Effects of L1 prosody on segmental contrast in L2: The case of English stop voicing contrast produced by Korean speakers

Jiyoun Choi, Sahyang Kim, and Taehong Cho

Citation: The Journal of the Acoustical Society of America **139**, EL76 (2016);
View online: https://doi.org/10.1121/1.4942021
View Table of Contents: http://asa.scitation.org/toc/jas/139/3
Published by the Acoustical Society of America

Effects of L1 prosody on segmental contrast in L2: The case of English stop voicing contrast produced by Korean speakers

Jiyoun Choi
Hanyang Phonetics and Psycholinguistics Lab, Hanyang University, Haengdang-dong 17, Seongdong-gu, Seoul 133-791, Korea
jiychoi@hanyang.ac.kr

Sahyang Kim
Department of English Education, Hongik University, Sangsu-dong, Mapo-gu, Seoul 121-791, Korea
sahyang@hongik.ac.kr

Taehong Cho[a)]
Hanyang Phonetics and Psycholinguistics Lab, Department of English Language and Literature, Hanyang University, Haengdang-dong 17, Seongdong-gu, Seoul 133-791, Korea
tcho@hanyang.ac.kr

Abstract: This study investigated how the L1 phonetics-prosody interface transfers to L2 by examining prosodic strengthening effects (due to prosodic position and focus) on English voicing contrast (*bad*-*pad*) as produced by Korean vs English speakers. Under prosodic strengthening, Korean speakers showed a greater *F*0 difference due to voicing than English speakers, suggesting that their experience with the *macroprosodic* use of *F*0 in Korean transfers into L2. Furthermore, Korean speakers produced voiced stops with low *F*0 and short voice onset time as English speakers did, although such a cue pairing is absent in Korean, showing dissociation of cues from L1 segments for L2 production.

Date Received: October 14, 2015 **Date Accepted:** January 24, 2016

1. Introduction

In the past decades, numerous studies have demonstrated that the phonetic realization of speech segments is fine-tuned by the prosodic structure of the utterance in which segments occur. For example, spatial and/or temporal expansion of segments (known as *prosodic strengthening*) has been found to occur in important landmarks of the prosodic structure such as in phrase-initial position and in prominent (stressed/accented) syllables (see Cho, 2011 for a review), possibly enhancing language-specific phonetic features (e.g., de Jong, 2004; Cho and McQueen, 2005).

Despite the growing awareness of the essential role of the phonetics-prosody interface in L1 speech production, our understanding of the L2 phonetics-prosody interface is quite limited. A large number of studies on L2 speech production have indeed vigorously illuminated how L1 phonetic knowledge transfers to L2 speech on the segmental level (e.g., Flege, 2003; Best and Tyler, 2007) or on the suprasegmental level (e.g., Munro, 1995; Trofimovich and Baker, 2006). But only a few studies have systematically investigated the interplay between the two levels (cf. Davidson, 2011). The present study therefore explores how speakers' experience with the phonetics-prosody interface in their native language may permeate into their L2 production with a view to understanding the nature of L1–L2 phonetic transfer that is attributable to the higher-order prosodic structure of L2 speakers' native language. To this end, the present study investigates the relationship between prosodic strengthening and stop voicing contrast in English (e.g., *bad*–*pad*) with special reference to prosodic strengthening that arises with prosodic position (phrase-initial/phrase-medial) and prominence (focused/unfocused) in L1 and L2, as produced by native American English (NAE) speakers vs native Korean (NK) learners of English.

[a)]Author to whom correspondence should be addressed

As an important phonetic cue to the initial stop voicing contrast, the present study examines variation of *F*0 of the following vowel since it serves as a useful testbed for exploring the phonetics-prosody interface: *F*0 not only functions as a cue to stop voicing contrast (higher for voiceless than for voiced stops, e.g., Kingston and Diehl, 1994) but it is also subject to intonational modification due to the prosodic structure of individual languages. In this regard, comparing L1 and L2 speech production by English vs Korean speakers is particularly interesting. In English, the *F*0 cue to stop voicing contrast does not reorganize the intonational structure, and thus the effect remains at the microprosody level. In Korean, on the other hand, the segmentally-induced *F*0 perturbation is "phonologized" at the macroprosody level, and is therefore integrated into the intonational phonology of the language. As a result, the initial syllable of the Accentual Phrase (AP) or of the Intonational Phrase (IP) is associated with a categorically low tone when the initial segment is a lenis stop, but a categorically high tone when the segment is an aspirated or a fortis stop (Jun, 2005). Such a categorical tonal contrast occurs only at the phrase-initial position, and disappears in the non-initial position as it is overridden by the intonational phonology which stipulates phrasal-internal tonal patterns within an AP (Jun, 2005). (See Cho *et al.*, 2002, for the Korean stop contrasts.)

Such cross-linguistic difference leads the present study to test a hypothesis that the macroprosodic use of *F*0 in L1 (Korean) transfers into L2 (English), so that the *F*0 difference due to stop voicing will be much greater in L2 speech (by NK) than in L1 speech (by NAE), reflecting the categorical difference in the macroprosodic vs the microspodic use of *F*0 in Korean vs English. This hypothesis builds on previous studies on the interaction between segment and higher-order phonological structure which showed the effects of word or syllable position and phonotactic restrictions on segmental realization (see Davidson, 2011, for a review). We will call this hypothesis the *macroprosody transfer hypothesis*. Crucially, the effect is expected to arise *only* when the prosodic condition is met—i.e., in a phrase-initial position with which the macroprosodic use of *F*0 is associated in Korean. Furthermore, given that focus in Korean is often realized by starting a new prosodic phrase with the focused word (Jun, 2014), the L1 use of *F*0 is more likely to transfer to L2 in the focused than in the unfocused condition.

Along with the segmentally-induced *F*0 cue, we also examine voice onset time (VOT). In Korean, the fortis stop in phrase-initial position is produced with a short VOT (comparable to a phrase-initial voiced stop in English), and both the lenis and aspirated are produced with a relatively longer VOT (comparable to a phrase-initial voiceless stop in English). But in phrase-medial position, it is the lenis stop (not the fortis stop) that becomes phonetically voiced intervocalically, comparable to the voiced stop in English in that position, while the fortis and the aspirated stops remain voiceless. We test whether NK speakers use the VOT cue paired with the *F*0 cue as they do when producing their L1 Korean stops in phrase-initial position (e.g., High *F*0–Long VOT for the aspirated, Low *F*0–Long VOT for the lenis, High *F*0–Short VOT for the fortis).

One might hypothesize that the use of paired cues to a particular segment type in L1 (Korean) transfers into L2 at the "segmental" level. If so, NK speakers would produce the initial English voiced stop either with Low *F*0–Long VOT (mapped onto the Korean lenis stop in terms of *F*0, but mismatched for VOT) or with High *F*0–Short VOT (mapped onto the Korean fortis stop in terms of VOT, but mismatched for *F*0), depending on which one of the two cues is weighted more (cf. Schertz *et al.*, 2015). (Note that the English voiceless stop is unequivocally mapped onto the Korean aspirated stop in the use of both cues.) In this case, NK speakers would find it difficult to produce English voiced stops with a paring of Low *F*0–Short VOT in phrase-initial position as none of the Korean stops in this position is produced with such a pairing. Alternatively, if *F*0 and VOT cues available for stop contrast in Ll transfer to L2 independently at the "phonetic" level, NK speakers would be able to dissociate the cues from the segmental representation in L1 and use the cues independently in L2. In such a case, NK speakers would be able to use any combination of *F*0 (High or Low) and VOT (Long or Short), including the non-native Low *F*0–Short VOT pairing for the voiced stop as produced by NAE speakers in phrase-initial position.

2. Method

2.1 Participants and recording

Twelve native speakers of American English (6 male, 6 female, aged 21–31) and 12 native speakers of Korean (6 male, 6 female, aged 21–28) participated in the study. The Korean speakers were all college students with English proficiency at a low-intermediate level (with an average TOEFL score of 74.5, which is around the

34 percentile according to a 2014 ETS report). They studied English at the elementary and secondary schools, and were exposed to a variety of spoken English materials for listening comprehension tests. At college, they have been exposed to English through EFL classes. The speech data were recorded in a soundproof booth at the Hanyang Phonetics and Psycholinguistics Lab with a Tascam HP-P2 digital recorder and a SHURE KSN44 microphone at a sampling rate of 44.1 kHz.

2.2 Speech materials and procedure

Four minimal pairs of CVC English words were used: *bad–pad*, *bed–ped*, *bat–pat*, *bet–pet*. As in Table 1, each target word was placed in a pair of carrier sentences where two experimental factors of Position [Intonational Phrase-initial (IPi) vs Intonational Phrase-medial (IPm)] and Focus (focused vs unfocused) were manipulated. The target words occurred either in initial or medial position of an IP, in one of the two focus conditions: phonologically contrastive (narrow) focus (*bad* vs *pad*), or no focus.

In the experiment, the speakers saw each pair of sentences on a computer screen with contrastive words written in uppercase in red. They then heard the question (A) which was pre-recorded by a female native speaker of American English, and answered it with the sentence (B) as in Table 1. The entire set of pairs was repeated three times in a randomized order. Out of 2304 tokens (8 target words $\times$ 2 boundaries $\times$ 2 focus types $\times$ 3 repetitions $\times$ 24 speakers), 5 tokens that were produced with a deviated prosodic rendition were discarded.

2.3 Measurements and statistical analyses

VOT of the target bilabial stop and *F*0 of the following vowel were measured. VOT included both voicing lag and voicing lead. Voicing lead included not only negative VOT (voicing initiates during the closure and continues into the release) but also a "bleed" pattern (voicing continues from the preceding vowel and dies out before the release) and a "hump" pattern (voicing initiates in the middle of the closure and dies out before the release) (see Davidson, 2016). *F*0 was measured at the vowel onset (V-Onset, 20 ms after the initial pulse) to test the microscopic local perturbation effect, and at the midpoint (V-Mid) and at the three-quarters point of the vowel (V-3Q) to test the macroscopic prosodic effect. *F*0 values were normalized across speakers, using *Z*-scores. Repeated measures (RM) analysis of variances (ANOVAs) on VOT and normalized *F*0 values (averaged over repetitions and items) were performed using SPSS v21, with Voicing, Position, and Focus as within-subject factors and (Native) Language as a between-subject factor. Only significant effects and interactions are reported below.

3. Results and discussion

3.1 F*0*

All three vowel points (V-Onset, V-Mid, and V-3Q) showed significant main effects of Voicing ($F[1,22] = 508.8$, 205.4, and 205.4, respectively, all at $p < 0.001$; /*p*/ > /*b*/), Position ($F[1,22] = 85.2$, 91.9, and 91.9, respectively, all at $p < 0.001$; IPi > IPm), and Focus ($F[1,22] = 123.1$, 226.3, and 226.3, respectively, all at $p < 0.001$; focused > unfocused). This indicates that the *F*0 effects are not limited to local perturbations (at V-Onset), but that they spread pervasively into the later parts of the vowel as evident in significant voicing-induced *F*0 differences at V-Mid and V-3Q in Fig. 1. What is also clearly noticeable from Fig. 1 is that in all cases, NK speakers produced /*b*/ with low *F*0, matched with the Korean lenis /*p*/ in terms of the *F*0 cue. There was no main effect of Language at any vowel point, nor was there any two-way interaction involving Language.

Language effects, however, were evident in three-way interactions. The Voicing $\times$ Position $\times$ Language interaction [Fig. 1(a)] did not reach significance at

Table 1. Sample sentences of a target word *bad*. Focused words are in uppercase letters and the target word (*bad*) is underlined.

IP-initial	Focused	A: Did you write "PAD fast again"? B: Not exactly. "<u>BAD</u> fast again" was what I wrote.
	Unfocused	A: Did you write "bad SLOWLY again"? B: Not exactly. "<u>bad</u> FAST again" was what I wrote.
IP-medial	Focused	A: Did you write "say PAD fast again"? B: No, I wrote "say <u>BAD</u> fast again".
	Unfocused	A: Did you write "say bad SLOWLY again"? B: No, I wrote "say <u>bad</u> FAST again".

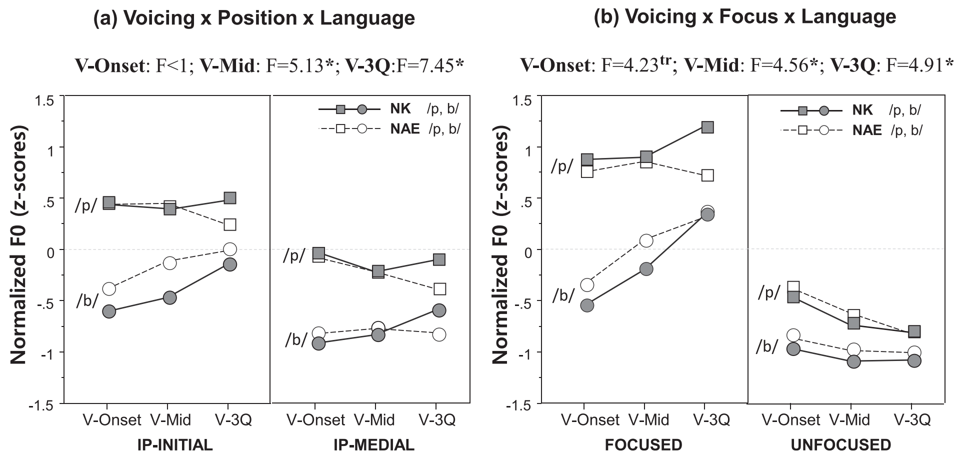

Fig. 1. Three-way interactions on *F*0 involving the Language factor: (a) Voicing × Position × Language and (b) Voicing × Focus × Language. "*" refers to $p < 0.05$, and "tr" to $p < 0.08$.

V-Onset, but it emerged at V-Mid and V-3Q. The interaction stemmed from the fact that only in IP-initial position, the *F*0 difference due to voicing was substantially greater for NK than for NAE (as confirmed by eta-statistics; at V-Mid, NK, $F[1,11] = 144.4$, $\eta^2 = 0.92$ NAE, $F[1,11] = 34.8$, $\eta^2 = 0.79$; at V-3Q, NK, $F[1,11] = 67.7$, $\eta^2 = 0.86$ vs NAE, $F[1,11] = 21.7$, $\eta^2 = 0.66$). This suggests that the NK speakers maintain the *F*0 difference throughout the entire vowel while the NAE speakers show localized perturbation. In other words, there was a language-specific modulation of the Voicing effect on *F*0 (larger for NK) as a function of prosodic position, which was manifested only at a stronger prosodic position (IPi), where a robust cross-linguistic difference in the phonetics-prosody interface exists between Korean and English.

A similar language-specific modulation of the Voicing effect was found in a three-way interaction that involved focus [i.e., Voicing × Focus × Language interaction; Fig. 1(b)]. It reached significance at all three vowel points [see Fig. 1(b) for a statistical summary]. The interaction stemmed from the fact that only when focused, the *F*0 difference due to voicing was substantially greater for NK than for NAE speakers at all three vowel points (as confirmed by eta-statistics; V-Onset: NK, $F[1,11] = 288.8$, $\eta^2 = 0.96$ vs NAE, $F[1,11] = 75.5$, $\eta^2 = 0.87$; V-Mid: NK, $F[1,11] = 144.4$, $\eta^2 = 0.93$ vs NAE, $F[1,11] = 53.3$, $\eta^2 = 0.83$; V-3Q: NK, $F[1,11] = 53.5$, $\eta^2 = 0.93$ vs NAE, $F[1,11] = 17.2$, $\eta^2 = 0.61$). These interaction patterns again indicate that the Voicing effect on *F*0 was modulated by the speakers' native language (larger for NK) in a strong prosodic environment, but this time under focus.

3.2 VOT

RM ANOVA yielded significant main effects of Voicing, Position, and Focus ($F[1,22] = 296.7$, 34.7, and 50.7, respectively, all at $p < 0.001$), while there was no main effect of Language ($F[1,22] < 1$). Crucially, as shown in Fig. 2, Voicing interacted with Language, Position, and Focus, respectively, reaching a significant two-way interaction in each case.

First, the Voicing by Language interaction [Fig. 2(a)] stemmed from the fact that the Voicing effect on VOT was larger (more polarized) by NAE than by NK (*/b/-/p/* mean diff., NAE 78.6 vs NK, 58.9 ms; $F[1,11] = 158.5$, $\eta^2 = 0.94$, $F[1,11] = 139.8$, $\eta^2 = 0.93$, respectively). As shown in Fig. 2(a), the greater polarization by NAE was attributable in part to the fact that they often produced */b/* with voicing lead (−8.7 ms), while NK produced */b/* as phonetically voiceless unaspirated (3.2 ms) which is largely matched with the Korean fortis */p*/* (rather than the lenis */p/*) at least in terms of VOT. (Recall that */b/* was better matched with the lenis */p/* as far as *F*0 was concerned.)

Second, the Voicing by Position interaction came from an asymmetric voicing effect on VOT as a function of prosodic position [Fig. 2(b)]: the Voicing effect was larger IP-medially than IP-initially. That is, VOT for */b/* was longer (positive) IP-initially, but shorter (negative) IP-medially (IPi 9.1 vs IPm, −14.9 ms; $F[1,22] = 44.07$, $p < 0.001$, $\eta^2 = 0.67$), whereas VOT for */p/* remains unaffected by Position (IPi 66.2 vs IPm 65.8 ms; $F[1,22] < 1$). Furthermore, there was a three-way interaction (Voicing × Position × Language) as a trend ($F[1,22] = 3.25$, $p = 0.086$). As can be inferred from Fig. 2(b), the trend was in part due to the tendency that NAE speakers produce */b/*

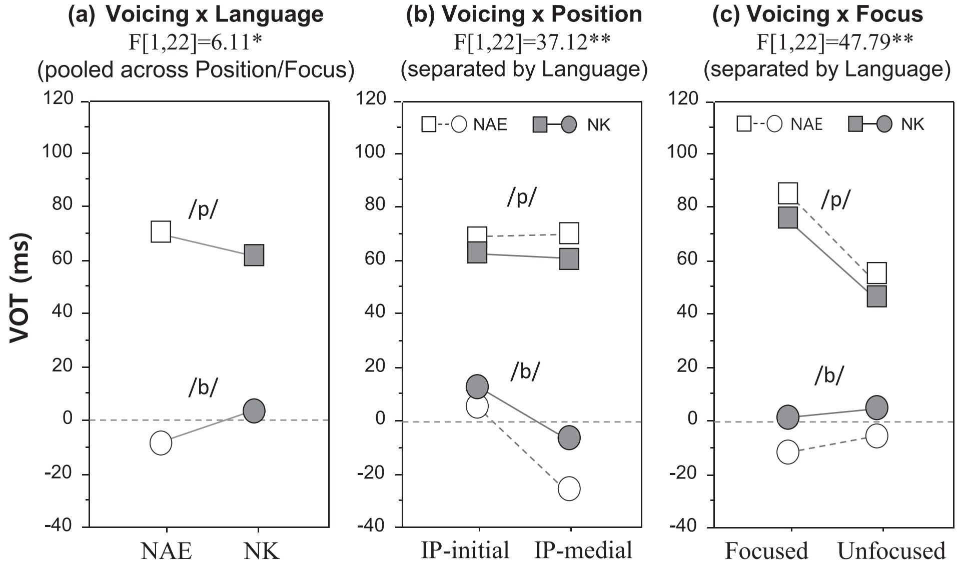

Fig. 2. Voicing (/*b*/-/*p*/) contrast in terms of VOT as a function of its interaction with (a) Language (NAE vs NK), (b) Position (IP-initial vs IP-medial), and (c) Focus (focused vs unfocused) "*" refers to $p < 0.05$, and "**" refers to $p < 0.005$.

with greater voicing lead than NK speakers in IP-medial position. The phrase-medial suppression of voicing lead for NK again suggests that NK speakers map English /*b*/ onto the Korean fortis /*p**/ (in terms of VOT) rather than onto the Korean lenis /*p*/, even though the latter is often produced with intervocalic voicing in a phrase-medial position, phonetically similar to English /*b*/ in that position. There was no further significant interaction that involves Voicing.

Third, the Voicing by Focus interaction [Fig. 2(c)] was due to the fact that the /*b*/-/*p*/ contrast was enhanced mainly by virtue of an increase in VOT for /p/ when focused vs unfocused ($F[1,22] = 84.2$, $p < 0.001$, $\eta^2 = 0.79$), while only a negligible focus-induced decrease in VOT was found for /*b*/ ($F[1,22] = 2.03$, $p > 0.1$, $\eta^2 = 0.09$). The three-way interaction (Voicing × Focus × Language) was not found, suggesting a comparable pattern for both NAE and NK.

Finally, the pattern of voicing for /*b*/ did not differ much between NAE and NK speakers. For IP-initial /*b*/, NAE speakers produced less than 5% of the tokens with voicing before the release, and NK speakers less than 3% of the tokens. IP-medially, both groups produced /*b*/ mostly with the bleed pattern for more than 75%, and about 7% of the tokens with full voicing during closure.

4. General discussion

One of the most important findings is that NK and NAE speakers are similar in terms of their use of *F*0 to mark initial stop voicing contrast in English (higher for the voiceless, lower for the voiced), but they are dissimilar in terms of the time course of the impact on *F*0. NK speakers show a pervasively categorical use of *F*0 throughout the entire vowel whereas NAE speakers show a relatively gradient use of *F*0 with the *F*0 difference being attenuated toward the end of the vowel. Most crucially, this language effect arises only in prosodic strengthening contexts which condition the language-specific (macroprosodic) use of *F*0, i.e., in phrase-initial position and under prominence. Recall that the use of *F*0 for stop contrast in Korean is integrated into the intonational phonology with *F*0 being either categorically High or Low for the aspirated/fortis vs the lenis in the phrase-initial position (the macroprosody), whereas English uses *F*0 largely as a cue to the stop voicing contrast at the segmental level (the microprosody). This finding therefore supports the *macroprosody transfer hypothesis*, suggesting that the L1 (Korean) use of the macroprosody in connection with higher-order prosodic structure transfers into L2 even though the use of prosody in L2 (English) is largely restricted to a lower-order segmental level.

Another important finding is that NK speakers produce English /*b*/ with lower *F*0 being mapped onto the Korean lenis stop, but with short VOT being mapped onto the Korean fortis stop in both phrase-initial and phrase-medial positions. In other words, despite the fact that the pairing of Low *F*0–Short VOT is hardly made for any Korean stop at the segmental level, particularly never in phrase-initial position, NK speakers produce the English voiced stop with such a pairing (Low *F*0–Short VOT), although the phonetic detail is further fine-tuned by L1 experience. This implies that, rather than engaging in one-to-one mapping between L1 and L2 phonetic categories at the segmental level, NK speakers (with low-intermediate English proficiency) dissociate the two phonetic cues from their segmental anchorage and use them independently in marking L2 phonological contrast. Moreover, the NK speakers' use of the two cues makes reference to higher-order prosodic structure in which segments occur: NK speakers enhance the VOT difference due to voicing particularly under focus and in phrase-medial position (which allows for negative VOT for voiced stops); and they use the *F*0 cue robustly when focused and in phrase-initial position just like NAE speakers do, although the level of *F*0 used is different between NK and NAE speakers (macroprosodic vs microprosodic).

In conclusion, the present study has made one step further toward understanding how L1 experience with the phonetics-prosody interface influences L2 production. The L1 transfer reflecting effects of higher-order prosodic structure on phonetic implementation was evident in the transfer of the L1 (Korean) macroprosodic use of *F*0 into L2 (English) in which *F*0 for stop contrast is used at the level of microprosody. Furthermore, it was found that the use of *F*0 and VOT cues could be dissociated from their segmental anchorage in L1, so that they can participate independently in marking L2 phonological contrast by making reference to the phonetics-prosody interface in both L1 and L2. These findings may also account for why Korean listeners show individual variation in cue weighting between *F*0 and VOT for voicing contrast in English (Schertz *et al.*, 2015), and one-to-many segmental mappings between English voiced stops and Korean stops (Park and de Jong, 2008). This also has further implications for theories of L2 speech (e.g., Flege, 2003; Best and Tyler, 2007; Davidson, 2011) which have focused on the effects of L1 experience on L2 production and perception without making explicit reference to the interplay between segments and prosodic structure. It is thus hoped that the present study has provided an impetus for further development of theories of L2 speech production by taking into account the interactions between the L1 and L2 phonetics-prosody interface.

Acknowledgments

The authors thank their RAs, Yuna Baek and Jiyoun Jang, for their assistance in acoustic measurements. This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (Grant No. NRF-2013S1A2A2035410).

References and links

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production*, edited by M. J. Munro and O.-S. Bohn (John Benjamins, Amsterdam), pp. 13–34.

Cho, T. (**2011**). "Laboratory phonology," in *The Continuum Companion to Phonology*, edited by N. C. Kula, B. Botma, and K. Nasukawa (Continuum, London/New York), pp. 343–368.

Cho, T., Jun, S., and Ladefoged, P. (**2002**). "Acoustic and aerodynamic correlates of Korean stops and fricatives," J. Phon. **30**, 193–228.

Cho, T., and McQueen, J. M. (**2005**). "Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress," J. Phon. **33**, 121–157.

Davidson, L. (**2011**). "Phonetic and phonological factors in the second language production of phonemes and phonotactics," Lang. Linguistics Compass **5**, 126–139.

Davidson, L. (**2016**). "Variability in the implementation of voicing in American English obstruents," J. Phon. **54**, 35–50.

de Jong, K. J. (**2004**). "Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration," J. Phon. **32**, 493–516.

Flege, J. (**2003**). "Assessing constraints on second-language segmental production and perception," in *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and N. Schiller (Mouton de Gruyter, Berlin), pp. 319–355.

Jun, S.-A. (**2005**). "Korean intonational phonology and prosodic transcription," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, edited by S.-A. Jun (Oxford University Press, Oxford), pp. 201–229.

Jun, S.-A. (**2014**). "Prosodic typology: By prominence type, word prosody, and macro-rhythm," in *Prosodic Typology II: The Phonology of Intonation and Phrasing*, edited by S.-A. Jun (Oxford University Press, Oxford), pp. 520–540.

Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Language **70**, 419–454.

Munro, M. J. (**1995**). "Nonsegmental factors in foreign accent: Ratings of filtered speech," Stud. Second Lang. Acquis. **17**, 17–34.

Park, H., and de Jong, K. J. (**2008**). "Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects in second language identification skills," J. Phon. **36**, 704–723.

Schertz, J., Cho, T., Lotto, A., and Warner, N. (**2015**). "Individual differences in phonetic cue use in production and perception of a non-native sound contrast," J. Phon. **52**, 183–204.

Trofimovich, P., and Baker, W. (**2006**). "Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech," Stud. Second Lang. Acquis. **28**, 1–30.