



Individual differences in categorical perception: L1 English learners' L2 perception of Korean stops*

Eun Jong Kong**

Department of English, Korea Aerospace University, Goyang, Korea

Abstract

This study investigated individual variability of L2 learners' categorical judgments of L2 stops by exploring English learners' perceptual processing of two acoustic cues (voice onset time [VOT] and f_0) and working memory capacity as sources of variation. As prior research has reported that English speakers' greater use of the redundant cue f_0 was responsible for gradient processing of native stops, we examined whether the same processing characteristics would be observed in L2 learners' perception of Korean stops ($/t/-/t^h/$). 22 English learners of L2 Korean with a range of L2 proficiency participated in a visual analogue scaling task and demonstrated variable manners of judging the L2 Korean stops: Some were more gradient than others in performing the task. Correlation analysis revealed that L2 learners' categorical responses were modestly related to individuals' utilizations of a primary cue for the stop contrast (VOT for L1 English stops and f_0 for L2 Korean stops), and were also related to better working memory capacity. Together, the current experimental evidence demonstrates adult L2 learners' top-down processing of stop consonants where linguistic and cognitive resources are devoted to a process of determining abstract phonemic identity.

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

1. Introduction

In processing auditory signals into phonetic categories, listeners are known to be sensitive to acoustic differences across categories but little (or not) so to differences within a category, even though the difference is physically the same along the acoustic dimension. This well-known speech processing mechanism, namely *categorical perception* (e.g., Liberman et al., 1957; Liberman et al., 1961; Pisoni & Tash, 1974), has emphasized the role of abstract phonetic labels in listeners' linguistic representation in warping relevant

acoustic space. While this characteristic of acoustic space distortion has been viewed by many researchers as listeners' insensitivity to within-category acoustic details masked by abstract phonological labels, other researchers also have demonstrated that fine-grained acoustic details persistently affect listeners' processing segments and words (Carney et al., 1977; McMurry et al., 2008; Pisoni & Tash, 1974; Schouten et al., 2003). That is, auditory stimuli with ambiguous or less typical acoustic properties delayed listeners' category decision or word recognition according to experimental studies measuring RTs or eye-movements during phonetic

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

** ekong@kau.ac.kr, Corresponding author

Received 29 October 2019; Revised 4 December 2019; Accepted 6 December 2019

© Copyright 2019 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

labelling tasks. Moreover, depending on the type of identification tasks, some listeners could be less categorical in assigning category labels in the task suggesting individual differences in phonetic encoding process (e.g., Massaro & Cohen, 1983; Schouten et al., 2003).

As one way to understand those processing delays and associated individual differences, several recent studies proposed that listeners' utilizations of acoustic cues would modulate perceptual manner of judging speech categories (Kapnoula et al., 2017, Kong & Edwards, 2016). The studies demonstrated that gradient (or less categorical) manner of perception was positively linked to listeners' greater sensitivities to redundant cues (e.g., *f0* in English /d-/t/ contrast) in the identification tasks. Also, listeners with gradient perception tended to have better working memory capacity than others. In this sense, gradient perception indicates listeners' ability of integrative phonetic encoding.

Inspired by the existing studies on gradient perception by English-speaking monolinguals, the current study examined the manner of category judgments by English-speaking adult learners of L2 Korean to investigate whether the gradient/categorical perception would also be modulated by learners' utilizations of multiple acoustic cues in L2 and individuals' general cognitive ability. This question is particularly interesting to be examined with the task of L1 English learners' perception of L2 Korean stops, since the roles of acoustic cues for laryngeal stop contrasts are defined differently between the languages. In English, voice onset time (VOT) 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* also differ systematically between voiceless and voiced stops (Abramson & Lisker, 1985; Francis et al., 2008; Gordon et al., 1993; Shultz et al., 2012). 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'/) (Kong & Lee, 2018; Kong et al., 2011; Lee & Jongman, 2011; Silva, 2006). Given existing evidence that L2 learners gradually learn to use cue weighting strategies of their second language (e.g., Escudero & Boersma, 2004; Escudero et al., 2009; Flege et al., 1997; Llanos et al., 2013), we wonder whether English-speaking L2 learners' individual variability of gradient perception, if any, might be associated with acoustic cues of VOT and *f0* in a different way from the pattern observed in monolingual English-speakers' perception of native stops.

The specific questions addressed in the present study were the following. One, would L2 learners show individual differences in their manner of judging L2 stop categories? If so, how would those individual differences in L2 perception be related to those in L1 perception? Two, would L2 learners' variability in gradient or categorical responses be explained by perceptual sensitivity to VOT and *f0*? How would re-weighted roles of *f0* and VOT for Korean stop contrasts affect learners' manner of judging the categories? The last question concerns an interaction between cognitive resources and L2 speech processing. Would L2 learners of better cognitive control differentiate auditory signals in a fine-grained scale (*gradient* perception) or map them into abstract labels (*categorical* perception)? We tested L2 learners' working memory capacity to explore this question.

In addressing the questions, we take L2 proficiency into consideration, given the well-known effect of L2 proficiency in categorical perception and relative cue-weighting strategies (Flege et al., 1997; Kong & Yoon, 2013; Piske et al., 2001). Experimental findings of the current study would broaden our understanding of how encoding processing of L2 categories interacts with language-specific uses of sub-phonemic acoustic details and with general cognitive ability.

2. Methods

2.1. Auditory Stimuli

The audio stimuli were adopted from Kong & Edwards (2016), two sets of synthesized CV syllables based on male speakers' natural productions. For the English stimuli, one token of /da/ and one of /ta/ (from *dot* and *tot*) productions by an adult male native English speaker were selected to process acoustic dimensions of VOT and *f0*. VOT values were manipulated by excising a portion of the burst release/aspiration from a /ta/ sample and pasting it before the voicing onset of the /da/ token: Six log-scale values from 9 ms to 59 ms of VOT were used (9 ms, 13 ms, 19 ms, 28 ms, 40 ms, and 59 ms). In 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 (98 Hz, 106 ms, 114 Hz, 122 Hz, and 130 Hz). This procedure yielded 30 different CV syllables (six VOT steps × five *f0* steps). The same procedure was implemented with a /t^ha/ and a /ta/ token (from '닷' [t^hat] and '타' [tat]) produced by a male Seoul Korean speaker. By pasting a portion of the burst release/aspiration from a /t^ha/ sample to a /ta/ token, we created another set of 30 different syllables (six VOT × five *f0* combinations) for the L2 Korean session of the speech perception task.

2.2. Tasks and Procedures

The participating learners were given two kinds of tasks: Speech perception tasks (visual analogue scaling (VAS) (Massaro & Cohen, 1983), and forced-choice identification (2/3AFC)) and executive function task (Digit N-Back) for working memory measure.

Carrying out the speech perception task of VAS, we exclusively employed task procedures of Kong & Edwards (2016). Listeners were presented a horizontal arrow with 'd' and 't' labels (in the English session; 'ㄷ' /t/ and 'ㅌ' /t^h/ in the Korean session) at two ends and were instructed to judge perceptual /d/- or /t/-likeness (in English; /t/- or /t^h/-likeness in L2 Korean) by clicking locations close or distant from the relevant labels. How precisely the learners tend to differentiate auditory tokens would indicate the individual listeners' manner of category judgement. The other speech perception task was an identification task where listeners choose either 'd' or 't' labels displayed on the monitor upon hearing auditory tokens in the L1 English session (2AFC). In the Korean session, listeners were to select one of the three consonant labels ('ㄷ' /t/, 'ㄷ' /t/ and 'ㅌ' /t^h/: 3AFC). The learners' categorical responses are used to estimate how listeners utilized acoustic information of VOT and *f0* in the stimulus tokens for phonetic labeling.

The VAS and AFC tasks were counter-balanced across the participants. In each language session of both tasks, a set of 30 different CV syllables was repeated three times in a random order, yielding 90 trials per listener. The speech perception tasks were administered in E-Prime2 so that responses were automatically recorded.

Besides the speech perception tasks, we administered Digit N-Back task where learners were asked to answer whether the digit in the current slide matches the digit in the one slide before (One-Back), one in the two slides before (Two-Back), and one in the three slides before (Three-Back). Performance of accuracy and speed of the N-Back task blocks is known to indicate one's working memory capacity, i.e., ability to hold and process information. Counts of accurate answers and response times were automatically obtained on E-Prime2 platform.

2.3. Participants

22 English native speakers (F:10, M:12 in their 20s-40s) participated in the experiment. All were living in Seoul or Gyeonggido at the time of recruitment and testing. Since the length of their residence in Korea widely varied from 18 months to 13 years, we evaluated their L2 proficiency by giving them a short test composed of question items from the TOPIK (*Test of Proficiency in Korean*, a test administered by the National Institute of International Education in Korea). The test was designed to assess listening comprehension, vocabulary comprehension, and reading comprehension. We used the composite scores from this test as a numeric index of L2 proficiency. Participants were monetarily compensated for their participation.

2.4. Statistical Analysis

Responses from the five tasks were processed to numerate individual-level perceptual (VAS, and 2/3AFC) or cognitive (WM) characteristics. In VAS, distributions of click locations along the arrow (i.e., distance from the two given labels) were quantified in terms of polarity of the distribution envelope (Kong & Edwards, 2016). Polynomial regression models were made for each individual listener, of which quadratic coefficients served to exhibit spread or steepness of the click increases toward the edges of arrow. In this assessment, smaller quadratic coefficients indicate more gradient judgment of the phonetic labelling.

To estimate listeners' sensitivities to the two acoustic cues (VOT and f_0) for identification (2/3AFC tasks), we built the mixed-effects logistic regression models separately for the L1 English and the L2 Korean tasks. In the models, the dependent variables were the binary responses of "da" and "ta" for the English session and the binary responses of "ㄷ" /ta/ and "ㄷ" /tʰa/ (lax-aspirated) excluding the responses of "ㅌ" /t/. The fixed effect variables were VOT and f_0 values of each stimuli as continuous variables, whose units were scaled to make beta coefficients comparable between two acoustic dimensions. The model included random intercepts and slopes of VOT and f_0 varying at the listener level. We used individual random slope coefficients as numeric index of individual listeners' sensitivity to VOT and f_0 .

Individual learners' performance in Digit N-Back task was evaluated by accuracy counts and speed of those correct responses. Accuracy and response times (RTs) were calculated in each

sub-block of the tasks: i.e., One-/Two-/Three-Back blocks – we later discarded values from One-Back due to its ceiling effect. Trials whose response time ranges out of two standard deviations of each participant's performance were excluded from the calculation. Averaged RTs of correct responses per listener in each block were used as WM scores for relevant analyses below.

Finally, all the numeric indices of perceptual and cognitive characteristics entered (partial-) correlation tests to figure out its relationship with perceptual manner of L2 category decision. L2 proficiency (i.e., test scores of Korean language) was considered as a control variable to partial out its potential effect on L2 sound category processing. We applied a relaxed criterion in interpreting the strength of correlation between variables to accommodate the relatively small sample size, treating correlation coefficients greater than .3 as a meaningful tendency to be noted.

3. Results

3.1. Individual Differences in Gradient Perception

Figure 1 presents click distributions of the English and Korean VAS tasks, which were collapsed across all listeners. When visually inspected, click distributions of the L2 Korean task were overall flatter than those of the L1 English task, suggesting more gradient responses of category judgement in L2 perception. It is noted that the clicks in the L2 Korean task form a peak at the center of the distribution, which might reflect learners' uncertainty in labelling L2 stop sounds. Regression curves overlaid on top of histograms confirm the observation by displaying a curve with steeper rises (i.e., categorical judgement) towards both ends of the distribution in the English task than in the Korean one.

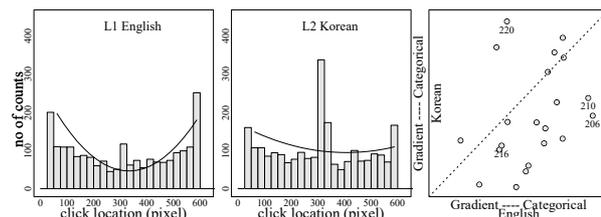


Figure 1. [Two left panels] Distributions of click locations in English and Korean VAS tasks with polynomial regression curves overlaid. [Rightmost] Individual listeners' quadratic coefficients in the English VAS as a function of those in the Korean VAS.

The scatter-plot at the rightmost panel of Figure 1 shows individual listeners' quadratic coefficients of English VAS responses as a function of those of L2 Korean VAS responses. It is observed that most data-points were clustered below a diagonal line, indicating many individual English listeners were more gradient in perceiving L2 Korean stops than native English stops. A correlation test between perceptual gradiency of English and Korean (quadratic coefficients of click histograms from VAS) returned a modestly meaningful coefficient with a positive sign ($r=.32$, $p=.16$), suggesting that perceptual gradiency in L1 is somewhat (but not strongly) suggestive of L2 gradiency.

Confirming this (consistent but) loose correlation between L1 and L2 perception in terms of gradient judgments, individual listeners' histograms in Figure 2 present examples of individual learners

whose gradiency in L1 English perception is similar to or different from degrees of gradiency in L2 Korean perception. For one, a listener ID220 was categorical and yet a listener ID216 was gradient in perceiving Korean stop although both were similarly gradient in the English stop perception.

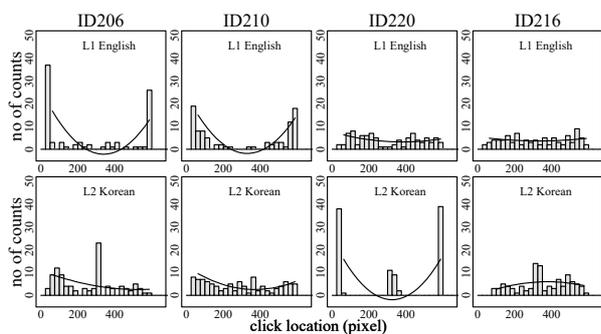


Figure 2. Four example individuals' click distributions in L1 English (top) and L2 Korean (bottom) VAS tasks with polynomial regression curves overlaid.

3.2. Acoustic Cue Weighting in L1 and L2 Stop Perception

Figure 3 [top panels] presents logistic curves of VOT and f_0 variables estimated from a logistic mixed-effects regression model of the English task (see parameter estimations summarized in Table 1). In the English session, the group-averaged curves for VOT and f_0 (black lines) rise as VOT and f_0 values increase, indicating that perception of the voiceless, and phonetically aspirated stop category is associated with greater VOT and higher f_0 . Relatively speaking, the group-averaged VOT curve switches from /d/ to /t/ more abruptly than the f_0 curve: $\beta_{\text{VOT}}=5.91$, $\beta_{f_0}=1.30$, which is also confirmed in the stimulus matrix (center panel), where darker cells (/t/ response) were distributed at the right side (longer VOTs). This pattern is in line with previous findings that VOT is the primary perceptual cue for monolingual English listeners (Abramson & Lisker, 1985; Francis et al., 2008; Gordon et al., 1993; Shultz et al., 2012). Looking into individual-level variations of VOT and f_0 sensitivities (rightmost panel), a correlation test between by-listener coefficients of VOT and f_0 slopes yielded a pattern that individual listeners' VOT uses were inversely correlated with those of f_0 although the magnitude of effect was modest ($r=-.33$, $p=.12$). English-speaking listeners who depended on VOT more than others tended to rely on f_0 less in perceiving L1 stops.

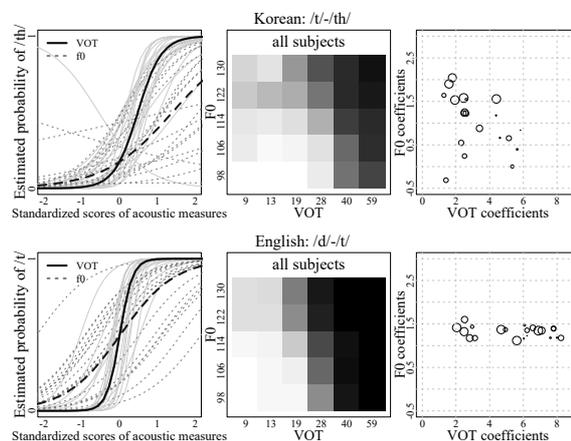


Figure 3. [leftmost] Estimated probability of /t/ (L1 English) from the mixed-effects logistic models. Thick lines indicate group-averaged coefficients of VOT and f_0 , and thin lines the individual learners' coefficients. Center panels present stimulus matrix where darker cells mean more /t/ responses. [rightmost] Individual learners' f_0 coefficients were plotted against VOT coefficients with L2 Korean proficiency indicated by the character size.

Figure 3 [bottom panels] presents regression curves of VOT and f_0 based on model outputs of /t-/t^h/ responses. Parameter estimations are summarized at Table 1. The curves based on a regression pattern of Korean /t-/t^h/ responses were similar to those of L1 English /d-/t/ responses, in that group-averaged curves rise as VOT or f_0 values increase, and a switch from /t/ to /t^h/ is more abrupt along a VOT dimension than along a f_0 dimension ($\beta_{\text{VOT}}=3.18$, $\beta_{f_0}=1.06$). One noticeably different pattern between the two models is that coefficient differences between VOT and f_0 were smaller in L2 Korean than in L1 English, suggesting that listeners' reliance on VOT was less dominant in L2 perception and their dependency on f_0 was relatively greater for a /t-/t^h/ contrast. This slightly adjusted cue-weighting pattern in L2 stop perception is visualized in the stimulus matrix (middle panel) where darker cells are located at the top-right corner (longer VOT and higher f_0) and lighter cells at the right bottom corner (shorter VOT and lower f_0). Although English-speaking L2 learners' reliance on f_0 was not as enhanced as Korean native speakers' primary use of f_0 , this re-weighting pattern suggests that f_0 is an important acoustic dimension (a targeted primary cue) that English-speaking L2 learners deliberately attend to enhance its use.

3.3. L2 Proficiency and Gradient Perception of L1 and L2 Stops

We examined a relationship between L2 test scores and manner of perceiving L1 and L2 stops. Considering a modest amount of correlation between L1 and L2 perceptual gradiency, we employed partial correlation tests to exclude effect of L1 perception in estimating the relation of L2 perception with L2 proficiency and vice versa. (*ppcor* package in R, Kim, 2012). Results revealed no significant relationship between L2 test scores and perceptual gradiency in L1 or L2 (with L1: $r=-.18$, $p=.43$; with L2: $r=-.20$, $p=.39$).

We conducted further correlation tests to examine whether L2 proficiency is related to learners' utilizations of VOT and f_0 in L1

English and L2 Korean stop perception. Table 2 summarizes test outputs. The relationship of L2 proficiency with L2 acoustic sensitivity was consistent in that higher L2 test scores were negatively correlated with VOT coefficients ($r=-.58, p<.01$) and positively with $f0$ coefficients ($r=.44, p<.05$). These results suggest that, as learners become proficient in L2, they become better able to attend to a primary information ($f0$) for a stop contrast.

Interestingly, L2 proficiency was also negatively correlated with L1 VOT use, although a strength of correlation was relatively weak. This finding that proficient L2 learners tended to utilize VOT (a primary cue in English) less ($r=-.37$) may suggest an effect of L2 learning on the perception of native stops (e.g., Lev-Ari & Peperkamp, 2013).

Table 1. The output of the mixed effects models presented in Figure 3

	Parameters	Estimate	Std.err	p-value
English	VOT	5.86	0.70	<.001
(/d/-t/)	$f0$	1.38	0.14	<.001
Korean	VOT	3.21	0.52	<.001
(/t/-t ^h /)	$f0$	1.11	0.21	<.001

Table 2. The summary of correlation tests between Variable 1 (individuals' acoustic coefficients) and Variable 2 (L2 proficiency)

Variable 1		Variable 2	Estimate	p-value
English	VOT coef.	L2 test scores	-0.37	.094
(/d/-t/)	$f0$ coef.	L2 test scores	0.25	.27
Korean	VOT coef.	L2 test scores	-0.58	<.01
(/t/-t ^h /)	$f0$ coef.	L2 test scores	0.44	<.05

3.4. Relationship between Gradient Perception and Acoustic Cue Utilizations

With L2 proficiency considered as a control variable, partial correlation tests showed no meaningful correlation coefficient between gradient responses in L1 task and L1 $f0$ utilizations ($r=-.12$). This is not congruent with observations in earlier studies that $f0$ utilizations was responsible for English monolinguals' gradient responses in the native stop perception (Kapnoula et al., 2017; Kong & Edwards, 2016). Instead, it was L1 VOT coefficients that were meaningfully correlated with L1 gradiency scores, suggesting that L2 learners who were sensitive to VOT, a primary cue in English /d/-t/ contrast, tended to be *categorical* in labelling the sounds.

Table 3. The summary of partial correlation tests between Variable 1 (L1 or L2 Gradiency) and Variable 2 (acoustic coefficients). L2 proficiency was a control variable

Var 2	L1 Gradiency (Var 1)	
	Estimate	p-value
L1 VOT coef. (/d/-t/)	.37	.111
L1 $f0$ coef. (/d/-t/)	-.12	.59
L2 Gradiency (Var 1)		
L2 VOT coef. (/t/-t ^h /)	-.22	.33
L2 $f0$ coef. (/t/-t ^h /)	.29	.21

Along the same lines, L2 gradiency scores were correlated with L2 $f0$ coefficients despite a weak strength of correlation ($r=.29$). That is, greater sensitivity to L2 $f0$ (a primary cue to a Korean /t/-t^h/ pair) was associated with *categorical* manner of judgment. We consider this direction of relationship in L2 analogous to the

relationship between categorical responses and VOT utilizations in L1, in that greater uses of a primary cue were responsible for categorical judgements of speech signals. In a general sense, the current finding conforms to existing evidence that listeners' utilizations of acoustic information modulate encoding process of speech categories accounting for individual variability in categorical perception (Kapnoula et al., 2017; Kong & Edwards, 2016).

3.5. Relationship between Perceptual Gradiency and EF Scores

Finally, partial-correlation tests were performed to examine the relationship between working memory capacity (WM: RTs and accuracy from N-Back tasks) and (1) perceptual gradiency scores (from VAS tasks) and (2) acoustic sensitivities to VOT and $f0$ in L1 and L2 stop identification tasks. L2 proficiency was also considered as a control variable in the tests to assess the correlation between cognitive control and speech processing independently of one's control of L2 language.

As summarized in Table 4 [top panels], faster RTs of N-Back task were significantly correlated with categorical judgements of the stops both in L1 and L2 perception, indicating an association of better WM with *categorical* perception. This direction of relationship is contrastive with the association of better WM with gradient responses observed in English-speaking monolinguals' native stop perception (Kapnoula et al., 2017). This discrepancy might reflect different nature of speech processing mechanism between L1 and L2 perception, as L2 learners' phonetic encoding would be more oriented to abstract level of speech (i.e., phonetic category decisions) due to established labels in L1 (e.g., Best & Tyler, 2007). Accordingly, in L2 speech perception, cognitive resources might operate in a different way from L1 perception by being linked to higher level of speech processing.

WM was also correlated with L1 and L2 acoustic coefficients. With L1 coefficients, higher accuracy of WM was associated with smaller L1 $f0$ coefficients and greater L1 VOT coefficients, indicating that listeners with better WM attended to VOT, a primary cue in L1, more than others and utilized $f0$, a redundant cue, less than others in identifying the English stops /d/-t/. The pattern suggests that cognitive resources are devoted to perceptual attention to a primacy cue.

Mirroring the pattern related to L1 acoustic coefficients, higher accuracy and faster RTs in the WM task were modestly correlated with greater use of $f0$ in L2 Korean perception. This echoes a tendency that better WM was associated with greater reliance on a primary cue, which is $f0$ in Korean /t/-t^h/ contrast. Again, this seems to support that bilinguals' or L2 learners' cognitive resources operated to maximize utilizations of primary acoustic information that is defined for a specific language and phonological contrast.

Table 4. The output of partial-correlation tests between WM scores and perceptual gradiency (top), VOT coefficients (middle) and $f0$ coefficients (bottom) measured in L1 and L2 stop perception tasks.

		L1 Gradiency		L2 Gradiency	
RT	2-Back	-0.037	0.878	-0.351	0.129
	3-Back	-0.328	0.158	-0.327	0.16
ACC	2-Back	0.091	0.702	-0.079	0.739
	3-Back	0.125	0.599	-0.208	0.379
		L1 VOT coefficients		L2 VOT coefficients	
RT	2-Back	0.178	0.452	0.126	0.595
	3-Back	0.119	0.618	-0.107	0.654
ACC	2-Back	0.306	0.189	0.061	0.798
	3-Back	0.378	0.1	-0.105	0.659
		L1 $f0$ coefficients		L2 $f0$ coefficients	
RT	2-Back	-0.16	0.499	-0.167	0.481
	3-Back	0.006	0.979	-0.356	0.124
ACC	2-Back	0.021	0.931	0.331	0.154
	3-Back	-0.381	0.097	0.419	0.066

4. Discussion

The current study explored English speaking L2 learners' individual differences in their manner of judging stop laryngeal contrasts in L1 Korean in order to investigate the roles of acoustic cue uses and cognitive controls as sources of individual variability in speech information processing. Experimental evidence showed that some individual learners of L2 Korean were gradient in judging the stop categories /t/ and /t^h/, while others were strictly categorical between the two consonant labels in category judgments. This pattern of individual variabilities in L2 perception is consistent with existing findings in English monolinguals' native perception (e.g., Kapnoula et al., 2017; Kong & Edwards, 2016, from which we employed methodological framework in a large scale). Importantly, those individual differences in gradient or categorical judgments of L2 stops were accounted for by learners' perceptual sensitivity to an acoustic cue primarily important for the stop contrasts, and they interacted with listeners' working memory (WM) control, whose relationship was assessed independently of learners' L2 proficiency.

4.1. Gradient Perception and Acoustic Cue Utilization

In L1 stop perception, the relationship of gradient/categorical responses with a primary cue to the /d-/t/ contrast (i.e., VOT) was meaningful, although the relation with a redundant cue ($f0$) was not significant. In a broad sense, this finding conforms to the proposal based on monolingual English speakers' L1 stop perception that acoustic sensitivity is responsible for listeners' manner of perception (Kapnoula et al., 2017; Kong & Edwards, 2016). In its details, however, the current finding slightly differs from evidence from monolingual listeners in that it was learners' acoustic sensitivity to a primary cue but not to a secondary cue that was modulated the manner of judging stop categories: Listeners with a greater sensitivity to a primary cue tended to be *categorical* in labelling the stimuli. We may argue for different nature of modulating acoustic cues for category judgments between monolinguals and bilinguals (adult L2 learners), where L2 learners modulate utilizations of primary acoustic information for the sake of perceptual differentiations of between-category identities.

The same account might be applied to explain the meaningful relationship between $f0$ utilizations and judgment manner of L2 Korean stops: Individual listeners' *categorical* responses were associated with a greater sensitivity to $f0$, a primary cue in perceiving Korean stops /t-/t^h/, analogous to how VOT modulated the listeners' categorical perception in English stop perception, $f0$ must have affected the L2 learners' categorical manner of labelling L2 Korean stops playing as a primary cue to the lax-aspirated stop contrast in Korean (e.g., Lee et al., 2013; Schertz et al., 2015). Although the current participants of English-speaking L2 learners poorly realized this relative cue-weighting between $f0$ and VOT in Korean stop perception, the increased role of $f0$ as their target primary cue functioned to affect the manner of categorical perception.

While a follow-up study of a larger sample size is called for to draw a robust conclusion in terms of categorical/gradient responses in L2 perception, it is still an important finding that L2 learners' manner of phonetic encoding is modulated by a primary cue to yield categorical responses of the stops. Existing models of L2 speech perception (e.g., PAM-L2, Best & Tylor, 2007) appears closely relevant to an understanding of the trend that adult L2 learners' signal processing had to do with a primary cue, but less so with a secondary cue. That is, L2 learning adults process speech inputs to map onto existing *categories*, difficulties of which depend on phonemic status and acoustic similarity of target phonetic categories in question between L1 and L2 languages. Weighing between within- and between-category processing, adult L2 learners might be oriented more toward differentiations of category identities, emphasizing a role of primary cue, than toward enrichment of category representation with within-category acoustic details. The current findings show that bilinguals (including adult L2 learners) utilize linguistic resources to operate higher level information relatively more than lower level information, which is probably different from how monolinguals utilized redundant cues to yield gradient responses.

4.2. L2 Proficiency and Acoustic Cue-Weighting

We could not find strong evidence that L2 proficiency modulated L2 learners' perceptual manner of judging stop categories. L2 proficiency, however, turned out to be an important factor in explaining individual differences in cue weighting strategies in L2, which is consistent with existing findings (e.g., Flege et al., 1997; Piske et al., 2001). Consistent trends were that, as the listeners with higher proficiency were relatively more sensitive to $f0$ (a primary information in Korean /t-/t^h/ contrast) and less so to VOT in differentiating the Korean stops. This direction of relationship suggests that more proficient learners are better able to adjust to the language-specific acoustic characteristics of the L2 phonetic categories in terms of relative cue-weighting. Kong & Yoon (2013) reported a similar finding that L1 Korean learners of L2 English showed differential degrees of dependency on VOT (a primary cue in English) between high-proficient and low-proficient learners in the L2 English stops identification. More generally, current results support that learning L2 sounds concerns fine-grained tuning of acoustic cue-weightings as examined in other languages than English and Korean (e.g., Escudero & Boersma, 2004; Escudero et al., 2009; Holt & Lotto, 2006).

4.3. Speech Perception and Executive Function Capacities

Relationship of gradient/categorical responses with EF control was meaningful both in L1 and L2 speech perception. Listeners with greater WM tended to judge the stop sounds in a *categorical* manner than others. Differently from observations in English-speaking monolingual listeners' native perception (Kapnola et al., 2017), L2 learners' WM functioned to facilitate *categorical* responses instead of *gradient* responses. As non-native listeners utilized cognitive resources less on processing low-level acoustic information (Mattys et al., 2010), current evidence also shows cognitive emphasis on a phonetic encoding of auditory signal. Together with L2 adult learners' manner of operating linguistic resources at an abstract level of category encoding, present findings related to cognitive resources serve to generalize that bilingual listeners' encoding process of speech input is different from that of monolinguals as a role of cognitive efforts is defined differently.

5. Conclusion

To conclude, the present study showed that individual differences in categorical/gradient responses existed in adult L2 learners' perception of stop consonants. Not only that, those variations were also systematic in terms of learners' perceptual utilizations of acoustic information and their general cognitive ability. It was characteristic of adult L2 learners' stop perception that greater utilizations of primary cue and better cognitive function facilitated *categorical* perception, supporting that *effective* L2 learners process speech in a top-down manner.

Acknowledgements

A subset of data in this work (VAS responses from 20 participants) has been previously presented at 18th International Congress of Phonetic Science (Kong & Edwards, 2015). The current version of manuscript is extensively updated from the earlier presentation by applying different analysis methods to a full set of response data.

References

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In: V. Fromkin (ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25-33). New York, NY: Academic Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13-34). Amsterdam, Netherlands: John Benjamins Publishing Company.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *The Journal of the Acoustical Society of America*, 62(4), 961-970.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26(4), 551-585.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, 37(4), 452-465.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437-470.

Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America*, 124(2), 1234-1251.

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, 25(1), 1-42.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, 119(5), 3059-3071.

Kapnola, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradience in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, 43(9), 1594-1611.

Kim, D., Clayards, M., & Goad, H. (2018). A longitudinal study of individual differences in the acquisition of new vowel contrasts. *Journal of Phonetics*, 67, 1-20.

Kim, S. (2012). ppcor: Partial and semi-partial (Part) correlation [Computer program]. <http://cran.r-project.org/package=ppcor/>

Kong, E. J., & Edwards, J. (2015). Individual differences in L2 learners' perceptual cue weighting patterns. *Proceedings of 18th International Congress of Phonetic Sciences (ICPhS 2015)*. Glasgow, UK.

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, 59, 40-57.

Kong, E. J., & Lee, H. (2018). Attentional modulation and individual differences in explaining the changing role of fundamental frequency in Korean laryngeal stop perception. *Language and Speech*, 61(3), 384-408.

Kong, E. J., & Yoon, I. H. (2013). L2 proficiency effect on the acoustic cue-weighting pattern by Korean L2 learners of English: Production and perception of English stops. *Journal of the Korean Society of Speech Sciences*, 5(4), 81-90.

Kong, E. J., Beckman, M., & Edwards, J. (2011). Why are Korean tense stops acquired so early: The role of acoustic properties. *Journal of Phonetics*, 39, 196-211.

Lee, H., & Jongman, A. (2011). Perception of initial stops in tonal and non-tonal Korean. *The Journal of the Acoustical Society of America*, 130(4), 2572.

Lee, H., Politzer-Ahles, S., & Jongman, A. (2013). Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of phonetics*, 41(2), 117-132.

Lev-Ari, S., & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics*, 41(5), 320-331.

Lieberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5),

358-368.

- Lieberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, 61(5), 379-388.
- Llanos, F., Dmitrieva, O., Shultz, A., & Francis, A. L. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *The Journal of the Acoustical Society of America*, 134(3), 2213-2224.
- Massaro, D. W., & Cohen, M. M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication*, 2(1), 15-35.
- Mattys, S. L., Carroll, L. M., Li, C. K. W., & Chan, S. L. Y. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication*, 52(11-12), 887-899.
- McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6), 1609-1631.
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29(2), 191-215.
- Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15(2), 285-290.
- Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, 52, 183-204.
- Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, 41(1), 71-80.
- Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, 132(2), EL95-EL101.
- Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, 23(2), 287-308.

- **Eun Jong Kong**, Corresponding author
Professor, Dept. of English, Korea Aerospace University
76 Hanggongdaehang-ro, Deogyang-gu, Goyang-City,
Gyeonggi-do, 10540, Korea
Tel: +82-2-330-0338
Email: ekong@kau.ac.kr
Areas of interest: Acoustic phonetics, L1 & L2 acquisition,
Speech perception