



Relationship between executive function and cue weighting in Korean stop perception across different dialects and ages*

Eun Jong Kong¹ · Hyunjung Lee^{2,**}

¹Department of English, Korea Aerospace University, Goyang, Korea

²Department of English Education, Incheon National University, Incheon, Korea

Abstract

The present study investigated how one's cognitive resources are related to speech perception by examining Korean speakers' executive function (EF) capacity and its association with voice onset time (VOT) and *f0* sensitivity in identifying Korean stop laryngeal categories (/t'/ vs. /t/ vs. /t^h/). Previously, Kong et al. (under revision) reported that Korean listeners (N = 154) in Seoul and Changwon (Gyeongsang) showed differential group patterns in dialect-specific cue weightings across educational institutions (college, high school, and elementary school). We follow up this study by further relating their EF control (working memory, mental flexibility, and inhibition) to their speech perception patterns to examine whether better cognitive ability would control attention to multiple acoustic dimensions. Partial correlation analyses revealed that better EFs in Korean listeners were associated with *greater* sensitivity to available acoustic details and with greater suppression of irrelevant acoustic information across subgroups, although only a small set of EF components turned out to be relevant. Unlike Seoul participants, Gyeongsang listeners' *f0* use was not correlated with any EF task scores, reflecting dialect-specific cue primacy using *f0* as a secondary cue. The findings confirm the link between speech perception and general cognitive ability, providing experimental evidence from Korean listeners.

Keywords: Korean stop perception, acoustic cue weighting, executive function capacity, voice onset time (VOT), fundamental frequency

1. Introduction

When listeners are faced with multiple dimensions of acoustic information in processing speech, some acoustic dimensions have greater perceptual impact than others in indicating a phonological

contrast (e.g., Lisker, 1986; Repp, 1983). To take an example of voice onset time (VOT) and *f0* cues for English stops (e.g., /d/ vs. /t/), VOT is perceptually dominant in distinguishing a voiced-voiceless stops, thus referred to as a primary cue (Abramson & Lisker, 1985; Francis et al., 2008; Gordon et al., 1993; Holt & Lotto,

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (No. NRF-2017 S1A5A2A03068448).

** hjlee123@inu.ac.kr, Corresponding author

Received 27 July 2021; Revised 25 August 2021; Accepted 26 August 2021

© Copyright 2021 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.

2006). For the same contrast, other cues such as *f0* after release are relevant but less informative than VOT, which makes *f0* as secondary cues.

Given the differential impact of multiple cues, how would listeners efficiently cope with redundant information in processing speech? Regarding this question, previous experimental studies have shown that even though a primary cue is most informative, listeners *do* accommodate secondary cues in perceiving phonetic categories in whichever listening conditions are present. Whether or not a primary cue was neutralized in a stimulus signal, listeners attended to secondary cues in perception, which could affect listeners' decisions of phonetic categories one way or the other (e.g., Abramson & Lisker, 1985; Idemaru & Holt, 2011; Whalen et al., 1990, 1993). When VOT was ambiguous between two possible stop categories (i.e., /d/ vs. /t/ in English), *f0* was used as a determining cue (Abramson & Lisker, 1985; Whalen et al., 1990). More interestingly, even with unambiguous primary information, auditory stimuli of conflicting *f0* values affected listeners by slowing down their responses and slightly moving category boundaries along the VOT dimension (Whalen et al., 1993). Listeners' accommodation of secondary cues seems to support that speech perception is a highly flexible and resilient process, where listeners should continuously monitor, select and update helpful acoustic cues among available ones.

Given the nature of listeners' flexible processing of multiple cues, we are interested in how one's cognitive resource would work to efficiently handle multiple acoustic cues in perception. Would cognitively adept listeners be better at holding or shifting between multiple cues in perception? Would those adept listeners inhibit less relevant cues? Targeting a specific area of cognitive capacity, several recent studies investigated the link between executive function (EF) control and acoustic cue weighting. EFs refer to a set of top-down cognitive processes that regulate one's thoughts and actions by resisting, selecting and monitoring behavior to attain goals in everyday life (Diamond, 2013; Miyake et al., 2000). Miyake et al. (2000) proposed that EFs consist of three somewhat correlated yet separable basic components including updating of working memory (WM), inhibitions and shifting attention. In English-speaking adults' identification of native stops, Kapnoula et al. (2017) found that listeners with better WM showed greater perceptual sensitivity to secondary cues (*f0* for the stop voicing contrast). This might indicate that cognitive resources in English speakers are allocated to efficient management of redundant information in processing their native language.

Also, in non-native perception where listeners were to re-weight acoustic cues between L1 and L2, learners' better inhibition control was meaningfully correlated with language-specific acoustic-phonetic encoding. In Lev-Ari & Peperkamp (2013), adult English-French speakers with lower inhibition controls were hindered by an L2 cue in processing L1 stops. Likewise, Darcy et al. (2016) also showed that adult Spanish-English bilinguals' better attention and inhibition ability were associated with accurate identifications of L2 vowels and consonants. Together, the findings suggest that better EFs may benefit listeners in terms of flexible acoustic-phonetic encoding in L1 and L2 speech perception by bringing relevant cues other than a single dominant cue under listeners' cognitive control.

As the research on listeners' EF capacity has not been widely conducted in relation to phonetic details in perception, we aim to contribute to the field by providing experimental evidence from

Korean listeners. Specifically, we examine the relationship between Korean listeners' EF scores and use of multiple acoustic cues in stop laryngeal category identification. Kong & Yoo (2017) conducted a study with Korean speaking elementary school children (N=15, aged between 7 and 9), finding that children's suppression of an irrelevant acoustic cue was associated with better cognitive ability. This study continues to investigate this by extending the age range to the older listeners (adults and adolescents as well as children) who are living in two different dialect regions of Korea.

To investigate acoustic cue weighting patterns in Korean stop perception across various social settings, authors' own recent study (Kong et al., under revision) examined stop perception by elementary, high school and college students (N=164) in two different dialect regions in Korea, namely, Changwon (speaking South Gyeongsang dialect of Korean) and Seoul (speaking standard Korean). Results were that the stop laryngeal categories (/t'/-/t/-/t^h/) in the standard Seoul Korean were differentiated both by VOT and *f0* (reflecting an enhanced *f0* role for a /t/-/t^h/ pair), while the stops were dominantly distinguished along the VOT dimension in Gyeongsang dialect (Kang, 2014; Kong & Lee, 2018; Lee et al., 2013; Lee et al., 2020; Silva, 2006, See Appendix 1 for the group coefficients). When their cue weighting patterns were further compared across student groups within each dialect, college students (age: mean=23.3, SD=2.6) differed from high school students (age: mean=16, SD=0) in that their local characteristic (i.e., using VOT as a primary cue) was weaker in perception. These findings suggest that college students were more willing to accommodate non-local variants when they are in a broader social spectrum after entering a college. Different from older listeners, child participants in both regions (age: mean=9, SD=0.6) did not show dialect-specific cue weighting patterns, indicating that they have not yet acquired social-indexical values associated with regional dialects. Children in both regions were also different from the older groups in that children used the two acoustic cues (VOT, and *f0*) in a positively correlated manner. That is, children sensitive to VOT were also sensitive to *f0*, while adults who used one cue more than others used the other cue less (Appendix 2 illustrates these individual pattern differences, plotting individuals' VOT coefficients against *f0* coefficients separated by dialect regions and institutions.) This confirms that elementary school students have not mastered adult-like sub-phonemic cue associations yet (e.g., Hazan & Barret, 2000). With the results summarized across participant sub-groups, cue weighting patterns of this study can be a useful dataset to test their relationship with listeners' cognitive ability.

Following upon Kong et al.' (under revision) study with Korean listeners, we will further explore the relationship between (within-group) individuals' use of sub-phonemic acoustic cues and their EF capacity to better understand how cognitive resources help utilize multiple acoustic information. Specific research questions are as follows. First, between dialect groups, we would like to examine how individuals' EFs are correlated with the dominant acoustic dimension in each dialect: Would Gyeongsang Korean listeners, who attend to *f0*, less dominant cue than VOT, have better EF ability? Second, for the child groups, who did not show dialect-specific weights between VOT and *f0*, we would like to examine how children's EFs are related to VOT and *f0*. Although we expect that better EF performance would be associated with an efficient use of multiple cues, it remains to be seen specifically how cognitive advantages would be realized in Korean stop perception. For

example, they may use a group-specific primary cue more than others. At the same time, they may also be better at holding redundant cues more than others, as in English stop perception (Kapnoula et al., 2017). In contrast, they may use a group-specific secondary cue less than others so that cognitive attention is devoted to a primary cue. We will explore the relationship by conducting correlation analyses regarding individuals' EF performance and stop perception data collected in Kong et al. (under revision). The findings of this study will broaden our understanding of cognitive ability in modulating acoustic-phonetic encoding in speech perception.

2. Methods

2.1. Speech perception tasks

The speech perception data comes from an existing study conducted by the authors (Kong et al., under revision), in which 75 Gyeongsang speakers (32 college, 23 high school and 21 elementary school students) and 79 Seoul speakers (39 college, 19 high school and 21 elementary school students) listened to /C/+a/ syllables to identify Korean stop categories: /t'/ vs. /t/ vs. /t^h/ (See Table 1). Gyeongsang and Seoul listeners were recruited in education institutions located in Changwon (South Gyeongsang), and Seoul, respectively.

To briefly review the experiment preparation, the consonant portion of the stimulus syllable was synthesized by combining 6 step-VOTs and 5 step-*f*0s. The vowel part was taken from a male speaker's production of /ta/ (‘ㄷㅏ’). In the identification task, there were 90 trials (30 stimulus syllables×3 repetitions) presented in a randomized order. For the elementary school listeners, the task was modified to be shorter. Stimulus syllables were combinations of 4 step-VOTs and 3 step-*f*0s, and there were 48 trials (12 stimulus syllables×4 repetitions) in total.

To analyze identification patterns of the Korean stops, mixed-effects regression models (fixed effect variables: VOT, *f*0, Dialect (Seoul vs. Gyeongsang), and School (High school vs. College)) were made where VOT and *f*0 were allowed to vary at the listener level (random effect variable). Note that this School variable is meant to represent the participants' age range but not their educational background or academic capacity. With the same variable settings, there were three regression models to explain identifications of /t'-/t^h/ (lax-aspirated), /t'-/t'/ (lax-tense) and /t'-/t^h/ (tense-aspirated), respectively. The regression models for elementary school listeners were constructed separately due to the difference in experimental details such as trial numbers and stimulus acoustics. The children's model had the same fixed/random variable specifications except that it did not have the School variable. For the goal of this study, individual listeners' sensitivity to VOT and *f*0 was calculated by adding by-listener random effect coefficients (i.e., individuals' coefficient differences from the group average) to fixed effect coefficients (i.e., group means) (Appendix 2 presents twelve reproduced panels of individual listeners' *f*0 coefficients as a function of VOT coefficients separated by Dialect and School factors.)

Table 1. Participant demographic information

Listener	Group	n	Mean age (SD) in years
Seoul	Elementary school	21	9.0 (0.6)
	High school	19	16.0 (0.0)
	University	39	23.3 (2.5)
Gyeongsang	Elementary school	21	9.0 (0.0)
	High school	23	16.0 (0.2)
	University	32	22.6 (2.6)

2.2. Executive function tasks

Three different tasks were administered to measure the sub-components of participants' EF capacities (Miyake et al., 2000): (1) the digit N-Back task for WM capacity, (2) the Dimensional Change Card Sort (DCSS) task for mental flexibility, and (3) the Stroop for inhibitory control. Figure 1 presents slide examples of each task.

In the digit N-Back task, the participants were asked to answer whether the number on the current slide was the same ('S') or different ('D') from the number on the previous slides in three different blocks (e.g., Owen et al., 2005). In the digit 1-Back block, the reference digit was the one shown on the immediately preceding slide. In the digit 2-Back and 3-Back blocks, the reference digit was the number in the two- and three-slides back from the current slide, respectively. There were 40 test trials in each block after six practice items with feedback. Elementary students were given the digit 1-Back block only due to a difficulty of the other two blocks. By design, 40 test trials consisted of 10 trials whose answers were 'the same' (*Target*), and 30 trials whose answers were 'different' (*Non-Target*). The EF tasks were carried out in the E-Prime 3 software for automatic recording of accuracy scores (ACC) and response time (RT). Higher accuracy scores and shorter mean RTs of the correct responses indicate better working-memory capacity.

In the dimensional change card sorting (DCCS), the participants selected one of the two picture cards displayed on the monitor, which differed in color or shape information (Zelazo et al., 2003). In the two initial blocks of DCCS (*color* and *shape* blocks), the participants chose a picture card that matched a single information dimension, e.g., color in a *color* block, and shape in a *shape* block. Then, in the test block, the target information dimension to match could switch or stay from one trial to next one (e.g., color - shape [*Switch*] - shape [*Stay*] - color [*Switch*]). There were 13 *Switch* trials and 17 *Stay* trials. The mean RTs of the correct responses in *Switch* and *Stay* trials of the test block were calculated in each individual participant and the mean RT differences between *Switch* and *Stay* condition were used to represent individuals' cognitive cost of mental flexibility.

Finally, the Stroop task was a color-word matching task where the participants pressed one of the color buttons (red, green, blue, and yellow) matching the font color of the words displayed on the monitor (e.g., van Maanen et al., 2009). There were 80 test trials in total after eight practice items with feedback. Out of 80 trials, 40 word items were color names congruent with the font color (Congruent-color condition: e.g., ‘빨강’, red), while 12 were color names incongruent with the font color (Incongruent-color condition: e.g., ‘빨강’, red). In addition to the color names, there were 16 trials of object names (Non-color word: e.g., ‘책상’, a table) and 12 trials of nonce words or meaningless characters (Nonce word: e.g., ‘ㄱㄱ’ /kk/). The individuals' mean RTs for the correct responses were calculated in each condition. To represent the individual participants' cognitive cost of inhibitory control, we estimated the mean

RT differences between Incongruent-color and Congruent-color conditions, and the mean RT differences between Incongruent-color and Non-color word conditions. The Stroop task was administered to high school and college students but not to elementary school students due to a practical reason of the time restriction: The whole experiments including speech perception and EF tasks took approximately an hour and a half to complete, and most children could not stay focused for the Stroop task.

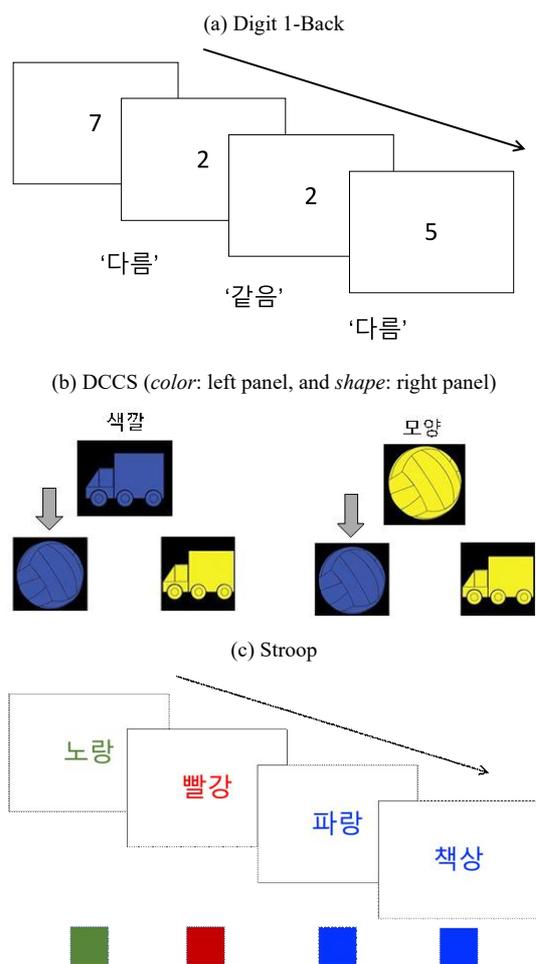


Figure 1. Illustrations of the three executive function tasks administered: (a) Digit N-Back, (b) DCCS (dimensional change card sorting), and (c) the Stroop task.

2.3. Statistical assessment for the relationship between EF and acoustic sensitivity

Partial correlation tests were conducted to quantify the relationship between perceptual sensitivity and EF scores. The partial correlation test was chosen to accommodate the fact that individuals' VOT coefficients were correlated with $f0$ coefficients as illustrated in Appendix 2. This analysis cancels the effect of a control variable in correlating two test variables, quantifying the strength of the relationship between the variables independent of a control factor. For instance, the partial correlation coefficients between EF scores and VOT coefficients with $f0$ coefficient controlled indicate how strongly VOT sensitivity is correlated with EF scores independently of $f0$.

In the partial correlation analyses for high school and college

students, we excluded accuracy scores of DCCS, Digit 1-Back, and Stroop tasks from the set of numeric indices of the EF capacity due to a ceiling effect. In these EF measures, the average counts of incorrect responses were less than two items with small standard deviations, providing too little variability to relate perceptual sensitivity to acoustic cues (see Appendix 3 for the summary of means and standard deviations in each task). The analyses were performed in R using *ppcor* package (Kim, 2015).

3. Results

3.1. Correlations between EF and acoustic sensitivity

Table 2 presents the results of partial correlation tests between EF task scores and acoustic cue coefficients (VOT and $f0$) from the mixed-effects regression models. We present the test results whose coefficients are statistically significant: $p < .05$.

As shown in Table 2, there were several test sets of significant partial correlation coefficients. Separated by a dialect factor, Table 2(a) shows that Gyeongsang speakers' EF scores were significantly correlated with VOT coefficients when $f0$ coefficients was controlled (but none of EF scores was correlated with $f0$ when VOT was controlled). Recall that VOT was a primary acoustic cue for Gyeongsang speakers across all the stop analysis pairs. Looking into the type of EF sub-component, college students with higher N-Back accuracy scores were more sensitive to VOT than others in identifying the lax-aspirated stops, indicating that those listeners with better WM capacity were better at utilizing a primary cue for the phonological contrast. Similarly, high school students with higher N-Back accuracies were more sensitive to VOT than others. For the lax-tense stop pair, VOT was as important as $f0$ in Gyeongsang dialect (see the coefficient summary in Appendix 1). High school students who scored greater cognitive cost (i.e., longer RTs in *Shift* than *Stay* trials) in the DCCS task also used VOT more than others, confirming that better cognitive flexibility was associated with a perceptually dominant dimension in the tense-aspirated stop perception. Elementary students in Gyeongsang did not show any consistent relationship between their EF scores and acoustic cue coefficients.

While Gyeongsang speakers' EF scores were relevant to VOT coefficients only, Seoul speakers' EF scores were correlated not only with VOT coefficients but also with $f0$ coefficients, as presented in Table 2(b). For the lax-aspirated stop perception, high school and college students' smaller cognitive cost (i.e., short RT differences between Incongruent-color and congruent or non-color conditions) in the Stroop task was associated with greater VOT. Likewise, elementary school students' higher N-Back accuracies were significantly correlated with greater $f0$ coefficients. The results from the lax-aspirated stop models show that Seoul listeners with better EF capacity such as WM and inhibition control were better at utilizing both available acoustic cues for the stop contrast than others. Results came from Seoul listeners of all three education institutions tested.

For the tense-aspirated model, less cognitive cost in the Stroop task was associated with college and high school students' greater sensitivity to VOT but with less sensitivity to $f0$. These opposite patterns of EFs associated with VOT and $f0$ may be due to the different role of each cue in differentiating the tense-aspirated stops in the perception. While VOT is a dominant acoustic dimension for the contrast, $f0$ is an acoustic dimension irrelevant to a distinction of

the tense-aspirated stops for the Seoul listeners. As a group, the fixed effect f_0 coefficient was not significant in explaining listeners' choice of /t'-/t^h/ (f_0 : $\beta=.179$, $SE=.146$, $p=.221$), and individual listeners' f_0 coefficients showed little variability (Appendix 2, rightmost panels). It may be that Korean listeners with better inhibition control were able to efficiently suppress unnecessary attention to irrelevant acoustic information for the phonological contrast. The result is congruent with Kong & Yoo (2017) where elementary school children with better EFs attended to f_0 less than others in differentiating the tense-aspirated stops. Finally, no significant relationship was found between acoustic cue coefficients and EF scores for the tense-lax stop models.

Table 2. Partial correlation coefficients between EF tasks (accuracy and RT measures from the sub-sessions) and acoustic sensitivity (VOT and f_0 coefficients). 'r', 'p' and 'N' indicate partial correlation coefficients, p-values, and sample numbers, respectively

(a) Gyeongsang						
Group	EF task	Measure	Cue	r	p-value	N
Lax-Aspirated						
College	N-Back (3-Back-Target)	ACC	VOT	0.38	0.04	32
Tense-Lax						
Highschool	N-Back (3-Back-NonTarget)	ACC	VOT	0.44	0.038	23
Highschool	N-Back (2-Back-NonTarget)	ACC	VOT	0.43	0.047	23
Tense- Aspirated						
Highschool	DCCS	RT	VOT	-0.48	0.034	20
(b) Seoul						
Group	EF task	Measure	Cue	r	p-value	N
Lax-Aspirated						
Elementary	N-Back (1-Back-Target)	ACC	f_0	0.56	0.01	21
Elementary	N-Back (1-Back-NonTarget)	ACC	f_0	0.49	0.03	21
Highschool	Stroop (Incongruent-Congruent)	RT	VOT	-0.47	0.04	19
College	Stroop (Incongruent-Non color)	RT	VOT	-0.36	0.02	39
Tense-Aspirated						
Highschool	Stroop (Incongruent-Congruent)	RT	f_0	0.495	0.036	19
College	Stroop (Incongruent-Non color)	RT	f_0	0.415	0.009	39
College	Stroop (Incongruent-Non color)	RT	VOT	-0.321	0.049	39

4. Discussion & Conclusion

The current study examined how Korean listeners' sensitivity to acoustic cues in stop perception was correlated with their EF capacity in order to understand the role of cognitive resources in explaining the use of multiple acoustic cues. Our specific goal was

to compare the Korean listeners in different social settings (dialect regions: Seoul vs. Gyeongsang and educational institutions or age: elementary, high school and college students), who are known to weight the two acoustic cues, VOT and f_0 differently in the stop perception. Individual level statistical analyses revealed that listeners' better EF control was positively associated with greater use of primary acoustic cues (VOT or f_0) reflecting dialect-specific cue weightings. Similarly, Korean listeners' better EF ability was consistently related to the effective suppression of an uninformative acoustic cue. That is, listeners with better EF control were better at making the most of primary acoustic information that helps differentiate the stop categories with little distraction by irrelevant information. The findings from the Korean listeners support that a general cognitive ability do play a role in speech perception by efficiently facilitating more use of primary acoustic cues.

One notable pattern was that Gyeongsang listeners' EF scores (regardless of education institution) were related to VOT only, while Seoul listeners' EF scores were significantly correlated with both VOT and f_0 . This asymmetrical pattern between the dialect regions may confirm the well-known differential perceptual weights on VOT and f_0 in the dialect groups when they identify Korean stop laryngeal categories. Previous studies have shown that Gyeongsang speakers primarily use VOT for the stop contrast in production and perception (Lee et al., 2013), although there is recent evidence that young adults adapt f_0 relatively more often than older adults in production (Lee, 2020). Similarly, the current dataset yielded a VOT dominant perception pattern across Gyeongsang listeners: f_0 coefficients were smaller than those of VOT across sub-groups both at the group and individual-level analyses. The only exception was when high school students discriminated the lax-tense stop categories ($\beta_{f_0}=2.24$, $\beta_{VOT}=2.14$, see Appendix 1), but the coefficient difference was too small to say f_0 was a dominant cue over VOT. Differently from Gyeongsang speakers, Seoul listeners did not use a single dominant acoustic dimension for the category identification but both dimensions of VOT and f_0 were relevant. At the group level, f_0 and VOT were the two comparably important dimensions either by yielding similar coefficients of VOT and f_0 variables or by using VOT or f_0 primarily depending on the type of stop contrast pairs (Appendix 2). In terms of individuals' patterns, some listeners weighted f_0 over VOT, and others used VOT over f_0 . Although an enhanced use of f_0 characterizes the sound change in the stop laryngeal contrast of the standard Seoul Korean, it is also true that VOT still is an important perceptual dimension for the contemporary Seoul listeners (Kong & Lee, 2018). The current results support that general cognitive ability measured via EF tasks is linked to the acoustic cue(s) that listeners dominantly attend to: Cognitive resources may be allocated to control of dominant acoustic dimensions in speech perception.

When we define f_0 as a secondary cue in Gyeongsang listeners' stop perception, there was no consistent evidence that sub- components of EF control were correlated with listeners' use of a secondary cue. Gyeongsang listeners who had better EF scores did not use any secondary acoustic cue more than others or suppressed them. This does not follow the pattern from English speakers' stop perception, where adult listeners' better EF control was associated with more use of f_0 , a secondary cue for the voiced-voiceless stop perception (Kapnoula et al., 2017). To take a conservative stance, despite this seeming discrepancy between two studies, it is premature to regard this as evidence for cross-linguistic difference between the

languages. It may be possible that the sample size in the present study is too small to observe a consistent relationship between individuals' secondary cue use and EF scores: Compare 131 English speaking adults in Kapnoula's study with at most 32 Gyeongsang Korean speakers in the current study. A follow-up study with enough participants needs to be done to find statistically reliable pattern of association (or a lack of association) between secondary cues and cognitive ability.

In addition to informative acoustic cues, Korean listeners' general cognitive ability was associated with an uninformative acoustic cue by suppressing potentially distracting information. Seoul high school and college students with better EF control consistently used *f0* less than others in differentiating the tense from the aspirated stops. *f0* does not help distinguish the tense-aspirated stop pair, although it is useful for Seoul listeners to identify the other stop pairs. This pattern is congruent with the earlier finding from Korean children (aged 7–9 years) in Kong & Yoo (2017), where a *f0* suppression in tense-aspirated stop perception was associated with better EF ability manifested in WM, mental flexibility, and inhibition. Since the same relationship was observed in adult listeners as well as children, it is suggested that the suppression of *f0*, a potential distractor in speech perception, is not likely to be a developmental characteristic only specific to child listeners but an efficient perceptual process modulated by one's cognitive control.

It is noted that while we explored a number of numeric measures from EF tasks in the partial correlation tests, there were only several EF sub-components that were significantly correlated with acoustic cues. This insufficient evidence makes us less confident to generalize the relationship described in the present study. The insufficient evidence of relationship may be due to a small sample size of the current study for individual analyses (e.g., 19 child listeners in the Seoul group). Or more generally, it may be that the EF scores and subtle phonetic details are related only weakly across individuals within the cohorts. Since there is not enough accumulated knowledge regarding the EF control and speech processing, we will take a conservative stance in discussing the current results from the partial correlation tests.

Finally, there was no child-specific pattern of relationship. The only correlated combination was the WM control associated with *f0* sensitivity in the Seoul children's perception of the lax-aspirated stops. Resembling the older Seoul listeners' pattern, Seoul children showed that higher EF scores were correlated with *f0*, which were weighted over VOT. Although Gyeongsang children's use of VOT or *f0* were not statistically different from Seoul children (Kong et al., under revision, Appendix 1), none of the cues was correlated with better EF controls. Given the small samples, it is not reasonable to generalize that the results reflect a dialect difference. The current investigation was exploratory, and we may need a larger sample of child participants to confirm that this null evidence is reliable.

To conclude, the present study explored the relationship between general cognitive ability and fine-grained sub-phonemic details across dialect/education institution groups speaking Korean. It turned out that Korean listeners having better EF control were able to better utilize dominant acoustic details in phoneme identification, and the relevant acoustic dimensions matched dialect-specific perceptually dominant cues regardless of institution groups. The findings confirm the link between speech perception and cognitive ability, adding experimental evidence from Korean language to the existing knowledge. While the present study has a value as an

attempt to explore understudied research area, it is hoped that further research with more diverse population and stronger statistical supports enables us to solidify the present findings.

Acknowledgement

The speech perception data from the Korean stop 3AFC tasks has been previously presented in authors' prior study (Kong et al., under revision). The present study uses it to pursue a different research goal, i.e., the role of cognitive capacity in speech perception. The main data of the research is the EF scores of the same Korean participants, which have never been reported before. We are deeply grateful for institutions and teachers at Changwon and Seoul, who allowed us to recruit participants and use test spaces. We also thank graduate students (Jieun Kang, Hyangwon Lee and Bokyoung Park), who helped with the data collection in both cities.

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.
- Darcy, I., Mora, J. C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning, 66*(4), 741-773.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168.
- 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.
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6 - 12. *Journal of Phonetics, 28*(4), 377-396.
- 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.
- Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance, 37*(6), 1939-1956.
- Kang, Y. (2014). Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics, 45*, 76-90.
- Kapnoula, 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, S. (2015). ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Communications for Statistical Applications and Methods, 22*(6), 665-674.
- 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., & Yoo, J. (2017). Speech processing strategy and executive function: Korean children's stop perception. *Phonetics and Speech Sciences*, 9(3), 57-65.
- Kong, E. J., Holliday, H., & Lee, H. (under revision) Post-adolescent changes in the perception of regional sub-phonemic variation. *Journal of Phonetics*.
- Lee, H. (2020). Cross-generational acoustic comparisons of tonal Kyungsang Korean stops. *The Journal of the Acoustical Society of America*, 148(2), EL172-EL178.
- Lee, H., Holliday, J. J., & Kong, E. J. (2020). Diachronic change and synchronic variation in the Korean stop laryngeal contrast. *Language and Linguistics Compass*, 14(7), e12374.
- 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.
- Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, 29(1), 3-11.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49-100.
- Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N back working memory paradigm: A meta analysis of normative functional neuroimaging studies. *Human Brain Mapping*, 25(1), 46-59.
- Repp, B. H. (1983). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication*, 2(4), 341-361.
- Silva, D. J. (2006). Variation in voice onset time for Korean stops: A case for recent sound change. *Korean Linguistics*, 13(1), 1-16.
- van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture – word interference are two sides of the same coin. *Psychonomic Bulletin & Review*, 16(6), 987-999.
- Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1990). Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica*, 47(1-2), 36-49.
- Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., Chiang, J. K., ... Carlson, S. M. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, 68(3), i-151.

• **Eun Jong Kong**

Professor, Dept. of English
 Korea Aerospace University
 76 Hanggongdaechang-ro, Deogyang-gu, Goyang 10540, Korea
 Tel: +82-2-300-0338
 Email: ekong@kau.ac.kr
 Fields of interest: Acoustic Phonetics, Speech Production & Perception, Language Acquisition

• **Hyunjung Lee**, Corresponding author

Associate Professor, Dept. of English Education
 Incheon National University
 12 Gaebeol-ro, Songdo-dong, Yeonsu-gu, Incheon 21999, Korea
 Tel: +82-32-835-8992
 email: hjlee123@inu.ac.kr
 Fields of interest: Acoustic Phonetics, Speech Production & Perception

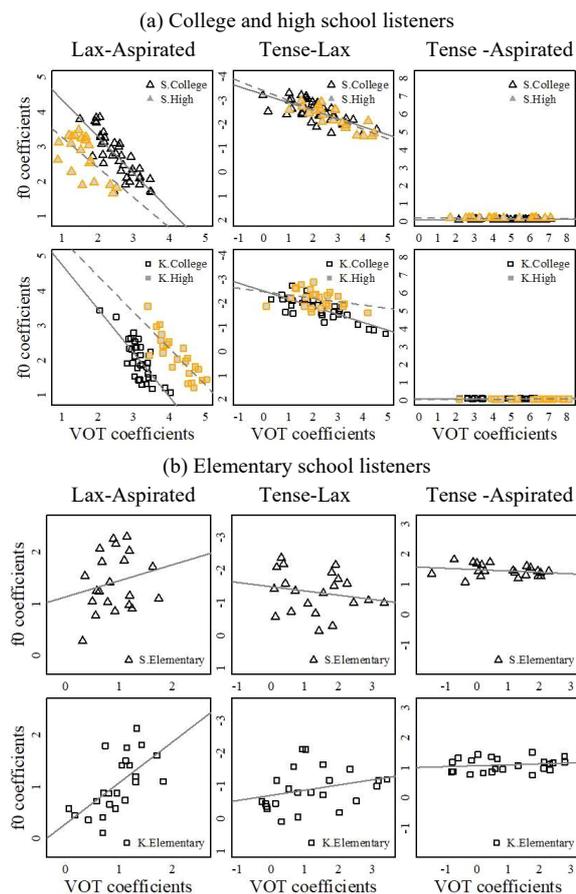
Appendix 1

A summary table of fixed effect coefficients of VOT and f_0 (in a standardized unit) from three types of mixed-effects regression models. The results were originally reported in Kong et al. (under revision) and are presented in this paper to supplement individual analyses. All the coefficients were significant except f_0 coefficients (in *italic*) for tense-aspirated models.

	Seoul		Gyeongsang	
	VOT	f_0	VOT	f_0
<Lax-Aspirated>				
College	2.576	2.733	3.194	2.025
High school	1.650	2.770	4.232	2.135
Elementary school	0.948	1.458	0.935	1.046
<Lax-Tense>				
College	2.160	-2.568	2.254	-1.831
High school	2.836	-2.331	2.143	-2.247
Elementary school	1.456	-1.372	1.22	-1.453
<Tense-Aspirated>				
College	5.288	0.106	4.749	0.106
High school	4.956	0.179	6.515	0.073
Elementary school	2.518	0.137	1.927	0.075

Appendix 2

Scatterplots of f_0 coefficients against VOT coefficients (re-structured panels based on Kong et al. (under revision)). Each datapoint indicates individual listeners. (a) VOT and f_0 coefficients of high school and college students, and (b) VOT and f_0 coefficients of elementary school students. Note that for the lax-aspirated models, panels for the elementary school students show positively correlated distributions between VOT and f_0 coefficients, while those for the older listeners display negatively correlated distributions.



Appendix 3

Means and standard deviations of accuracy scores and response time (RT) across participant groups in each executive function tasks.

(a) Digit 1-Back					
Dialect	School	Mean	SD	Mean	SD
<i>Target</i>		Accuracy		RT	
Gyeongsang	Elementary	6.69	2.49	1,011.95	281.81
	Highschool	8.21	1.44	641.59	157.36
	College	8.09	1.70	558.24	137.12
Seoul	Elementary	6.09	2.50	946.78	229.17
	Highschool	7.95	1.53	597.21	244.94
	College	8.51	1.73	561.33	111.67
<i>Non-target</i>		Accuracy		RT	
Gyeongsang	Elementary	17.86	6.91	1,012.06	234.08
	Highschool	27.30	2.43	703.54	170.04
	College	25.69	4.85	573.70	133.47
Seoul	Elementary	21.57	5.68	912.21	143.20
	Highschool	26.52	3.53	598.48	156.94
	College	28	1.23	556.04	110.31
(b) Digit 2-Back					
Dialect	School	Mean	SD	Mean	SD
<i>Target</i>		Accuracy		RT	
Gyeongsang	Highschool	7.60	1.67	850.83	186.65
	College	6.45	2.82	740.09	217.15
Seoul	Highschool	7.09	2.64	681.93	275.02
	College	7.82	2.23	875.45	309.36
<i>Non-target</i>		Accuracy		RT	
Gyeongsang	Highschool	26.04	2.88	952.65	209.57
	College	24.09	5.34	781.28	267.11
Seoul	Highschool	25.04	6.19	744.35	289.77
	College	25.76	2.74	884.69	288.39
(c) Digit 3-Back					
Dialect	School	Mean	SD	Mean	SD
<i>Target</i>		Accuracy		RT	
Gyeongsang	Highschool	6.78	2.08	863.13	244.50
	College	5.15	2.04	873.00	270.77
Seoul	Highschool	6.00	2.42	753.84	362.95
	College	5.92	2.15	961.73	308.42
<i>Non-target</i>		Accuracy		RT	
Gyeongsang	Highschool	24.30	2.68	905.09	202.72
	College	21.75	4.97	866.30	296.92
Seoul	Highschool	21.71	5.67	825.91	364.93
	College	23.51	3.29	985.70	316.89
(d) DCCS					
Dialect	School	Mean	SD	Mean	SD
<i>Switch</i>		Accuracy		RT	
Gyeongsang	Elementary	9.47	1.88	1,004.51	182.69
	Highschool	11.30	0.97	588.46	166.49
	College	11.71	1.30	625.20	177.23
Seoul	Elementary	9.04	1.82	921.58	327.82
	Highschool	9.88	1.96	653.28	173.03
	College	11.23	1.13	573.44	111.00
<i>Stay</i>		Accuracy		RT	
Gyeongsang	Elementary	12.85	1.79	933.74	216.25
	Highschool	14	1.07	520.94	110.21
	College	14	0.91	571.38	136.98
Seoul	Elementary	13.28	1.70	814.09	232.15
	Highschool	14.05	1.86	554.59	99.93
	College	14.35	0.66	532.05	105.16

(e) Stroop					
Dialect	School	Mean	SD	Mean	SD
<i>Congruent color word</i>		Accuracy		RT	
Gyeongsang	Elementary	39.20	1.87	937.11	342.73
	Highschool	39.20	1.23	557.75	75.10
	College	39.68	0.59	574.63	69.06
Seoul	Elementary	39.31	1.05	803.32	116.81
	Highschool	37.84	1.83	513.63	56.79
	College	39.33	0.98	528.05	79.05
<i>Incongruent color word</i>		Accuracy		RT	
Gyeongsang	Elementary	11.60	0.84	1,074.75	526.36
	Highschool	11.60	0.68	616.46	95.32
	College	11.53	0.71	626.23	127.66
Seoul	Elementary	11.52	0.69	886.29	175.19
	Highschool	11.21	0.78	545.34	90.92
	College	11.46	0.94	573.29	106.60
<i>Non-color word</i>		Accuracy		RT	
Gyeongsang	Elementary	15.50	1.26	1,010.71	505.05
	Highschool	15.75	0.44	572.18	82.42
	College	15.81	0.39	590.80	74.00
Seoul	Elementary	15.73	0.56	844.96	128.90
	Highschool	14.89	1.41	550.81	77.35
	College	15.56	0.68	551.84	92.05
<i>Nonce word</i>		Accuracy		RT	
Gyeongsang	Elementary	11.40	0.84	1,021.14	243.97
	Highschool	11.45	0.82	589.27	86.07
	College	11.87	0.33	592.52	83.97
Seoul	Elementary	11.73	0.56	864.03	135.91
	Highschool	11.15	1.06	572.08	76.72
	College	11.53	0.88	559.00	86.99