previous comparisons focused on closure overlap and did not find much group difference. Thirdly, the current study examines two effects of phonetic detail (i.e., POA and voicing) on English and Mandarin speakers' production of heterosyllabic clusters. Of the two phonetic effects, voicing is not employed in the native language of L2 (Mandarin in this case) [15]. As all Mandarin stops are voiceless but English stops utilize voicing contrast, most research to date has focused on Mandarin speakers' acquisition of word-final clusters or coda consisting of voiced stops [16,17,18,19]. No study has examined the voicing effect on heterosyllabic cluster production. This research will fully examine the effects of phonetic detail of the upcoming word on cross-word planning, and provide us with a broader view of stop coarticulation under interacting phonetic and prosodic (e.g., boundary) effects. With the comparison between the two phonetic effects, the results in this study will also likely contribute to a general explanatory speech model that accounts for both cross-linguistic commonalities and language-specific details. More importantly, findings of this research could offer a better understanding of speech planning scope [5,6,8,9].

2. Experiment

2.1. Research questions

This study examines the speech planning window through phonetic output of C1#C2 across words. The phonetic reduction is measured through three acoustic features: C1's release likelihood (see the burst energy in Figure 1), closure duration of C1#C2 (see the highlighted duration in Figure 1), and C1#C2's closure overlap (measured by how much closure duration is shortened in C1#C2 coarticulation compared to when C1 and C2 were produced individually; see [3, 14]).

  http://dx.doi.org/10.21437/Interspeech.2019-1412

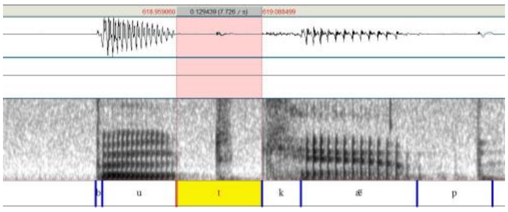

Figure 1. *A release example*

We hypothesize that the beginning of the second word will already be phonologically encoded at the time when the first word is planned, and thereby phonetic detail of C2 would affect the three phonetic cues. Otherwise, the effect of locality planning would be expected in that the external sandhi patterns would not depend on C2.

Hypothesis One: C2's Place of Articulation (POA) would affect release likelihood of C1 in C1#C2, closure duration of C1#C2, and the closure overlap of C1#C2.

Hypothesis Two: C2's voicing would affect these three cues.

2.2. Participants

In this study, participants included a group of 15 native Canadian-English speakers (NE, 9F, 6M, mean = 26.3Y), a group of 25 native Mandarin speakers (NNH, 14F, 11M, mean = 27.4Y, living in Canada, higher English proficiency), and another group of 12 native Mandarin-speaking college students (NNL, 12F, mean = 19.4Y, living in China, lower English proficiency). All NE speakers were born and raised in British Columbia, Canada, to control for regional dialect influence. The Mandarin speakers were selected and grouped based on their English test scores (e.g., IELTS), accent rating (2-min monologue, judged by native speakers), and years of living in an English-speaking country. All Mandarin speakers spoke Standard Mandarin, a language that prohibits stop coda or stop voicing contrast [15]. No speakers reported speech or hearing disorders.

2.3. Stimuli

The data consisted of all possible combinations of stop-stop clusters (C1#C2, C= /b, p, k, t, d, g/). This also gave us all the voicing combinations (e.g., *keep pace*, *lap dance*, *bad kid*, *good game*). Meanwhile, a list of nonce phrases with similar vowel environments was designed to correspond to the real lexicon (e.g., *keep pac*e vs. *peep pate*).

Each participant was instructed to produce 144 phrases (6 C1 * 6 C2* 2 Compound conditions * 2 Lexical conditions) that were embedded in four dialogues. Participants were asked to read each question (i.e., the prompt) silently, then produce the given answer to elicit natural speech instead of simply repeating the carrier sentences. Stimuli were presented in random order in E-Prime. The NE and NNH groups were recorded in the Speech Research Lab in the Department of Linguistics at the University of Victoria, Canada. The NNL group was recorded at the Speech, Perceptual and Cognitive Lab at Yangzhou University, China.

2.4. Analysis

A total of 29952 (576 * 52 participants) sound files were acoustically analyzed using Praat [20]. There were 45 missing tokens, and any phrase with a silence of 350 ms or more in C1#C2 was defined as disfluent (also see [3]). Based on this criterion, 1125 tokens were excluded (3.8% of the total collected). All the remaining data (28782 sound files) were coded by the author, and 1.7% of the tokens were additionally coded by a trained Linguistics researcher. Interrater reliability was 96.1%.

The three cues—release likelihood, closure shortening ratio and closure duration—served as the dependent variables. They were analyzed using one generalized linear mixed-effects model, predicting release likelihood; and two linear mixed-effects models, predicting closure overlap and closure duration. The mixed-effects regression was run in R [21] using the lme4 package [22]. For all the three variables, fixed effects included Group (NE, NNH, NNL), POA (Homorganic, Front-Back, Back-Front), C2's Voicing (voiceless C2 vs. voiced C2), Lexical (real words vs. nonwords), and all interactions of the above. For all three models, we had the maximal by-subject random intercepts and by-subject random slopes for POA, Voicing and Lexical. Posthoc analyses were conducted using the lsmeans package [23] with a tukey correction for multiple comparisons..

3. Results

3.1. Results summary for the three phonetic cues

The overall results of the three phonetic measurements for each group in real words and in nonwords are summarized in the Table 1 below. As the table shows, the NE group released C1 more often in nonwords than in real words, and produced longer closure durations and smaller closure shortening ratios (i.e., more closure overlap). It can also be seen that the NNH group did not differentiate closure shortening between real words and nonwords, and the NNL group did not differentiate C1 release in the two lexical conditions.

Table 1: *Results summary*

Group/ Lexical	**Release Percentage**		**Shortening Ratio**		**Closure Duration (ms)**	
	Real	Non	Real	Non	Real	Non
NE	32.5%	40.6%	58%	64%	144	160
NNH	29.9%	36.9%	71%	80%	168	188
NNL	66.2%	70.8%	92%	101%	166	181

On average, the NNL group had the highest C1 release percentage (68.5%), followed by the NE group (36.6%), then the NNH group (33.4%). The results found that the NNL group released C1s significantly more often than the NE group (*$\beta = 1.96$, $se = 0.54$, $z = 3.65$, $p = 0.000267$).* The differences between the two groups were significant in Realwords ($\beta = 2.07$, $se = 0.39$, $z = 5.26$, $p < 0.0001$) and in Nonwords ($\beta = 1.92$, $se = 0.43$, $z = 4.4$, $p = 0.0002$*)*. This result indicates that the L2 group with lower proficiency deviated significantly from the L1 group in terms of releasing C1.

As for the closure shortening ratio (the smaller the value, the more the closure overlap), the NE group shortened 61% of closure durations (i.e., 39% closure overlap) in C1#C2, the NNH group shortened 76% (i.e., 24% closure overlap) and the NNL group produced 97% (i.e., 3% closure overlap), compared to when C1 and C2 were produced without coarticulation. The closure ratio differences between NNL and NE were significant ($\beta = 0.3$, $se = 0.1$, $t = 2.88$, $p = 0.00578$*)*, which were consist across lexical conditions (Realwords: $\beta =$

0.33, se = 0.09, t = 3.52, p = 0.011; Nonwords: *β =0.36, se = 0.1, t = 3.43, p = 0.0142*). In terms of closure duration, the NNH group produced longer closure durations than the NE group, and the difference was marginally significant (*β = 19, se = 10.3, t = 1.87, p = 0.068*). This deviation of the NNH group mainly came from nonwords (*β = -29.39, se = 8.9, t = -3.3, p = 0.0208*), and was mitigated by lexical frequency in real words (*β = -24.4, se = 8.9, t = -2.7, p = 0.088*).

Overall, the NE group produced consistent phonetic output across lexical conditions. The NNL group with lower L2 proficiency significantly deviated from the L1 group in C1 release likelihood and closure shortening, while the NNH group with higher proficiency exhibited less nativelike performance in terms of closure durations and in nonwords.

3.2. Hypothesis 1

3.2.1. POA effect on release likelihood

To examine the place of articulation (POA) effect on cross-word coarticulation, we firstly expect the place order effect (more C1 releases in front-back than in back-front clusters) would be shown in both real words and nonwords. The overall POA effect on the release likelihood is plotted in Figure 2 below, which shows a similar release trend for all groups. This trend is visible in that they all had the highest release percentage in front-back clusters, followed by back-front clusters, then homorganic clusters. Although the NNL group did not differentiate front-back and back-front clusters as much as the other two groups did and released C1s significantly more often than the NE group across POA clusters (all *p < 0.04*), their performance was still in line with this trend.

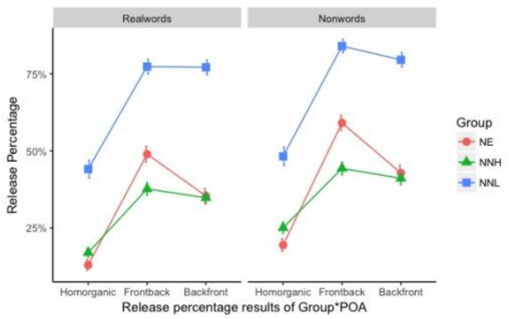

Figure 2. *Release likelihood summary of Group*POA (95% CI error bar)*

As predicted, the NE group released significantly more often in front-back than in back-front clusters in both real words (*β = 0.72, se = 0.2, z = 3.5, p = 0.0142*) and nonwords (*β = 0.85, se = 0.2, z = 4.1, p = 0.0011*). In contrast, neither of the L2 groups distinguished the two types of clusters under either lexical condition (all *p > 0.9*).

This result directly supports our prediction, meaning that both C1 and C2's place of articulation affected C1's release likelihood in native speech. More importantly, such an effect was strongly shown in both familiar words and nonsense words in native English speakers' execution, suggesting consistent external sandhi patterns without speakers being subject to extra cognitive load. However, we should note that both L2 groups kept in line with the release trend (frontback > backfront > homorganic clusters), even in nonwords. This finding suggests that although the POA effect does not influence the L2 groups as much as the it does on the L1 group, the L2 groups were also conditioned by the place of articulation of C2 when releasing C1.

3.2.2. POA effect on durational cues

The effect of POA on closure overlap and closure duration was summarized in Table 2 below. As the table shows, all the three groups had slightly smaller closure shortening ratios (i.e., more overlap) in frontback than in backfront clusters, but no group statistically differentiated the two types of clusters with closure overlap (all *p > 0.05*).

Table 2: *Closure duration (Dur.) and closure overlap (Ratio) results summary under POA*

Group/ POA	**Homorganic**		**Frontback**		**Backfront**	
	Dur.	Ratio	Dur.	Ratio	Dur.	Ratio
NE	156	64%	147	60%	155	62%
NNH	181	75%	176	75%	182	77%
NNL	173	96%	169	93%	180	101%

However, all groups produced significantly shorter closure durations in frontback than in backfront clusters (all *p < 0.0014*). Both NE and NNL groups extended this durational pattern in real words and nonwords (all *p < 0.0211*). The NNH group performed similarly in real words (*p < 0.0001*) but not in nonwords (*p = 0.3*).

To summarize, the results found that POA of C1 and C2 significantly affected release percentage and closure durations for the three groups in the same fashion. All groups exhibited highly comparable patterns with more C1 releases, shorter closure durations and more overlap in frontback than in backfront clusters. Although the place order effect was only significantly shown in L1 speakers' release patterns, the L2 groups followed the same trend. Moreover, the NNL group produced similar closure duration patterns to the native group across a range of lexical frequencies, which was quite unexpected.

3.3. Hypothesis 2

3.3.1. Voicing effect on release likelihood

On average, all groups released a voiceless C1 more often than a voiced C1 (NE: 40.4% vs. 32.7%; NNH: 37.4% vs. 29.2%; NNL: 69.6% vs. 67.3%). To examine the prediction of the voicing effect on phonetic output of C1#C2, we assume that C1's release likelihood is influenced by C2's voicing.

Table 3: *Results summary of C2's voicing*

Group/C2	**Release Percentage**		**Closure Duration (ms)**		**Shortening Ratio**	
	VC	VD	VC	VD	VC	VD
NE	39.3%	33.7%	148	158	62%	62%
NNH	33.8%	32.8%	172	188	75%	76%
NNL	67.9%	69%	163	185	97%	97%

As Table 3 summarizes, all groups released C1 more often when C2 was voiceless than when it was voiced. The analyses found that the difference was significant only in the NE group

(β = *0.31, se = 0.07, z = 4.47, p = 0.0001*). More importantly, this pattern was shown in both real words (β = *0.29, se = 0.09, z = 3.3, p = 0.0122*) and nonwords (β = *0.33, se = 0.09, z = 3.75, p = 0.0024)*. In contrast, neither of the L2 groups was influenced by C2's voicing (both $p > 0.9$). This finding again supports that the native speakers paid simultaneous attention to the C2's phonetic detail when releasing C1, while L2 groups were affected solely by the "local" voicing of C1.

3.3.2. Voicing effect on durational cues

According to Hypothesis 2, voicing is expected to influence closure duration and closure overlap of C1#C2. The statistical analyses found all groups produced significantly longer closure durations of C1#C2 when C2 was voiced than when C2 was voiceless in both real words and nonwords (all $p < 0.0001$).

However, the post-hoc analyses found different responses to the voicing effect in L2 groups. When C2 was voiced, NNH group produced significantly longer closure durations than NE group ($\beta = 30$, $se = 8.84$, $t = 3.37$, $p = 0.017$). When C2 was voiceless, the difference was marginally significant ($p = 0.081$). The cognitive load also worsened this nonnative deviation. The closure duration differences between NNH and NE groups were significant in nonwords across C2's voicing conditions (voiceless C2: $\beta = 26.2$, $se = 9$, $t = 2.92$, $p = 0.055$; voiced C2: $\beta = 32.5$, $se = 9.1$, $t = 3.59$, $p = 0.0091$). On the other hand, the NNL group mainly deviated from the NE group in terms of larger closure ratios (i.e., less closure overlap) irrespective of C2's voicing (both $p < 0.02$).

3.4. Summary

To recapitulate, the results reported various acoustic consequences in external sandhi of C1#C2 made by learners at different L2 proficiency levels. The NNH group with higher L2 proficiency produced significantly longer closure durations than the NE group, especially under the voicing effect in less familiar words. The NNL group with lower proficiency exhibited consistent differences in both release and closure overlap measurements across voicing and lexical conditions.

In examining our hypotheses, the results discovered that only the L1 group attended to both C1 and C2's phonetic information. This can be seen in how the native group displayed significant release patterns under POA, as well as the fact that C1's release was determined by C2's voicing. It directly indicates that L1 group extended speech planning scope to both C1 and C2 at phrase edges, while L2 groups demonstrated local planning.

4. Discussion

As a continuing effort to explore the phonetics-phonology interface, the current study examines heterosyllabic consonant cluster coordination produced by English and Mandarin speakers. The results found that POA effect significantly affected release likelihood and closure duration, and the effect was quite similar on all groups. In this study, the fact that even L2 speakers with lower proficiency exhibited similar release trends suggests that a constriction transition from front to back articulators is strongly favored and articulatorily grounded even in a boundary position where gestures are relatively loosely coordinated [24]. This also suggests that the effect of POA, a language-independent articulatory constraint, seems to be universal on speakers with different language backgrounds. Similarly, the results also reported that all groups out-released voiceless C1s compared to voiced ones. This observation of inter-group similarity likely reflects the articulatory conflict involved in sustaining voicing word-finally [25,26], and provides strong support for the integration of articulatory constraints into L2 models of language development [27].

However, the results also observed systematic nonnative deviations under the voicing effect, which links voicing to more predictive powers when compared to POA. In this study, both L2 groups deviated significantly from the native norm in terms of durational cues of voicing contrast. The finding corroborates previous results suggesting that closure duration is an essential and reliable acoustic means to differentiate L1 and L2 speech [19]. This difficulty is likely due to the fact that voicing as the phonetic detail is not employed in the native language of the L2 speakers in the experiment.

Further, the current study highlights an important finding in examining external sandhi: C2's phonetic detail significantly affected the surface realization of the whole cluster in the native speech. It suggests that the L1 group realized cross-word coarticulation while the L2 groups' production was subject to the local word-final segment. Moreover, these results demonstrate that detailed phonetic information in long-distance, or at least cross-word, coarticulation is encoded prior to motor execution in native speech [28, 29].

More importantly, the patterns in L1 speech were consistently shown across lexical frequency in this study. It strongly supports that external sandhi applies blindly across word boundaries and is independent of lexical cognitive load in planning and execution. We thus argue that phonetic encoding does not entirely rely on stored exemplars. Rather, it is a highly sensitive process that is stored in native grammar and can be applied to pseudo-lexicon, if the lexicon conforms to native phonological rules. Interestingly, language-specific phonetic encoding is executed locally in L2 speech. This encourages further examination of more fine-grained differences in coarticulation, and whether localized planning of phonetic information extends into perception in L2 speech.

5. Conclusions

The core findings in this study indicate that automatic phonetic encoding interacts with prosodic encoding in native speech's long-distance planning scope whereas locality of production planning is shown in nonnative speech. Contrary to the view that speakers prefer a minimal planning window that is less likely to cross boundaries [8,9], we argue the planning scope extends across words since speakers treat them as a phrasal chunk.

The major theoretical implication of this study is that prosodic structure not only provides a framework in which phonetic detail is nested, but also involves simultaneous encoding with phonetic information. We hope this study fills an important and underexplored niche in the phonetics-phonology interface and calls further attention to the role of phonetic detail in fully describing and explaining boundary-modulated variations and advanced speech planning.

6. Acknowledgements

The work is funded by Jiangsu Provincial Department of Education Grant (2018SJA1163) and "LvYangJinFeng" Talent Plan (YZLYJFJH2017YB131).

7. References

[1] D. Byrd and S. Choi, "At the juncture of prosody, phonology, and phonetics—The interaction of phrasal and syllable structure in shaping the timing of consonant gestures," *Laboratory Phonology*, vol. 10, pp. 31–59, 2010.

[2] T. Cho, "Prosodic boundary strengthening in the phonetics–prosody interface," *Language and Linguistics Compass*, vol. 10, no. 3, pp. 120–141, 2016.

[3] E. C. Zsiga, "Articulatory timing in a second language," *Studies in Second Language Acquisition,* vol. 25, no. 3, pp. 399–432, 2003.

[4] M. Tamminga, "Modulation of the following segment effect on English coronal stop deletion by syntactic boundaries," *Glossa: A Journal of General Linguistics*, vol. 3, no. 1, pp. 1–27, 2018.

[5] W. J. Levelt, A. Roelofs, and A. S. Meyer, "A theory of lexical access in speech production," *Behavioral and brain sciences*, vol. 22, no. 1, pp. 1–38, 1999.

[6] M. Wagner, "Locality in phonology and production planning," *McGill Working Papers in Linguistics,* vol. 22, pp. 1–18, 2012.

[7] J. Tanner, M. Sonderegger, and M. Wagner, "Production planning and coronal stop deletion in spontaneous speech," *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 8, no. 1, pp. 1–39, 2017.

[8] O. Kilbourn-Ceron, M. Wagner, and M. Clayards, "The effect of production planning locality on external sandhi: a study in /t/. In *Proceedings of the 52nd Annual Meeting of the Chicago Linguistic Society*, 2016.

[9] L. Wheeldon, "Producing spoken sentences: The scope of incremental planning," In Perrier, P. and Verlag, P. L. (Eds.,) *Cognitive and physical models of speech production, speech perception, and production-perception integration,* pp. 93–114, 2013.

[10] D. Byrd, "A phase window framework for articulatory timing," *Phonology,* vol. 13, no. 2, pp. 139–169, 1996.

[11] L. Davidson, "Characteristics of stop releases in American English spontaneous speech," *Speech Communication*, vol. 53, no. 8, pp. 1042–1058, 2011.

[12] J. B. Henderson, and B. H. Repp, "Is a stop consonant released when followed by another stop consonant?" *Phonetica*, vol. 39, no. 2-3, pp. 71–82, 1982.

[13] E. C. Zsiga, "Phonetic alignment constraints: consonant overlap and palatalization in English and Russian," *Journal of Phonetics*, vol. 28, no. 1, pp. 69–102, 2000.

[14] S. Luo, "Gestural overlap across word boundaries: Evidence from English and Mandarin speakers," *Canadian Journal of Linguistics*, vol. 62, no. 1, pp. 56–83, 2017.

[15] S. Duanmu, *The phonology of standard Chinese*. Oxford University Press, 2007.

[16] T. Bent, A. R. Bradlow, and B. L. Smith, "Production and perception of temporal patterns in native and non-native speech," *Phonetica,* vol. 65, no. 3, pp.131–147, 2008.

[17] E. Broselow, S. I. Chen, and C. L. Wang, "The emergence of the unmarked in second language phonology," *Studies in Second Language Acquisition,* vol. 20, no. 2, pp.261–280, 1998.

[18] R. Hayes-Harb, B. L. Smith, T. Bent, and A. R. Bradlow, "The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts," *Journal of Phonetics,* vol. 36, no. 4, pp. 664–679, 2008.

[19] X. Xie, and C. A. Fowler, "Listening with a foreign-accent: The interlanguage speech intelligibility benefit in Mandarin speakers of English," *Journal of Phonetics,* vol. 41, no. 5, pp. 369–378, 2013.

[20] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," Version 6.0.25, http://www.praat.org/, 2017.

[21] R Core Team, "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2018.

[22] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software,* vol. 67, no. 1, pp. 1-48, 2015.

[23] R. V. Lenth, "Least-Squares Means: The R Package lsmeans," *Journal of Statistical Software*, vol. 69, no. 1, pp. 1–33, 2016.

[24] L. Davidson, and K. Roon, "Durational correlates for differentiating consonant sequences in Russian," *Journal of the International Phonetic Association,* vol. 38, no. 2, pp. 137–165, 2008.

[25] L. Davidson, "Variability in the implementation of voicing in American English obstruents," *Journal of Phonetics,* vol. 54, no. 1, pp. 35–50, 2016.

[26] J. Westbury, and P. Keating, "On the naturalness of stop consonant voicing," *Journal of Linguistics,* vol.22, no.1, pp. 145–166, 1986.

[27] L. Colantoni, and J. Steele, "Integrating articulatory constraints into models of second language phonological acquisition," *Applied Psycholinguistics,* vol. 29, no. 3, pp. 489–534, 2008.

[28] P. Keating, "Phonetic encoding of prosodic structure," *Speech production: Models, Phonetic Processes, and Techniques*, pp. 167–186, 2006.

[29] T. Cho, Y. Lee, and S. Kim, "Communicatively driven versus prosodically driven hyper-articulation in Korean," *Journal of Phonetics*, vol. 29, no. 3, pp. 344–361, 2011.