Phonetic encoding of coda voicing contrast and its interaction with information structure in L1 and L2 speech

Jiyoun Choi[a], Taehong Cho[b], Sahyang Kim[c], Yuna Baek[b] & Jiyoung Jang[b]

[a]Hanyang Phonetics & Psycholinguistics Lab, Hanyang University, Seoul; [b]Hanyang Phonetics & Psycholinguistics Lab, Dept. of English Language & Literature, Hanyang University, Seoul; [c]Hongik University, Seoul
jiychoi@hanyang.ac.kr; tcho@hanyang.ac.kr; sahyang@hongik.ac.kr; yunab@hanyang.ac.kr; jiyoungljang@hanyang.ac.kr

ABSTRACT

This study examines the coda voicing effect in English on production of preceding vowel by comparing L1 and L2 speech. Particular attention is paid to how coda voicing contrast is encoded in temporal *vs.* spectral dimensions, how the effect interacts with information structure (with varying focus types), and how the use of information structure differs in L1 *vs.* L2. Both English (native) and Korean (non-native) speakers show a comparable pattern in the temporal dimension (vowel duration), but Korean speakers fail to encode the spectral attribute (vowel F1/F2) due to coda voicing. Furthermore, while native speakers make use of information structure in an efficient way that is accountable by the output-oriented and the system-oriented constraints, non-native speakers do not use information structure as efficiently. Alongside the L1/L2 issues, a ‘long-distance’ coda voicing effect is observed on VOT of the preceding initial stop, which illuminates the nature of the coda voicing effect.

Keywords: coda voicing, focus, L1, L2, English, Korean, vowel lengthening, VOT, formant

1. INTRODUCTION

The coda voicing effect that vowel duration is longer before a voiced than before a voiceless stop occurs in multiple languages ([1],[2],[3]). While the commonly recurring pattern may have originated from physiological and biomechanical properties of speech, the effect may be encoded in the phonological system of the language, such that a language like English (and also German) shows a more robust coda voicing effect on vowel duration than other languages (e.g., Arabic and Catalan) ([4],[5]). The coda voicing contrast, at least in English, is further manifest in the fine-phonetic detail of the formant structure of the preceding vowel, showing a decrease in F1 and an increase in F2 due to coda voicing ([6],[7]). But a language like Korean does not exhibit the coda voicing effect at all as consonantal laryngeal contrasts in Korean are completely neutralized to be an unreleased voiceless stop in the coda position (cf. [8]).

The present study examines how speakers of Korean who do not experience the same coda voicing contrast in their first language encode the L2 contrast in both the temporal and the spectral (formant) dimensions of the vowel, and how their L2 phonetic encoding pattern compares to the way that it is encoded by native speakers of English.

We take this question to be exploratory, so that no specific hypotheses are formulated here. But considering the differences of the sound system between Korean (L1) and English (L2), one might envisage a few possibilities on how native and non-native (Korean) speakers may differ in encoding the coda voicing contrast. For example, given that Korean does not employ the voicing contrast in the coda position, Korean learners of English may find it relatively hard to learn the way that it is phonetically encoded by native speakers. This is particularly likely under the assumption that the coda voicing effect in English is internalized or ‘phonologized’ in the grammar of the language ([4],[5]).

More specifically, however, Korean speakers are expected to do better in encoding coda voicing in the temporal dimension than in the spectral dimension for some independent reasons. First, the coda voicing effect on the preceding vowel duration is a near-universal tendency (e.g., [5]) and such a universally-applicable phonetic attribute may be learned relatively easily. Second, languages like Arabic and Czech employ phonological vowel-length (quantity) contrast, which impose constraints on the temporal realization of the vowel (thus showing no robust coda voicing effect on the vowel) ([4],[5],[9]). But there is no such functional demand from phonology of Korean, which would otherwise hinder learning the temporal effect. Third, although Korean does not make the same coda voicing contrast, the voicing-induced vowel lengthening effect is not entirely new as the vowel in Korean lengthens before a phonetically voiced lenis stop in an intervocalic VCV context [10]. These universal and language-specific features taken together suggest that the coda voicing effect on the vowel duration may be learned quite easily and therefore reflected in L2 speech by Korean learners. On the other hand, Korean learners of English often fail to make spectral distinctions between English vowels (e.g., tense *vs*. lax vowels or /ɛ/ *vs*. /æ/) [11], resonating with the fact that Korean has a smaller vowel system than English does (cf. [12]). Thus, Korean speakers' sensitivity to spectral differences may not be as refined as English speakers, and therefore they may not effectively use or learn fine-grained spectral differences due to coda voicing.

Another important question to be addressed is how the coda voicing effect interacts with focus in L1 and L2. It has already been observed that native speakers of (American) English exaggerate the durational difference of the preceding vowel due to coda voicing when the voicing contrast is emphasized under focus [13]. We extend this to L2 production by Korean learners of English. Following [13], two types of focus are employed—i.e., lexical focus (LexFoc, in which semantic contrast between words was emphasized, as in *bed-chair*), and phonological focus (PhFoc, in which coda voicing contrast was emphasized, as in *bed-bet*). Furthermore, in order to disentangle the focus effect from a potential effect of boundary strength (e.g., [14],[15]), prosodic boundary is added as another factor. With these additional factors, the present study explores how Korean learners of English make reference to higher-order information structure (as reflected in the focus information) in signalling the coda voicing contrast in comparison with how native speakers do. It will be particularly interesting to understand how phonetic encoding of the coda voicing contrast is modulated by information structure in a way to reflect the output-oriented *vs*. system-oriented constraint (contrast maximization *vs*. effort minimization principle) ([16]) in L1 *vs*. L2 speech. This study will therefore add new empirical data on this understudied aspect (i.e., the interplay of phonetics, phonology, and higher-order linguistic structure) to the L2 phonetics literature, informing existing theories of L2 phonetics (e.g., [17],[18], [19],[20])

Finally, alongside testing the coda voicing effects in L1 and L2, the present study will also test a possible 'long-distance' effect of coda voicing on VOT of the initial (onset) stop beyond the preceding vowel. Various accounts have been advanced regarding the underlying mechanism for the coda voicing effect in both perceptual and articulatory terms (e.g., [1], [3], [5], [21]). Whatever the underlying mechanism might be, if the effect has been internalized in the articulatory system of the language, the timing of the entire vocalic gesture may be subject to modification due to the coda voicing. Under this possibility, VOT of the syllable initial consonant before a vowel is also expected to be lengthened (a 'long distance' coda voicing effect), given that the onset of VOT coincides with the onset of articulatory vocalic gesture. Testing this possibility will therefore illuminate the nature of the coda voicing effect with implications for theories of phonetic encoding of coda voicing.

2. METHOD

2.1. Participants, materials and procedure

Ten native speakers of American English (5 females and 5 males, M_{age} = 29 years, range: 23 – 34 years) and 16 native speakers of Korean (8 females and 8 males, M_{age} = 24 years, range: 22 – 29 years) participated in the experiment. Korean participants were undergraduate and graduate students whose English proficiency was ranging from an intermediate to an advanced level (with the average TOEFL score of 90.66, which is around 66.5% in percentile rank, according to 2013 ETS report). (Koreans' English proficiency may reveal more information about the difference within the non-native group, but it was not included as a factor as it is tangential to the purpose of the present study.)

Table 1: Sample sentences of a target word 'bed'. Focused words are in uppercase letters.

IP	LexFoc	A: Did you write 'CHAIR fast again'? B: Not exactly. 'BED fast again' was what I wrote.
	PhFoc	A: Did you write 'BET fast again'? B: Not exactly. 'BED fast again' was what I wrote.
	Unfoc	A: Did you write 'bed SLOWLY again'? B: Not exactly. 'bed FAST again' was what I wrote.
Wd	LexFoc	A: Did you write 'say CHAIR fast again'? B: No, I wrote 'say BED fast again'.
	PhFoc	A: Did you write 'say BET fast again'? B: No, I wrote 'say BED fast again'.
	Unfoc	A: Did you write 'say bed SLOWLY again'? B: No, I wrote 'say bed FAST again'.

Four minimal pairs of English monosyllabic CVC *words* were used as in (1):

(1) *bed–bet*, *bad–bat*, *ped–pet*, *pad–pat*

Within each pair, items varied only in the voicing of final stop (/d/ *vs.* /t/) with the initial (onset) consonant being either /b/ or /p/, and the vowel being either /ɛ/ or /æ/. Each of the eight target words (in four pairs) was embedded in carrier sentences, which consist of a pair: a prime sentence (A in Table 1) and a test sentence (B in Table 1). The test sentence contained the target words. The prime sentence was always a question, used to induce the intended focus types for the target words. The target word was placed in phrase-initial position (IP) or in phrase-internal position (Wd), and was either lexically focused (with lexical contrast between the target word and a word in the prime sentence, e.g., *BED-CHAIR*), phonologically focused (with phonological contrast on the coda voicing contrast, e.g., *BED-BET*) or unfocused (with lexical contrast falling on the following word).

Prior to the experiment, all prime sentences were recorded by a female native speaker of American English (23 years old). She produced the sentences one by one, while placing focus on intended words. In the experiment, speakers were presented with both prime and test sentences on a computer screen, while two focused words (one from a prime and the other from a test sentence) were highlighted in red colour. The pre-recorded prime sentence was played, and then the speaker produced the test sentence as an answer to the question. The speech data were recorded in a soundproof booth with a Tascam HP-P2 digital recorder and a SHURE KSN44 microphone at a sampling rate of 44 kHz.

2.2. Measurements and statistical analyses

We measured vowel duration, VOT of the onset stop and the first (F1) and second formant (F2) at the midpoint of the vowel. These four measures plus F2-F1 were included as dependent variables in statistical analyses. For statistical evaluations of the results on the acoustic measures, repeated measures Analyses of Variance were carried out with Coda Voicing (voiced/voiceless), Boundary (IP/Wd), Focus (LexFoc/PhFoc/Unfoc) as within-subject factors and Native Language (ENG/KOR) as a between-subject factor. (Note that VOT for the initial stop was examined only with /p/, as it turned out that /b/ was produced with very short VOT with no variation.)

3. RESULTS

We will limit our report to the directly relevant results that are only to the purpose of the present study—i.e., main effects of Coda Voicing and its interaction with other factors.

3.1. Vowel Duration and VOT

Vowel duration. There was a significant main effect of Coda Voicing (F[1,24]=66, p<.01), so that vowel duration was longer before a voiced than before a voiceless stop. Coda Voicing did not significantly interact with other factors, except that there was a three-way interaction between Coda Voicing, Focus and Boundary (F[2,48]=4.4, p<.05). This interaction was due to the fact that the coda voicing effect on vowel duration was reliable in all conditions except for the unfocused condition for the native English speaker group (p>.1) (see Fig.1a).

VOT (only for the initial /p/). There was a significant main effect of Coda Voicing (F[1,24]=19.4, p<.01). For both language groups, VOT was significantly longer before a voiced than before a voiceless stop, similar to what has been found for the vowel duration (Fig.1b). This Coda Voicing effect did not interact with any other factors.

Fig. 1: Effects of Coda Voicing on (a) Vowel duration and (b) VOT for the native English and the non-native Korean speakers.

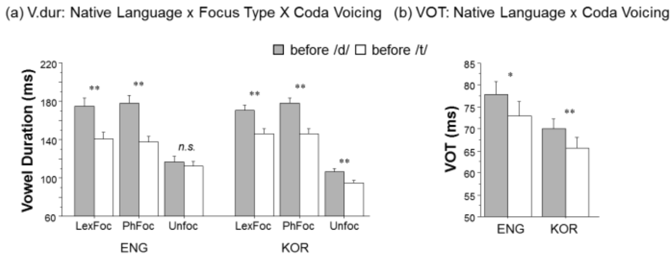
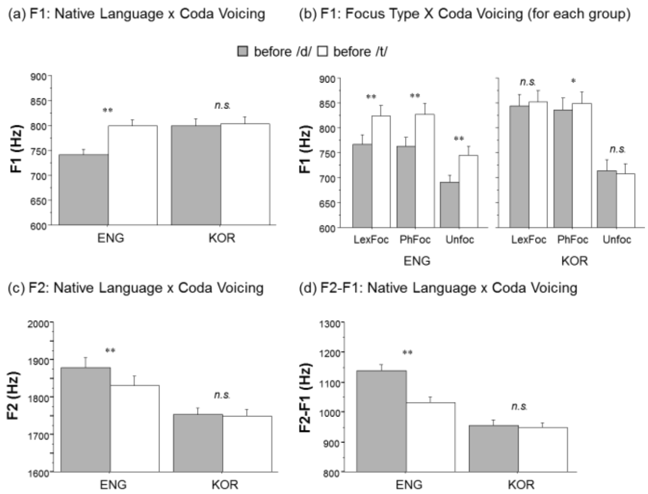


Fig. 2: Effects of Coda Voicing on (a, b) F1, (c) F2 and (d) F2-F1 for the native English and the non-native Korean speakers.



3.2. F1, F2, and F2-F1

F1. F1 showed a significant main effect of Coda Voicing (F[1,24]=52.8, p<.01): F1 was lower before a voiced than before a voiceless stop, showing that the coda voicing raised the non-high vowel (/ɛ/ or /æ/) in the acoustic vowel space. The Coda Voicing effect, however, significantly interacted with Native Language (F[1,24]=38.6, p<.01) and Focus (F[2,48]=5.1, p<.05). As can be seen in Fig. 2a, the interaction between Coda Voicing and Native Language stemmed from the fact that the F1 lowering effect before a voiced stop was significant only by native English speakers (ENG: F[1,9]=57.9, p<.01; KOR: F[1,15]<1).

The Coda Voicing by Focus interaction was analyzed for each group of Native Language. (Although there was no significant three-way interaction between Coda Voicing, Focus and Native Language, we decided to analyze the data separately for each language group, given that there was a two-way interaction between Coda Voicing and Native Language, and given that some noteworthy patterns indeed emerged from the separate analyses as reported below.) As shown in Fig. 2b, for English speakers, the F1 lowering effect due to coda voicing was consistently observed across focus types, whereas Korean speakers showed the coda voicing effect on F1 only in a phonologically focused condition (F[1,15]=5.2, p<.05).

F2. There was a significant main effect of Coda Voicing **(**F[1,24]=29.2, p<.01): F2 was higher (thus the vowel was more fronted in the acoustic vowel space) before a voiced than a voiceless stop. Coda Voicing interacted with Native Language (F[2,48] = 4.7, p<.05), but not with Focus. As can be seen in Fig. 2c, the interaction was attributable to the coda voicing effect being significant only with the native English speakers (ENG: F[1,9]=39.6, p<.01; KOR: F[1,15]<1).

F2-F1. There was a significant main effect of Coda Voicing (F[1,24]=62.3, p<.01). F2-F1 was larger before a voiced than before a voiceless stop. Again, as was the case with F1 and F2, there was an interaction between Coda Voicing and Native Language (F[1,24]=44.8, p<.01). The interaction, as shown in Fig. 2d, was due to the Coda Voicing effect on F2-F1 being significant only for the native English speakers (ENG: F[1,9]=59.4, p<.01; KOR: F[1,15]=1.23, p=.29).

4. DISCUSSION

One of the basic findings of the present study is that Korean and English speakers showed a similar coda voicing effect on the preceding vowel in the temporal dimension (longer before a voiced stop). This suggests that Korean learners of English encode the coda voicing contrast successfully in the temporal dimension in a way comparable to the native speakers' phonetic encoding, despite the fact that Korean speakers do not have their first language experience with the coda voicing effect. However, as for the spectral dimension, native speakers showed a robust effect on F1 and F2 of the preceding non-high front vowel (positioning it being lower and fronter in the vowel space), whereas Korean learners failed to encode the spectral attribute of coda voicing.

As was discussed at the outset of the paper, Korean speakers' successful encoding of the coda voicing contrast in the temporal dimension may be due to the universally-applicable tendency of the coda voicing effect which may be easily learned (and thus easily reflected in L2 speech). Alternatively, it may also have to do with Korean speakers' first language experience. Korean does not employ phonological vowel-length (quantity) contrast which would otherwise impose constraints on temporal realization of the vowel ([4],[5],[9]), and Korean lengthens the vowel before a phonetically voiced stop in an intervocalic context [10]. This experience may facilitate learning the coda voicing effect. Similarly, the failure of encoding the coda voicing contrast in the spectral dimension is also attributable to their first language experience. Since Korean has a smaller vowel system, Korean speakers are likely to be less sensitive to fine-grained spectral differences due to coda voicing, which may not be easily learned. (This does not mean that they are entirely unaware of the spectral attributes. They showed the effect on F1 under phonological focus, indicating that learning spectral attributes would simply require more attention and effort.)

Another important finding comes from how information structure (with different focus types) is used by native English speakers *vs.* non-native Koran speakers. In encoding the coda voicing contrast in the temporal dimension, the native English speakers exaggerated the coda voicing effect under focus (with no difference between phonological and lexical focus) (cf. [13]), but they did not signal the coda voicing contrast in the unfocused condition—i.e., when the target word was no longer the locus of information (carrying 'old' information). This implies that native speakers of English make use of information structure in a communicatively efficient way, so that they 'hyperarticulate' (driven by an output-oriented constraint) when necessary, but make no extra effort to generate redundant information (driven by a system-oriented constraint) (e.g., [16]). Korean learners, on the other hand, consistently lengthened the vowel duration before a voiced stop regardless of whether the target word was focused or not. This suggests that non-native (Korean) speakers did not make full use of information structure as efficiently as the native English speakers do. This also has implications for theories of L2 phonetic learning (e.g., [17],[18],[19],[20]) in that the fine-grained phonetic encoding of phonological contrast in L2 cannot be fully understood and therefore modelled adequately without taking into account its interplay with higher linguistic structure (e.g., information structure and prosodic structure).

Finally, aside from the L1/L2 issues on the coda voicing contrast, the 'long-distance' coda voicing effect on VOT of the preceding stop (i.e., longer VOT before a voiced stop) was observed with both native English and non-native Korean speakers. Given that the onset of VOT coincides with the onset of the vocalic gesture, this result suggests that the coda voicing effect is not simply a local acoustic effect for enhancing the coda voicing contrast (e.g., [21]) or due to the local modulating of speed of constriction formation (e.g., [1]), but rather it is encoded in the timing of the vocalic gesture which may be better understood as a global effect in the articulatory dynamical system (e.g., [22],[23]).

5. REFERENCES

[1] Chen, M. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22, 129-159.

[2] Lisker, L. 1974. On "explaining" vowel duration variation. *Glossa* 8, 233-246.

[3] Maddieson, I. 1997. Phonetic universals. In: Laver, J., Hardcastle, W. J. (eds.), *The Handbook of Phonetic Science*. Oxford: Blackwells, 619-639.

[4] Keating, P. A. 1985. Universal phonetics and the organization of grammars. In: Fromkin, V.A. (ed), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Orlando FL: Academic Press, 115–132.

[5] Cho, T. 2015. Language effects on timing at the segmental and suprasegmental levels. In: Redford, M.A. (ed), *The Handbook of Speech Production.* Wiley-Blackwell., 505-529.

[6] Moreton, E. 2004. Realization of the English postvocalic [voice] contrast in F1 and F2. *Journal of Phonetics* 32, 1-33.

[7] Kluender, K.R., Lotto, A.J., Jenison, R.L. 1995. Perception of voicing for syllable-initial stops at different intensities: does synchrony capture signal voiceless stop consonants? *J. Acousti. Soc. Am.* 97, 2552-2567.

[8] Cho, T., McQueen, J. Phonological versus phonetic cues in native and non-native listening: Korean and Dutch listeners' perception of Dutch and English consonants. *J. Acoust. Soc. Am.* 119(5), 3085-3096

[9] de Jong, K. J., & Zawaydeh, B. A. 2002. Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics* 30, 53-75.

[10] Cho, T. 1996. Vowel correlates to consonant phonation: An acoustic-perceptual study of Korean obstruents. M.A. Thesis, University of Texas at Arlington.

[11] Ingram, J. & S. Park. 1997. Cross-language vowel perception and production by Japanese and Korean listeners of English. *Journal of Phonetics* 25, 343-370.

[12] Cho, T., Jun, S.-A., Jung, S., Ladefoged, P. 2001. The vowels of Cheju. *Korean Journal of Linguistics* 26(4), 801-819.

[13] de Jong, K. 2004. Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics* 32, 493-516.

[14] Cho, T., Keating, P. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37, 466-485.

[15] Cho, T., Lee, Y., Kim, S. 2014. Prosodic strengthening on the /s/-stop cluster and the phonetic implementation of an allophonic rule in English. *Journal of Phonetics* 46, 128-146.

[16] Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H & H theory. In: Hardcastle, W.J., Marchal, A. (eds), *Speech Production and Speech Modeling*. Dordrecht: Kluwer Academic Publishers, 403-440.

[17] Best, C. T. 1995. A direct realist perspective on cross-language speech perception. In: Strange, W. (ed), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. Timonium, MD: York Press, 167–200.

[18] Best, C. T., Tyler, M. D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: Bohn, O.-S., Munro, M. J. (eds), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, Philadelphia: John Benjamins, 13–34.

[19] Flege, J. E. 1995. Second language speech learning: Theory, Findings, and Problems. . In: Strange, W. (ed), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. Timonium, MD: York Press, 233–277.

[20] Flege, James E., Munro, Murray J. & McKay, Ian R. A. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16, 1–26

[21] Kluender, K. R., Diehl, R. L., & Wright, B. A. 1988. Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16, 153–169.

[22] Goldstein, L., Byrd, D., Saltzman, E. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In: Arbib, M. A. (ed) *Action to language via the mirror neuron system*. Cambridge: Cambridge University Press, 215–249.

[23] Goldstein, L. & Fowler, C.. 2003. Articulatory Phonology: A phonology for public language use. In: Schiller, N. O., Meyer, A. S. (eds), *Phonetics and phonology in language comprehension and production: Differences and similarities*. Berlin & New York: Mouton de Gruyter, 159–207.