Individual differences in L2 learners' perceptual cue weighting patterns

Eun Jong Kong[a], Jan Edwards[b]

[a]Korea Aerospace University, [b]University of Wisconsin-Madison
ekong@kau.ac.kr, jedwards2@wisc.edu

ABSTRACT

This study investigated how English learners of Korean weigh multiple acoustic cues, namely VOT and *f0*, in perceiving the Korean stop laryngeal contrast. This question is of interest because VOT is more central to this contrast in English relative to Korean. Our goal was to examine whether L2 cue-weighting strategies were influenced by individual differences in cue-weighting strategies in L1 and by individual learner's L2 proficiency. The phoneme identification tasks [English /t/ vs. /d/; Korean /t/ vs. /t^h/] with 20 English learners showed that listeners who relied more on VOT in English relied less on *f0* in Korean. Further, listeners who relied more on *f0* in English in an ambiguous VOT condition also relied more on *f0* in Korean. L2 proficiency was a significant predictor of both of these relationships. These results suggest that both attention to secondary cues in L1 and L2 proficiency influence cue-weighting strategies in L2.

Keywords: perceptual cue weighting, stop laryngeal contrast, voice onset time, fundamental frequency

1. INTRODUCTION

Phonetic contrasts are typically realized by multiple acoustic cues (e.g., [13]) and the relative importance of these cues in speech perception often differs across languages. For example, in English, voice onset time (VOT, [14]) is the primary cue that differentiates voiceless stops (e.g., /t/ [t^h]) and voiced stops (e.g., /d/ [t]) although secondary cues such as *f0* and F1 cut-back also differ systematically between voiceless and voiced stops [1,6,7,16]. By contrast, in Korean, VOT and *f0* are equally important in distinguishing aspirated stops (e.g., /t^h/) from the other two phonation-type categories, lax and tense stops (e.g., /t/, /t'/) [17].

In the context of second language learning, these language-specific cue-weighting strategies may conflict between the native language (L1) and the target language (L2). There is evidence that L2 learners gradually learn to use the cue weighting strategies of their second language (e.g., [3,4,5]). There are at least two factors that seem to influence L2 perceptual learning: (1) L1 background [4], and (2) L2 experience or L2 proficiency [5,15] (but also see [2]). Given the different cue-weighting strategies for the phonation type contrast in English and Korean, it seems likely that L2 proficiency will influence the cue-weighting strategies of English L1 speakers who are learning Korean. Our particular interest in this paper was to examine whether L2 cue weighting strategies were *also* influenced by individual differences in cue weighting strategies in L1. This interest was inspired by earlier findings that individual variability is observed in L1 perception [9,10,11,18,19] as well as in L2 perception [4]. Several researchers have shown individual differences in cue weighting strategies for perceiving the voicing contrast by L1 English speakers: some listeners are more sensitive to the secondary acoustic cue, *f0*, than other listeners [9,11,19]. We wondered whether these individual differences in L1 cue-weighting patterns might carry over into L2 speech processing. In a similar sense, the current study examines the role of individual L2 proficiency in modulating the degrees of L1-to-L2 transfer on cue weighting. That is, we also examined whether greater L2 experience might help learners to inhibit L1 influence on L2. The specific questions addressed in this study were the following. One, does greater reliance on the redundant acoustic cue of *f0* in L1 result in more attention to *f0* and less attention to VOT in L2 cue weighting? We hypothesized that individuals who relied more on the redundant cue of *f0* in L1 might have an advantage in mastering the phonation-type contrast in L2 Korean where the role of *f0* is as important as VOT. We were also interested in whether the relationship between cue weighting strategies in L1 and L2 was modulated by L2 proficiency. We hypothesized that the effect of L2 proficiency would be found in the individual level analysis in a way that more proficient L2 learners of Korean could suppress their reliance on VOT and rely on *f0* more than less proficient L2 listeners.

2. PERCEPTION EXPERIMENT

2.1. Stimuli

The audio stimuli were synthesized using natural productions. For the English stimuli, a token of /da/ and of /ta/ (from *dot* and *tot*) produced by an adult male native English speaker were used. VOT values

were manipulated by excising a portion of the burst release/aspiration from /ta/ and pasting it before the voicing onset of the /da/ token. Six log-scale values from 9ms to 59ms of VOT were used (9ms, 13ms, 19ms, 28ms, 40ms, 59ms). For each VOT step, the *f0* dimension was manipulated by replacing the original *f0* value during the vowel of /da/ with one of five *f0* values (98Hz, 106ms, 114Hz, 122Hz, 130Hz). This procedure yielded 30 different syllables (six VOT steps × five *f0* steps). The same procedure was used with a /t^ha/ and a /ta/ token (from '탓' [t^hat] and '닷' [tat]) produced by a male Seoul Korean speaker, producing another set of 30 different syllables (six VOT × five *f0* combinations).

2.2. Tasks and Procedure

For the English session, there was a two-alternative forced-choice task and for the Korean session, there was a three-alternative forced-choice task. In the English session, the listeners were asked to decide whether the consonant in the syllable played was /d/ or /t/, and to respond by clicking one of two consonant labels ("d" or "t") shown on the screen. Similarly for the Korean session, listeners were asked to decide whether they heard /t/, /t'/ or /t^h/ and were asked to respond by clicking one consonant labels shown in Hangul ("ㄷ", "ㄸ", "ㅌ") on the screen. Responses of tense consonants for the Korean session were excluded from the analysis.

The English and Korean sessions were counter-balanced across the participants. In each session, a set of 30 different CV syllables was repeated three times in a random order, yielding 90 trials per person.

2.3. Participants

20 English native speakers (F:10, M:10) participated in the experiment. All were living in Seoul at the time of testing; the length of their stay varied from 18 months to 13 years. L2 proficiency was assessed with a short test composed of items (i.e., listening comprehension, vocabulary comprehension, and reading comprehension) from the TOPIK (*Test of Proficiency in Korean*, a test administered by the National Institute of International Education in Korea). We used the score from this test as a measure of L2 proficiency. Participants received monetary compensation for their participation.

2.4. Statistical Analysis

The mixed-effects logistic regression models were constructed separately for the English and the Korean tasks in order to measure listeners' reliance on the two acoustic cues, VOT and *f0,* in perceiving the aspirated stop [t^h] in each language. The dependent variables were the binary categorical responses of "da" and "ta" for the English session and the binary responses of "다" /ta/ (lax) and "타" /t^ha/ (aspirated) for the Korean session. The independent variables were VOT and *f0*. VOT and *f0* values were standardized so that we could directly compare the model coefficients. The model included random intercepts and slopes. Therefore, the random slope coefficients are a measure of the deviation of individual listener' coefficient from the group average (i.e., the fixed effect coefficient). We used these individual slope coefficients to represent as numeric individual listeners reliance on VOT and *f0* [14].

Simple linear regressions were performed to test whether these individual slope coefficients in L1 (i.e., how much individual listeners relied on VOT and *f0* to differentiate voiceless and voiced stops) explained individual differences in cue weighting in L2. Separate regressions were run for the two acoustic parameters (VOT and *f0*) in each language. In each regression model, an additional independent variable (L2 proficiency) was added later to examine whether this multiple regression model outperformed the simple regression model.

3. RESULTS

Fig.1 presents the logistic curves of VOT and *f0* and Table 1 summarizes the parameter estimation of the mixed effects regression models. In both language sessions, the group-averaged curves for VOT and *f0* (black lines) increase as VOT and *f0* values increase, indicating that the perception of phonetically aspirated stop categories was associated with greater VOT and higher *f0*.

In the English session (left panel), the group-averaged curve for VOT variable was observed to switch to /t/ more abruptly than the *f0* curve: β_{VOT} = 5.89, β_{f0} = 1.26. This is consistent with previous finding that VOT is the primary perceptual cue for English listeners [1,6,7,16]. However, it can be observed that the curves of individual listeners (random coefficients, grey lines) show that there are substantial individual differences in L1 cue weighting strategies, especially with respect to *f0*. Some listeners are much more sensitive to *f0* than other listeners.

The regression patterns from the Korean task (L2, right panel) were similar to those from L1 English session in that the group-averaged curves showed a increase as VOT or *f0* values increase, and in that the switch from the lax to the aspirated stop occurred more abruptly in VOT dimension than in *f0* dimension: β_{VOT} = 3.02, β_{f0} = 1.11. Despite these

similarities, the perceptual sensitivity to VOT patterned differently between the two language sessions. The group-averaged curve for VOT in the Korean session showed a less abrupt rise than the curve in the English session, indicating that listeners relied relatively less on VOT in L2 perception. This coefficient difference of VOT between the two languages was statistically significant ($\beta_{VOT:diff}$ = -2.25, S.E. = 0.22, p<.001). Unlike VOT, the *f0* curves between the two languages were not statistically different ($\beta_{f0:diff}$ = -0.15, S.E. = 0.14, p = .263).

Figure 1: Estimated probability of /t/ (L1 English, left panel) and /t^h/ (L2 Korean, right panel) from the mixed-effects logistic models. Black lines indicate group-averaged coefficients, and thin grey lines indicate the individual learners' coefficients.

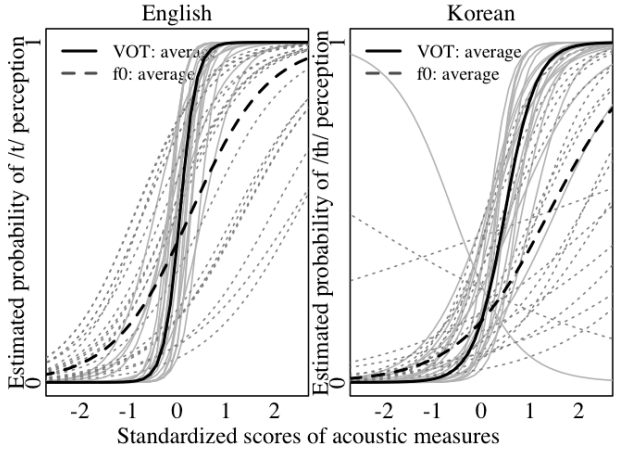


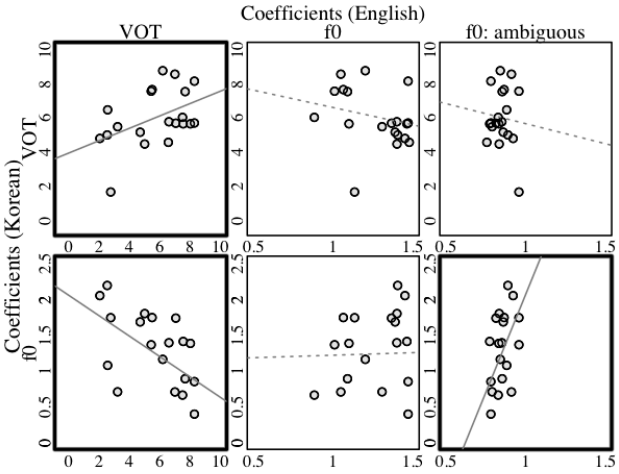
Table 1: The output of the mixed effects models presented in Fig.1

	parameters	estimate	std.err	p-value
English	VOT	5.89	0.69	<.001
	f0	1.26	0.14	<.001
Korean	VOT	3.02	0.49	<.001
	f0	1.11	0.20	<.001

The random coefficients of the individual listeners in L2 from the regression model were regressed against those in L1 as presented in Fig.2. VOT coefficients from L1 were positively correlated with VOT coefficients ß from L2 (β=.36, p=.044, r^2=.16) and they were negatively correlated with $f0_{L2}$ coefficients from L2(β=-.14, p=.024, r^2=.20). The direction of correlation in the leftmost panels suggests that the listeners who relied less on VOT in L1 also relied less on VOT in L2. The listeners who relied less on VOT in L1 were more sensitive to *f0* in L2. Note that the coefficient of VOT_{L2} against VOT_{L1} was smaller than 1, indicating that the use of VOT was less active in L2 than in L1 for all listeners. L2 proficiency improved the model fit (r^2=.34). Individuals with better L2 proficiency relied less on VOT in L2, relative to those with poorer L2 proficiency.

Unlike VOT, reliance on $f0_{L1}$ in L1 was not significantly correlated with either reliance on VOT_{L2} or $f0_{L2}$ in L2 (center panels) [VOT_{L2}: β=-2.03, p=.28; $f0_{L2}$: β=.07, p=.91]. Because *f0* is a secondary cue to voicing in English, we decided to examine the effect of *f0* in ambiguous VOT conditions (the 19 ms and 28 ms VOT conditions combined), as it is in these conditions when listeners should be most sensitive to *f0* (rightmost panels). In this condition, reliance on *f0* in L1 was positively correlated with reliance on *f0* in L2 [β=5.46, p=.015, r^2=.24]. That is, the listeners who were more sensitive to the redundant cue in ambiguous VOT conditions in L1 were more sensitive to the same cue in L2 speech processing. In addition, the linear slope (β=5.46) greater than 1 indicates that sensitivity to *f0* was strengthened as the relative importance of the *f0* cue become greater. Similar to the VOT models, adding L2 proficiency improved the model fit (r^2=.43 in the complex model compared to r^2=.24 in simple model). Again, listeners with better L2 proficiency were more attentive to *f0* than listeners with less good L2 proficiency.

Figure 2: Scatter-plots of L2 coefficients as a function of L1 coefficients: VOT_{L1} (leftmost), $f0_{L1}$ (center), and $f0_{L1}$ in the ambiguous VOT condition (rightmost). Solid lines indicate that the regression model was significant (p< .05)



4. DISCUSSION

The current study explored individual differences in cue-weighting patterns for the stop voicing contrast in L1 English speakers and examined whether they predicted individual differences in L2 cue weighting patterns in Korean. The findings showed that there was a relationship between cue-weighting patterns in L1 and L2. Individuals who relied more on *f0* and

less on VOT in L1 showed a similar pattern in L2, observed in individuals' L2 speech perception, confirming the existence of the L1-to-L2 transfer effect at the individual listener level. Consistent with previous findings (e.g., [5]), L2 proficiency also turned out to be an important factor in explaining individual differences in cue weighting strategies in L2. As L2 proficiency in Korean increased, VOT became relatively less important and *f0* became relatively more important in differentiating the Korean phonation-type contrast relative to how listeners processed the voicing contrast in their native language.

The case of L1 English speakers learning Korean is a particularly interesting example of cue weighting because *f0* is a secondary cue to the voicing contrast in English and one of several primary cues to the voicing contrast in Korean. At least, in this specific situation of L2 perceptual learning, greater attention to the redundant acoustic cue in L1 resulted in more native-like cue weighting in L2. It will be of interest to examine individual differences in attention to other redundant cues in L1 and L2 perception, where the secondary cue is redundant in both languages (or neither). It is possible that language learners with better L2 perceptual cue weighting strategies attend more to redundant cues in L1 more generally, or it may be that this result holds only when the redundant cue in L1 is relatively more important in L2. More generally, this result provides additional evidence for the claim in [10,11] that individual differences in cue-weighting strategies are systematic.

6. ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012S1A5A8022655).

7. REFERENCES

[1] Abramson, A. S., and Lisker, L. 1985. Relative power of cues: F0 shift versus voice timing. In: V. Fromkin *(eds), Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Academic: New York, 25–33.

[2] Cebrián, J. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics.* 34, 371-387.

[3] Escudero, P., & Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition.* 26, 551-585.

[4] Escudero, P., Benders, T., and Lipski, S. 2009. Native, non-native and L2 perceptual cues weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*. 37, 452–466.

[5] Flege, J.E., O.-S. Bohn, & S. Jang. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics,* 25, 437-470.

[6] Francis, A. L., Kaganovich, N., and Driscoll-Huber, C. 2008. Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *J. Acoust. Soc. Am.* 124(2), 1234-1251.

[7] Gordon, P. C., Eberhardt, J. L., and Rueckl, J. G. 1993. Attentional modulation of the phonetic significance of acoustic cues. *Cogn. Psychol.* 25, 1-42.

[8] Haggard, M., Ambler, S., and Callow, M. 1970. Pitch as voicing cue. *J. Acoust. Soc. Am.* 47, 613–617.

[9] Holt, L. L., and Lotto, A. J. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *J. Acoust. Soc. Am.* 119(5), 3059–3071.

[10] Idemaru, K., Holt, L.L., and Seltman, H. 2012. Individual differences in cue weights are stable across time: the case of Japanese stops lengths. *J. Acoust. Soc. Am.* 132, 3950-3964.

[11] Kong, E., and Edward, J. 2011. Individual differences in speech perception: Evidence from visual analogue scaling and eye-tracking. *Proceedings of the 17th International Congress of Phonetic Sciences*, 1126-1129.

[12] Lisker, L., and Abramson, A. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Words.* 20, 384-442.

[13] Lisker, L. 1986. "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech.* 29, 3-11.

[14] Morrison, G. S., and Kondaurova, M.V. 2009. Analysis of categorical response data: Use logistic regression rather than endpoint-difference scores or discriminant analysis (L) *J. Acoust. Soc. Am.* 126(5), 2159-2162.

[15] Piske, T., Mackay, I.R.A, and Flege, J.E. 2001. Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics* 29, 191-215.

[16] Shultz, A. A., Francis, A. L., and Llanos, F. 2012. Differential cue weighting in perception and production of consonant voicing. *J. Acoust. Soc. Am.* 132, EL95– EL101.

[17] Silva, D.J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology.* 23, 287-308.

[18] Walley, A.C., and Carrell, T.D. 1983. Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *J. Acoust. Soc. Am.* 73, 1011-1022.

[19] Whalen, D.H., Abramson, A.S., Lisker, L. and Mody, M. 1993. F0 gives voicing information even with unambiguous voice onset times. *J. Acoust. Soc. Am.* 93, 2152-2159.