ISSN 2005-8063
2017. 9. 30.
Vol.9 No.3
pp. 57-65

말소리와 음성과학

Phonetics and Speech Sciences

한국음성학회지

http://dx.doi.org/10.13064/KSSS.2017.9.3.057

Speech processing strategy and executive function: Korean children's stop perception

Eun Jong Kong · Jeewon Yoo*

Abstract

The current study explored how Korean-speaking children processed the multiple acoustic cues (VOT and *f0*) for the stop laryngeal contrast (/t'/, /t/, and /t^h/) and examined whether individual perceptual strategies could be related to a general cognitive ability performing executive functions (EF). 15 children (aged from 7 to 8) participated in the speech perception task identifying the three Korean laryngeal stops (3AFC) on listening to the auditory stimuli of C-/a/ with synthetically varying VOT and *f0*. They completed a series of EF tasks to measure working memory, inhibition, and cognitive shifting ability. The findings showed that children used the two cues in a highly correlated manner. While children utilized VOT consistently for the three laryngeal categories, their use of *f0* was either reduced or enhanced depending on the phonetic categories. Importantly, the children's processing strategies of a *f0* suppression for a tense-aspirated contrast were meaningfully associated with children's better cognitive abilities such as working memory, inhibition, and attentional shifting. As a preliminary experimental investigation, the current research demonstrated that listeners with inefficient processing strategies were poor at the EF skills, suggesting that cognitive skills might be responsible for developmental variations of processing sub-phonemic information for the linguistic contrast.

Keywords: child speech processing, Korean stop laryngeal contrast, acoustic cue-weighting, executive function, working memory, inhibition control, attention shift

1. Introduction

While multiple acoustic information exists for linguistic contrasts in the speech stream, listeners do not pay equal amounts of perceptual attention to all the available acoustic cues (e.g., Summerfield & Haggard, 1974; Lisker, 1978; Repp, 1983; Walley & Carrell, 1983; Kingston & Diehl, 1994; Holt & Lotto, 2006; Francis *et al*., 2008). The listeners weigh one acoustic dimension more than others in identifying contrastive speech categories. Prior experimental studies have shown that, in a typical listening condition, English-speaking adult listeners, on average, attended to voice onset time (VOT, Lisker & Abramson, 1964) more than fundamental frequencies (*f0*) at the post-consonantal vocalic onset for a voice-voiceless contrast (e.g., Gordon *et al*., 1993; Francis *et al*., 2008), and relied on spectral information (vowel formants) more than on duration for a tense-lax vowel contrast (e.g., Flege *et al*.,1997; Cebrián, 2006; Escudero & Boersma, 2004).

The current research explored the related fact that the order of perceptual primacy among the multiple cues is language-specifically defined, which becomes what language learners are to acquire (or re-adjust) in the course of a mastery of the target language (Yamada & Tohkura, 1992; Flege *et al*.,1997; Cebrián, 2006; Mayr & Escudero, 2010 for L2 mastery; Walley & Carrel, 1983; Nittrouer & Miller, 1997; Hazan & Barret, 2000; Nittrouer, 2002; Mayo & Turk 2004; Li, 2012 for L1 mastery). Specifically focusing on the developmental characteristics of perceptual cue-weightings, the current study explored Korean-acquiring children's (aged from 7 to

* Ewha Womans University, jeewonyoo@ewhain.net, corresponding author
Received 12 August 2017 ; Revised 18 September 2017; Accepted 22 September 2017

8) processing strategies of two acoustic cues, VOT and *f0*, in identifying three Korean laryngeal stops (/t'/, /t/, and /t^h/). We were mainly interested in examining Korean language-specific patterns of children's acoustic cue processing and in exploring the child listener's cognitive ability as one factor to account for the children's perception strategies.

1.1. Children's acoustic cue weightings

Developmentally, children are known to acquire native-like relative cue-weighting strategies with experiences although initially attended acoustic dimensions might not necessarily be identical between child and adult populations (Walley & Carrel, 1983; Nittrouer, 1992, 2002; Nittrouer & Miller, 1997; Hazan & Barret, 2000; Li, 2012). Nittrouer and colleagues (1992, 1997, 2002) showed that in perceiving /s/-/ʃ/ in a prevocalic condition, children (aged 3, 5 and 7) tended to attend to a F2 transitional cue present in the following vowel more than to a spectral cue (i.e., fricative energy concentration) present in the fricative signal. They argued that this processing pattern reflected a developmental state of perceiving phonetic units where they were not analytic enough to separate out the consonant independently from the vowel. The formant transition cue is to be more important for young language learners than the consonantal cue because of children's global processing of the speech signal.

Prior studies documented experimental evidence of developmental weighting shifts that cue-processing strategies interacted with the age (e.g., Walley & Carrel, 1983; Hazan & Barret, 2000; Mayo & Turk, 2004). Mayo & Turk (2004) showed that English-speaking children put more weights on a transitional cue than a static cue for a /ta/-/da/ contrast as adults did but the magnitude of children's reliance on the primary cue was not as great as that of adults' and the magnitude of children's reliance on the secondary cue was not as small as that of adults'. Similarly in Riki *et al*. (2007), Hebrew-learning children (aged 4~10) utilized the multiple acoustic cues (VOT, formant transition, and burst) for a voicing contrast in the same order as adults did but the magnitude of perceptual weights on each acoustic cue was smaller in younger listeners, indicating children's less sensitivity to linguistic information in the perception. In Hazan & Barret (2000), English-speaking children (aged 3~12) significantly underperformed the categorical perception task (/t/-V vs. /d/-V) in the single-cue condition (i.e., condition with limited acoustic information) compared to the multiple-cue condition. The finding suggested that child listeners in this age were less able than adults to flexibly adapt to available acoustic cue information.

Given the background, the current study questioned whether and how Korean-learning children would show language-specific patterns of processing the stop categories with a three-way laryngeal contrast, i.e., tense, lax and aspirated stops. This research question with Korean children's stop perception deserves attention because of Korean language–specific roles of acoustic cues (VOT and *f0*, in particular) in representing a stop laryngeal contrast. A number of production studies have documented that Korean laryngeal categories were differentiated by *f0* as well as VOT, as the lax stops (long-lag VOT) were distinguished from the aspirated stops (also long-lag VOT) by having a lower *f0* at the following vowel onset (Silver, 2006; Kang, 2014; Kong *et al*., 2011). The *f0* cue also played an important role in the adults' perception of the lax stops (e.g., Schertz *et al*., 2015). The prior findings mostly with adult listeners indicate that it is not a single cue (e.g., VOT in English) but at least two cues (VOT and *f0*) that are equally primarily activated in distinguishing the three stop categories one after another. It has never been reported, to our knowledge, how children process these two cues in perceiving the stops so far. As a preliminary step to investigate how Korean children learn to deal with multiple acoustic cues, the current research examined Korean-speaking children aged from 7 to 8 (who are mature enough to handle the behavioral perception tasks) by conducting the identification task of the stop categories.

1.2. Speech processing and executive function

The current study attempted not only to describe the processing strategies in Korean-speaking children, but also to explore the children's cognitive ability via Executive Function (EF) tasks as an explanatory factor of the processing patterns.

EF refers to the high level cognitive processes involving a self-directed thought and action toward the goal (Barkley, 2012; Diamond, 2013; Friedman *et al*., 2006). A widely accepted EF model is the integrative model, suggested by Miyake *et al*. (2000). In this model, EF is proposed as three separable and also moderately correlated components including updating of working memory (WM), inhibition and shifting attention. WM refers to the ability to maintain the verbal or nonverbal information while involving other cognitive activity (Baddeley & Hitch, 1974; Baddeley, 2003). Some examples of WM task are backward-digit span task, reading span task, corsi block test, counting span task, etc. (Alloway, 2007; Daneman & Carpenter, 1980; Lezak, 1983). Inhibition refers to the ability to focus attention to the given target while ignoring the irrelevant stimuli (Diamond, 2013) and this ability is measured by stroop task, simon task, flanker task, etc. (Eriksen & Eriksen, 1974; Hommel, 2011). Finally, shifting refers to the ability to flexibly change perspective or thought to adjust to the given demands (Diamond, 2013). The frequently used tasks to measure the shifting ability are Wisconsin Card Sorting task and Dimensional Change Card Sort (DCCS) Test (Stuss *et al*., 2000; Zelazo *et al*., 2003).

EF has been of great interest in various fields of developmental research because of its close link to speech and language abilities. The associations between the perceptual level of speech processing and cognitive abilities have been reported in the bilingual population (Lev-Ari & Peperkamp, 2013) and in the population with hearing impairment (Kronenberger *et al*., 2013; Nittrouer *et al*., 2013). For instance, Kronenberger *et al*. (2013) recruited children with Coclear Implant (CI) and children with Normal Hearing (NH) to examine the relationship between speech-language skills and EF. All children were administered the speech perception tasks (isolated word recognition task and sentence recognition tasks), the language tasks, and the EF tasks (verbal and nonverbal WM tasks, fluency-speed tasks, and inhibition tasks). Correlational analyses revealed that there were significant relationships between speech perception scores and EF skills including verbal WM capacity and fluency-speed, suggesting EF may play an important role in transmitting the speech signals to the auditory nerve system by rapid encoding and processing in children with CI.

Also, the indirect link between more fine-grained speech perception processing (perceptual weighting strategies) and cognitive ability has been reported (Nittrouer & Burton, 2005). Nittrouer & Burton (2005) investigated whether the early language experience would facilitate the efficient usage of language-specific perceptual weighting strategies, and further examined whether these perceptual strategies would enhance storage and retrieval of words in working memory system. They manipulated language experience by recruiting children aged from 4 to 5 years, who were divided into four groups: low socio-economic status (SES) group, chronic otitis media (OME) group, both SES and OME group, and control group. Children completed a series of speech and language tasks including speech perception task, phonological awareness task, verbal WM task, and comprehension task. Results showed that the less language experienced group (OME and SES groups) weighted a formant transition cue more than the control group and the OME group made more errors on the verbal WM task and the comprehension task than the control group, suggesting the circular link among language input, differential perceptual strategy use of speech, working memory skill, and comprehension. Based on these findings, it is reasonable to hypothesize that there may exist the bidirectional link between these two components of speech system and cognition system although the relationship between perceptual strategy (cue weighting in speech) and EF was not directly tested.

Together, it seems to be clear that the sub-phonemic level of speech processing is the fundamental component that subserves the language development and that children's EF skills and language processing are closely linked. However, little is known about how Korean-speaking children perceive and process native speech sounds at the sub-phonemic level. Especially, it is hard to find research about the associations between EF and speech processing in Korean children. Thus, we planned to investigate the potential relationship between EF and sub-phonemic level of speech processing in Korean young children, which would be the first step to fill the gap in literature of Korean speech processing and cognition system in young children.

1.3. Current study

Despite the critical role of speech processing strategies in efficiently accessing to phonological system and its impact on appropriate language development, it is rare to find research examining Korean-speaking children's perceptual processing strategies of the speech and the developmental characteristics of Korean-speaking children's perception patterns. Likewise, with respect to the Korean-speaking children's performance of EFs, studies scarcely exist examining the relationship between the EF scores and speech perception — although there has been studies that investigated the EF and literacy skills as a global indication of language skills.

Therefore, in the current study, we attempted to document experimental evidence of the relationship between speech processing strategy and cognitive ability by testing how child listeners would utilize cue weighting strategy in perceiving Korean stop categories and whether the individuals' sensitivity to the acoustic information would be related to their performance of EF tasks.

2. Method

2.1. Participants

15 Korean-speaking elementary school children (7:25~8:83, mean = 8.07, first or second grade) participated in the speech perception tasks and a series of EF tasks. All of participants were within normal ranges in their vocabulary ability with a mean of 91.80 (SD=9.87) for expressive vocabulary and a mean of 95.47 (SD=13.54) for a receptive vocabulary indexed by Receptive & Expressive vocabulary Test (REVT; Kim *et al*., 2009). Also, their non-verbal IQ was in a normal range with a mean of 112.47 (SD=13.40) indexed by Kaufman Assessment Battery for Children (K-ABC, Moon & Byun, 1997).

2.2. Auditory stimuli

We used a subset of the existing stimuli (stop-/a/) that has been created and tested by the first author (Kong *et al*., 2011; Kong & Yoon, 2013; Kong & Kang, 2015; Kong & Edwards, 2016). The original stimulus set was prepared by acoustically manipulating a natural production of /ta/ by a 30-year old Korean male speaker. For the VOT steps, varying duration of the burst release portion in a /t^{h}a/ production were copied and pasted immediately after the burst release of the /ta/ token. For the *f0* steps, flat values were overlaid over the vowel. The manipulations were implemented in Praat (Boersma & Weenink, 2013) using a built-in Pitch Tier Manipulation function. Out of the complete set of 30 CV syllables (6-step VOTs [9ms, 13ms, 19ms, 28ms, 40ms, 59ms] X 5-step *f0*s [98Hz, 106Hz, 114Hz, 122Hz, 130Hz]), twelve different stop-/a/ syllables were chosen for the sake of child listeners' attention span by considering distributions of the acoustic variables: Each of four-step VOTs (9ms, 19ms, 28ms, and 59ms) was combined with three-step *f0*s (98Hz, 114Hz and 130Hz).

2.3. Tasks and procedure

Participants performed both a speech perception task and multiple EF tasks. A three-alternative forced-choice (3AFC) identification task was designed to test the speech perception. Sitting in front of a lap-top computer, child participants listened to auditory stimuli played over a headphone and chose one response out of three given consonant character options displayed on the monitor, namely, "ㄷ" (/t/), "ㅌ"(/t^h/), and "ㄸ"(/t'/). This 3AFC task was programmed and conducted at E-prime 2.0 software platform (Psychology Software Tools, Pittsburgh, PA) to present the stimuli at regular intervals and to record the responses automatically. Twelve different CV syllables were repeated three times in a random order.

For the executive tasks, working memory, inhibition, and shifting abilities were measured. To measure the WM ability, a digit n-back task for verbal working memory and a *visual n-back* task for visual working memory were used (Owen *et al*., 2005; Vanderplas & Garvin, 1959). For these tasks, child participants were asked to respond to the target stimulus (i.e., digit or shape) located in a n position before in a sequence. The participants decided whether the given stimulus is identical to the one in a n position before by pressing either 'the same' or 'different'. *N* increased from 1 to 3 for this study. They were asked to respond as fast and accurately as possible. From the tasks, we obtained accuracy and response time of each trial.

To measure the inhibition ability, a *flanker* task and a *stroop* task were used (Kopp *et al*., 1994; van Maanen *et al*., 2009). For the tasks, three different conditions of stimuli were constructed: (1) a congruent condition (where both flanker fish and a middle fish faced the same direction [*flanker*] vs. the color and the word were matched [*stroop*]), (2) an incongruent condition (where the flanker fish faced the opposite direction to the middle fish [*flanker*] vs. the color and the letter were not matched [*stroop*]), and (3) a neutral condition (where the flanker fish were replaced by squares [*flanker*] vs. where the letter was replaced by the meaningless symbol 'xxx' [*stroop*]). Again, participants were to respond as fast and accurately as possible to be reliably reflective of their inhibition ability. Finally, to measure the shifting ability, a *trail-making* test (TMT) was used. In this task, three conditions were presented including number condition (1~14), letter condition (14 Korean letters) and mixed condition (14 numbers and 14 letters were mixed), and participants were asked to draw a line to connect (a) each number, (b) each letter and mixed (c) letter and number according to their order. The total duration of the task was measured to represent the relevant cognitive ability. All of EF tasks were presented using the SuperLab 5 software (Cedrus Corporation, 2015). It did not take more than 15 minutes, on average, to finish each task.

2.4. Analysis

Mixed-effects logistic regression models (Bates *et al.*, 2015) were constructed to estimate perceptual roles of acoustic parameters (VOT and *f0*) in differentiating the three-way contrastive stops (. Dependent variables were the child participants' response categories of /t'/, /t/ and /t^h/. Since a binary dependent variable was considered, we made three separate models by pairing the categories into tense vs. aspirated, tense vs. lax, and lax and aspirated responses. Fixed effect variables were the two acoustic dimensions, VOT and *f0,* treated as continuous variables. The models included random intercepts and random slopes for VOT and *f0* at the listener level. The analyses were implemented in R (R Core Team), using *lme4* package (Bates *et al*., 2015).

In interpreting the model outputs, the coefficients of each fixed effect parameter indicate the group-averaged perceptual dependency on the cue in identifying one category over the other. Besides the group-level assessments, the individual listeners' deviations from the group-average also have been calculated in the model as random slope coefficients of VOT and *f0*. These individual coefficients represent child listeners' sensitivities to the acoustic cues in the speech perception and were used to correlate with various measures of EF.

For each EF task, accuracies and response times were obtained automatically by the stimuli presentation software. Since there was no ceiling effect either in the target trials or in the non-target trials as well as the accuracies for each condition differed in the n-back tasks, two trial sections were collapsed, and the overall accuracies (global accuracies) for each task across conditions were obtained to capture the accurate WM abilities. In addition, to maintain consistency in analyzing EF tasks, the two numeric measures were averaged across the sub-conditions of each task to have global figures to indicate their cognitive capacities. Final analyses used these global accuracies (from *n-backs*, *flanker,* and *stroop*) and response times (from *n-backs*, *flanker, stroop*, and *trail-making*) as individual characteristics.

A series of partial correlation tests were conducted between individual coefficients from mixed-effects logistic models (based on the speech perception tasks) and individual scores of the EF tasks in order to address relationship between perceptual strategy and cognitive ability. The analysis entered the standardized IQ scores, express vocabulary scores, and receptive vocabulary scores as control variable to consider variability of intelligence and language proficiency among children. The analyses were implemented in R using a *ppcor* package (Kim, 2015).

3. Results

3.1. Speech perception: 3AFC tasks

3.1.1. Descriptive Patterns

The distributions of response categories showed that the lax, the aspirated, and the tense types were frequent in that order (/t/: 47%, /t^h/: 29% and /t'/: 24%), presumably due to the stimuli proportion at the middle range of VOT (19ms and 29ms). The response distributions along VOT and *f0* (<Figure 1>-top panels) illustrated that the auditory stimuli of 9ms VOT × 130 Hz *f0*, 19ms VOT × 98 Hz *f0* and 59ms VOT × 130 Hz *f0* were associated with the tense, lax and aspirated types most frequently, respectively. Consistent with existing knowledge of the Korean stop perception, child listeners' categorical decisions of tense category decreased as VOT values increased, and those of lax category decreased as *f0* values increased (<Figure 1>-bottom). Decision of aspirated categories increased as VOT and *f0* increased.

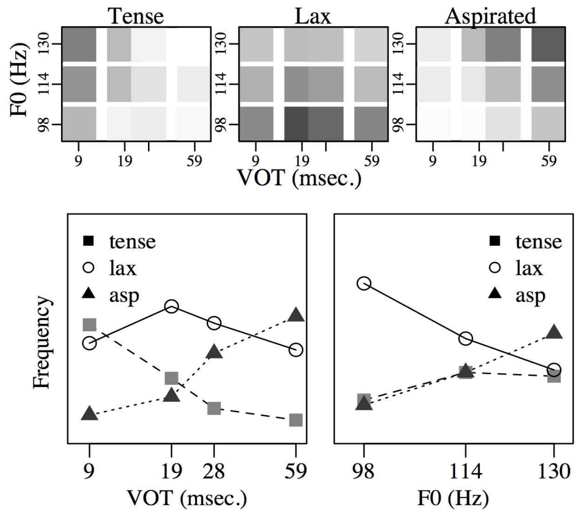

Figure 1. Response distributions of /t'/(tense), /t/(lax) and /t^h/(aspirated) along VOT and *f0* dimensions (top). Response frequencies of the three types as VOT and *f0* change (bottom).

3.1.2. Logistic regression model outputs

The logistic curves are drawn based on the coefficients estimated from the mixed-effects regression models of the three contrast pairs (<Figure 2>). The outputs are summarized in <Table 1>. In differentiating the aspirated stop /t^h/ from the tense stop /t'/ (<Figure 2>-left), child listeners were sensitive to changes of VOT values but not to changes of *f0* values -- the *f0* curve was much shallower than the VOT curve. This pattern is in line with patterns of adult Seoul Korean speakers, reported in Schertz *et al*., (2015) and Kong *et al*., (2011) in that perceptual reliance on *f0* was relatively minimal compared to their reliance on VOT. For the other two models with the lax type involved, listeners were sensitive to both VOT and *f0* — the curves for VOT and *f0* either steeply fell or rose as VOT and *f0* increased (<Figure 2>-center & right) and model-estimated coefficients of VOT and *f0* were statistically significant (<Table 1>). With respect to the relative cue-weighting between VOT and *f0*, the VOT curve was steeper than the *f0* curve in the tense-lax model, indicating that child listeners perceptually depended on VOT more than on *f0* in distinguishing /t/ from /t'/ (β_{VOT}=1.119, β_{f0}=-0.825). By contrast, in the lax-aspirated model, the *f0* curve was steeper than the VOT curve suggesting greater perceptual dependency on *f0* than on VOT in differentiating /t^h/ from /t/ (β_{VOT}=0.991, β_{f0}=1.255). The finding of greater *f0* dependency in child listeners' perception of the lax category is important in that the enhanced perceptual role of *f0* has not been experimented and confirmed in the Korean child listeners' perception.

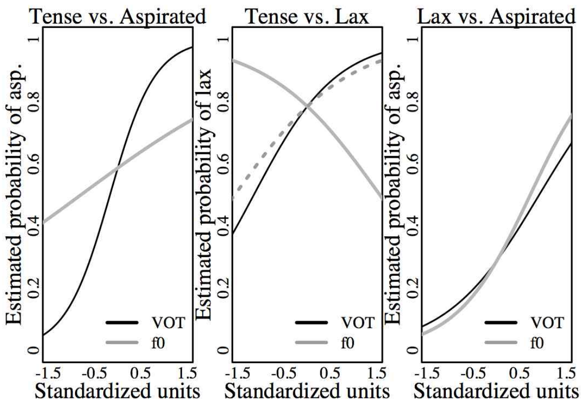

Figure 2. Logistic curves based on the group-averaged estimations from the three mixed-effects logistic regression models. Note that a grey dashed line in the center panel is identical to a grey solid line except the signs.

Besides a group tendency, we examined individual differences of utilizing the acoustic cues in the stop perception by assessing individual listeners' deviations from the group-averaged coefficient estimations of VOT and *f0* parameters from the regression models (<Figure 3>). In the three separate regression models, correlations between VOT and *f0* coefficients were statistically significant, indicating that the two acoustic cues are not completely independent with each other: $r_{(T\ vs.\ A)}$=-0.785 (df =13, p<.001); $r_{(T\ vs.\ L)}$=-0.981 (df =13, p<.0001); $r_{(L\ vs.\ A)}$=0.978 (df =13, p<.0001). With reference to adults' patterns (Kong & Kang, 2017: $r_{(T\ vs.\ A)}$= 0.248, df = 51, p=.072, $r_{(T\ vs.\ L)}$= 0.241, df = 51, p=.081; $r_{(L\ vs.\ A)}$= -0.651, df = 51, p<.0001), child data showed relatively greater degrees of correlation between the cues.

In the model of tense vs. aspirated contrast, the negative direction of the relationship was observed suggesting a trade-off between the two cues across the listeners. That is, listeners who were sensitive to VOT did not rely on *f0* much and vice versa in differentiating the tense from the aspirated types. Given that VOT (but not *f0*) was estimated as an important perceptual dimension for a tense vs. aspirated contrast according to the group-level analysis (see <Table 1>), the VOT listeners appear to be *tactful* listeners who do not attend to the *f0* dimension in the irrelevant linguistic context.

Table 1. Parameter estimations of fixed effect variables from the mixed-effects logistic regression models.

	tense vs. aspirated			tense vs. lax			lax vs. aspirated		
	Estimate	(S.E)		Estimate	(S.E)		Estimate	(S.E)	
intercept	0.348	0.467		1.291	0.362	***	-0.910	0.371	*
VOT	2.072	0.444	***	1.119	0.278	***	0.991	0.259	***
f0	0.439	0.292		-0.825	0.260	**	1.255	0.228	***

Signif. codes: 0 '***'0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The two models of tense vs. lax and lax vs. aspirated contrasts revealed the same relationship between coefficient magnitudes of VOT and *f0* where child listeners who relied on VOT more than others also depended on *f0* cue more (i.e., *positive association*). It is noted that, although this *positive association* between VOT and *f0* was depicted as a downward regression line for tense vs. lax contrast pair in <Figure 3>-center due to a negative sign of *f0* coefficients, the trend line indicates an association of greater use of VOT (plus coefficients) with greater use of *f0* (minus coefficients) in terms of magnitude of perceptual dependency. This finding is different from adult patterns reported in Kong & Kang (2017) where the same experimental protocol (including the same auditory stimuli) has been used with 53 Korean adult native speakers in their 20s. Specifically, the adults' coefficients of VOT and *f0* from the lax-aspirated model were highly correlated similar to children's but the relationship between VOT and *f0* coefficients was negative indicating that adult listeners depended on one cue and suppress the other cue, r=-0.651, df=51, p<.0001. The relationship between VOT and *f0* coefficients from the tense vs. lax contrast was also different from adult patterns (r=0.241, df = 51, p=.081 in Kong & Kang, 2017) in that adults' coefficients of VOT and *f0* were not significantly correlated. Supported by the differences between child and adult listeners' use of the two cues, the correlation patterns of the coefficients suggest that Korean child listeners did not make independent uses of the multiple acoustic cues in the stop perception.

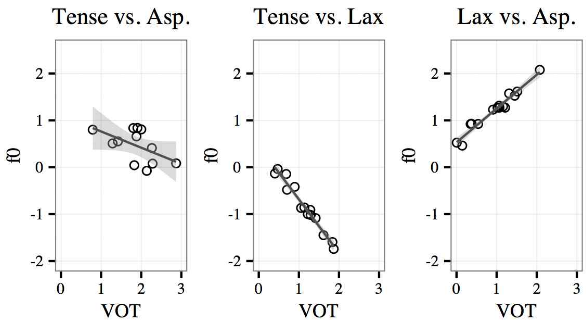

Figure 3. Scatter-plots of VOT and *f0* coefficients estimated from the three mixed-effects logistic regression models. Each data-point represent individual child listeners. In each panel, linear regression lines are overlaid with 95% confidence intervals indicated in grey.

3.2. Relationship between 3AFC and EF tasks.

3.2.1. Descriptive Patterns: EF tasks

In general, children performed each EF task well. As expected, the n-back WM tasks seemed to be more cognitive resource-demanding tasks than the inhibition tasks. Also, the mixed condition seemed to be more demanding condition than the single-stimuli condition (letter or number) in the TMT task. To compare the performance of two different modalities within WM task and inhibition task, the paired *t*-tests were conducted. In the WM tasks, children performed better on verbal WM task (*digit n-back*) than non-verbal WM task (*visual n-back*), t (14) = 2.83, p < 0.05, while there was no difference in RTs between two tasks, t (14) = 0.13, p = 0.90. It seemed to be relatively easier for children to retrieve the verbal information positioned n before the target stimuli compared to the non-verbal information due to the familiar verbal information (digit) in WM task. In the inhibition tasks, children performed better on the non-verbal inhibition task (*flanker*) than the verbal inhibition task (*stroop*) both in accuracies, t (14) = 3.32, p < 0.05, and in RTs, t

(14) = - 7.13, $p < 0.05$. For children, it seemed to be harder to inhibit an activation of the familiar semantic information (color) than the non-verbal stimuli, so that they made more errors and took longer when performing the verbal inhibition task than non-verbal inhibition task. These results seemed to reflect each EF task used in the current study was well constructed to index independent EF components as suggested by Miyake *et al*. (2000).

Table 2. Descriptive statistics of each executive function task: Means, standard deviations (S.D.), minimal scores, and maximal scores.

Task	Mean	S.D.	Min.	Max.
digit n-back accuracy	0.706	0.143	0.377	0.888
digit n-back RT	1193.186	349.037	600.342	1874.675
visual n-back accuracy	0.603	0.134	0.357	0.829
visual n-back RT	1180.699	259.868	626.485	1661.105
flanker accuracy	0.959	0.079	0.683	1
flanker RT	699.054	129.142	510.305	935.236
stroop accuracy	0.890	0.100	0.675	0.975
stroop RT	1280.90	372.258	818.269	2317.355
trail-making task: number	53.502	25.549	25.33	120.37
trail-making task: letter	75.946	26.948	28.58	126.83
trail-making task: num-let	161.960	106.543	43.33	391.02

3.2.2. Partial correlations: EF tasks and 3AFC

A series of partial correlation tests were conducted between the individual (random-effect) coefficients of VOT and *f0* from the mixed-effect regression model based on 3AFC perception task and individual scores of the tasks assessing the EF capacities. <Table 3> summarizes the outputs of the correlation tests.

Correlation with WM capacity: The correlation between WM capacity (*digit* and *visual n-back* task scores) and perceptual reliance on the acoustic cues was meaningful, although it was not comprehensively so, in that only one test pair of tense-aspirated contrast revealed significant relationship (β_{f0} & β_{VOT} and *visual n-back* accuracy). Looking into the sign of correlation, the positive correlation between β_{VOT} and *visual n-back* accuracy (r=0.52, p <.1) shows that greater dependencies on VOT for a tense-aspirated contrast were associated with better WM. Also, the negative sign of correlation coefficient (r=-582) between β_{f0} and *visual n-back* accuracy indicates that listeners relying on the *f0* cue more than others for a tense-aspirated contrast tended to have lower WM. These relationships are expected given the group-averaged trend in the speech perception that *f0* did not play a significant role in differentiating the aspirated stops from the tense stops (<Table 1>). That is, child listeners who inefficiently utilized the unimportant *f0* cue more than others had lower WM capacities.

Correlation with inhibition control: There were significant correlations between accuracy scores of the *flanker* task and individual VOT and *f0* coefficients across the contrast pairs. It is necessary to mention that we related the accuracy scores to the acoustic coefficients (although it is common in the field to treat the response time as a numeric representation of inhibition control), because we found no evidence of compromise between accuracy and response time -- positive correlations between accuracy and response time in the *flanker* tasks might imply that listeners tried to be accurate although it took time.

For the tense-aspirated contrast pair, greater individual VOT coefficients were associated with higher accuracies indicating that listeners relying on VOT more than others had better inhibition control. For the same contrast pair, individual *f0* coefficients were negatively correlated with the accuracy scores of the *flanker* task, suggesting that listeners who depended on the *f0* cue in the tense-aspirated identification (i.e., an unimportant cue for the perception according to the group trend) tended to have poor inhibition control. This association of poor inhibition control with use of unimportant acoustic cue (i.e., *f0* in tense-aspirated contrast pair) was also found in the *stroop* task accuracies with a marginal statistical significance (r=-0.503). For the tense-lax, and lax-aspirated contrast pairs, VOT and *f0* coefficients were significantly correlated with the *flanker* accuracy scores in a consistent direction where individuals with more strengthened use of the acoustic cue (VOT) were associated with better inhibition control.

Table 3. Summary of the correlation tests. Variables of speech perception are presented in columns and those of executive function are in rows. Bold indicates p<.05, and italic indicates p<.1.

	tense-aspirated		tense-lax		lax-aspirated	
	β_{VOT}	β_{f0}	β_{VOT}	β_{f0}	β_{VOT}	β_{f0}
digit n-back accuracy	0.097	-0.35	0.261	-0.179	-0.288	-0.335
digit n-back RT	0.146	0.219	0.159	-0.237	0.345	0.362
visual n-back accuracy	*0.52*	**-0.582**	0.422	-0.351	0.338	0.307
visual n-back RT	-0.089	0.248	-0.384	0.359	-0.102	-0.098
flanker accuracy	**0.775**	**-0.749**	**0.749**	**-0.714**	**0.595**	*0.556*
flanker RT	0.297	-0.173	0.128	-0.166	*0.512*	*0.506*
stroop accuracy	0.518	*-0.503*	0.377	-0.301	0.076	0.046
stroop RT	0.187	-0.035	-0.157	0.132	0.276	0.281
trail-making task: number	**-0.638**	**0.678**	*-0.588*	*0.543*	-0.274	-0.246
trail-making task: letter	-0.17	0.235	-0.376	0.361	-0.247	-0.258
trail-making task: num-let	**-0.639**	**0.614**	**-0.701**	**0.684**	-0.273	-0.209

Correlation with attention shift control: Consistent with the other EF tasks, the TMT response time of the mixed condition (number-letter) were significantly correlated with beta coefficients of VOT and *f0* for the tense-aspirated and tense-lax contrast pairs. The correlation signs consistently suggest that the enhanced uses of informative cues were associated with shorter duration of TMT. Children with shorter RTs in TMT tended to rely on VOT (higher β_{VOTs}) and *f0* (smaller β_{f0} in Tense-Lax pair) more than others, yielding a negative relationship for the VOT cue and a positive relationship for the *f0*. More importantly, children's use of an uninformative cue (*f0* for a tense-aspirated contrast) was associated with shorter RTs in TMT.

To summarize, the results of partial correlation tests consistently supported that the listeners attending to uninformative cue (less tactful processing strategy) underperformed in the EF tasks and those with greater attention to informative acoustic cues (tactful processing strategy) outperformed in the EF tasks.

4. Discussion

The current study investigated how Korean-speaking children utilize the multiple acoustic cues in perceiving the stop laryngeal contrasts and whether their perceptual attention to the acoustic cues is related to the general cognitive abilities by testing their executive function (EF). The findings showed that the children aged from 7 to 8 depended on the VOT cue consistently across the laryngeal contrast and also employed multiple cues by either ignoring or activating the *f0* cue depending on the nature of the contrast pair. Individual level analysis revealed that selective uses of the acoustic cue (i.e., *f0*) were related to children's EF in a way that suppression of uninformative cue and enhancement of useful cue for the contrasts (e.g., *f0* and VOT for the tense-aspirated contrast, respectively) were associated with better cognitive controls. The current findings converge to support that listeners' general cognitive ability is responsible for listeners' speech processing patterns.

While cautious interpretations of the observations should be made due to a relatively small sample size (N=15) of the present study, the results revealed consistent tendencies of how children (of this age range) processed sub-phonemic information present in the speech signals to access the phonetic categories. The child participants' effective uses of multiple acoustic cues reflected the enhanced role of *f0* in the context of a sound change in the Korean stops (Silva, 2006; Kang, 2014). In accommodating phonetic characteristics of the three laryngeal stops, children selectively employed the *f0* cue as acoustic information, while the role of VOT was effective across the contrastive pairs. Specifically, the active use of *f0* was noticeable in perceiving the lax type (e.g., lax-tense and lax-aspirated stops). The role of *f0* was even greater than that of VOT in differentiating the lax from the aspirated stops, of which VOT values were known to overlap according to the production studies (Kim, 2004; Kong *et al*., 2011). By contrast, when the *f0* characteristics between tense and aspirated stops were not distinct, the children tended to ignore the *f0* cue. The tendency in children's speech perception confirms that the changed acoustic nature of the Korean laryngeal stops have widely spread across generations in the speech community.

The study revealed age-related characteristics of speech perception, which were not quite adult-like based on prior research. For one, it was observed that children used the two cues (VOT and *f0*) in a highly correlated manner across individuals, indicating that internal organizations of the multiple cues were different between child and adult listeners. For child listeners, the use of *f0* systematically covaried with that of VOT, even when *f0* was not an useful cue for a category distinction: For a tense-aspirated pair, greater perceptual reliance on VOT was associated with less use of *f0*. This differs from an inconsistent relationship in adults' perception of tense-aspirated stop contrast observed in Kong & Kang (2017). Compared to adults, children showed a poor control of activating one cue independently of other cues in processing the laryngeal stops.

Another novel finding in the current research is that individual children's perceptual patterns tended to be related to their performance of the EF tasks. Importantly, the nature of relationship was insightful in that individuals with better cognitive ability processed available acoustic cues more tactfully than others in identifying the phonetic categories of the Korean stops. That is, Korean child listeners who could suppress less informative, thus *not* primary acoustic cue (e.g., *f0* for tense-aspirated contrast) turned out to have better controls of working memory, inhibition capacity and attention shift. Similarly, child listeners who paid attention to VOT across the stop contrast pairs excelled in the EF tasks. These results support that child listeners' cue-weighting strategies in speech perception are modulated by their general cognitive function in order to make selective use of multiple available information.

With respect to the association of better cognitive ability with less use of *f0* for the tense-aspirated in the current research, it is likely that this relationship reflects developmental characteristics of how and what child listeners at this age range can do with multiple information given that their cognitive abilities are still developing. Contrary to our findings with child listeners, Kong & Edwards (2016) demonstrated that English adult listeners with better EF controls were able to attend to redundant acoustic cues more than others in processing the phonetic categories. Their findings suggest that adult listeners who can hold both primary and redundant information for the task of speech processing are considered efficient listeners. It appears that efficiency of speech processing needs to be assessed differently between cognitively fully developed adults and cognitively still developing children, because of dynamic interaction between language and cognition.

One might be skeptical of the current interpretation of the relationship between EF and speech perception since the correlations were not found across the board in the entire sets of EF tasks we administered. Indeed, *digit n-back* (a measure of WM), and *stroop* (a measure of inhibition) were either *not* or *only minimally* correlated with any uses of acoustic parameters for the three contrast pairs. While understanding one's conservative position seeking for more robust demonstration of the relationship, we note that the complex nature of EF tasks makes it unrealistic to obtain significant correlations across the board, since each type of EF tasks stimulates different aspects of participants' cognitive processing and carries different degrees of performance difficulties. The current experimental evidence is still valued as an inspiring preliminary attempt to consider the cognitive ability as a source of individual variations in speech perception, which has not been investigated over Korean-speaking populations.

To conclude, the current study showed that, for the child listeners aged from 7 to 8, there was a meaningful relationship between speech processing strategy and general cognitive ability in a way that the inefficient listeners who are poor at managing the acoustic information for phonetic categories had poor cognitive abilities related to executive function. Further studies on this relationship needs to follow up in order to better understand the interplay of cognitive ability and language acquisition in the course of life-time, in general.

References

Alloway, P. (2007). Working memory, reading and mathematical skills in children with developmental coordination disorder. *Journal of Experimental Child Psychology*, 96, 20-36.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829-839.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 47-90). New York: Academic Press.

Barkley, A. (2012). *Executive functions: What they are, how they work, and why they evolved*. New York: Guilford Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by

computer [Computer program]. Version 5.3.51. Retrieved from http://www.praat.org/ on August 15, 2017.
Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34(3), 372-387.
Cedrus Corporation. (2015). SuperLab. Version 5.0.5. San Pedro, CA. [Computer program]
Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
Diamond, A. (2013). Executive functions. *Annual review of psychology*, 64, 135-168.
Eriksen, B., & Eriksen, C. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143-149.
Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26, 551-585.
Flege, J., Bohn, O., & 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., 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.
Friedman, N., Miyake, A., Corley, R., Young, S., DeFries, J., & Hewitt, J. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17(2), 172-179.
Gordon, P., Eberhardt, J., & Rueckl, J. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, 25, 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., & Lotto, A. (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.
Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica*, 136, 189-202.
Kang, Y. (2014). Voiceonset time merger and development of tonal contrast in Seoul Korean stop: Acorpus study. *Journal of Phonetics*, 45, 76-90.
Kim, M. (2004). Correlation betweenVOT and F0 in the perception of Korean stops and affricates. *Proceedings of the 8th International Conference on Spoken Language Processing* (pp. 49-52).
Kim, S. (2015). ppcor: partial and semi-partial (part) Correlation. R package version 1.1. Retrieved from https://CRAN.R-project.org/package=ppcor on August 15, 2017.
Kim, Y., Hong, G., Kim, K., Jang, H., & Lee, J. (2009). *Receptive & expressive vocabulary test (REVT)*. Seoul: Seoul Community Rehabilitation Center.
Kingston, J., & Diehl, R. (1994). Phonetic knowledge. *Language*, 419-454.
Kong, E., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, 59, 40-57.
Kong, E., & Kang, J. (2015). L2 proficiency and effect of auditory source in processing L2 stops. *Phonetics and Speech Sciences*, 7(3), 99-105.
Kong, E., & Kang, S. (2017). Individual variability in processing English source input into Korean stops. *Studies in Phonetics, Phonology and Morphology*, 23(2), 201-219.
Kong, E., & Yoon, I. (2013). L2 proficiency effect on the acoustic cue-weighting pattern by Korean L2 learners of English: production and perception of English stops. *Phonetics and Speech Sciences*, 5(4), 81-90.
Kong, E., Beckman, M., & Edwards, J. (2011). Why are Korean tense stops acquired so early?: The role of acoustic properties. *Journal of Phonetics*, 39(2), 196-211.
Kopp, B., Mattler, U., & Rist, F. (1994). Selective attention and response competition in schizophrenic patients. *Psychiatry Research*, 53(2), 129-139.
Kronenberger, W., Pisoni, D., Henning, S., & Colson, B. (2013). Executive functioning skills in long-term users of cochlear implants: A case control study. *Journal of Pediatric Psychology*, 38, 902-914.
Lev-Ari, S., & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics*, 41, 320-331.
Lezak, M. (1983). *Neuropsychological Assessment.* New York: Oxford University Press.
Li, F. (2012). Language specific developmental differences in speech production: A cross-language acoustic study. *Child Development*, 83(4), 1303-1315.
Lisker, L. (1978). Rapid vs. rabid: A catalogue of acoustic features that may cue the distinction. *Haskins Laboratories Status Report on Speech Research*, 54, 127-132.
Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384-422.
Mayo, C., & Turk, A. (2004). Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *The Journal of the Acoustical Society of America*, 115(6), 3184-3194.
Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, 13(3), 279-297.
Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A., & Wager, T. (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.
Moon, S., & Byun, C. (1997). *Korean Kaufman assessment battery for children (K-ABC).* Seoul: Hakjisa.
Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20(3), 351-382.
Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *The Journal of the Acoustical Society of America*, 112(2), 711-719.
Nittrouer, S., & Burton, L. (2005). The role of early language experience in the development of speech perception and phonological processing abilities: Evidence from 5-year-olds with histories of otitis media with effusion and low socioeconomic status. *Journal of Communication Disorders*, 38, 29-63.
Nittrouer, S., & Miller, M. (1997). Developmental weighting shifts for noise components of fricative-vowel syllables. *The Journal of the Acoustical Society of America*, 102(1), 572-580.
Nittrouer, S., Caldwell-Tarr, A., & Lowenstein, J. (2013). Working memory in children with cochlear implants: problems are in

storage, not processing. *International Journal of Pediatric Otorhinolaryngology*, 77, 1886-1898.
Owen, A., McMillan, K., Laird, A., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, 25(1), 46-59.
Psychology Software Tools, Incorporated. [E-Prime 2.0]. (2012). Retrieved from http://www.pstnet.com on August 15, 2017.
R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/ on August 15, 2017.
Repp, B. (1983). Coarticulation in sequences of two nonhomorganic stop consonants: perceptual and acoustic evidence. *The Journal of the Acoustical Society of America*, 74(2), 420-427.
Riki, T., Minka, H., & Liat, K. (2007). The effect of age on acoustic cue weighting in the perception of initial stop voicing contrast in Hebrew. *Proceedings of the 16th International Congress of Phonetic Sciences*.
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.
Silva, D. (2006). Acoustic evidence for the emergence of tonal contrast in contemporaryKorean. *Phonology*, 23, 287-308.
Stuss, D., Levine, B., Alexander, M., Hong, J., Palumbo, C., Hamer, L., Murphy, K., & Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, 38, 388-402.
Summerfield, Q., & Haggard, M. (1974). Perceptual processing of multiple cues and contexts: effects of following vowel upon stop consonant voicing. *Journal of Phonetics*, 2, 279-295.
van Maanen, L., van Rijn, H., & Borst, J. (2009). Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin & Review*, 16(6), 987-999.
Vanderplas, J., & Garvin, E. (1959). Complexity, association value, and practice as factors in shape recognition following paired-associates training. *Journal of experimental psychology*, 57(3), 147-154.
Walley, A., & Carrell, T. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 73, 1011-1022.
Yamada, R., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Attention, Perception, & Psychophysics*, 52(4), 376-392.
Zelazo, P., Mueller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, 68, 1-137.

• **Eun Jong Kong**
Department of English
Korea Aerospace University
76 Hanggongdaehang-ro, Deogyang-gu, Goyang–city
Gyeonggi-do 412-791, Korea
Tel: 02-300-0338
Email: ekong@kau.ac.kr
Fields of interest: Acoustic phonetics, L1 & L2 phonological acquisition, speech perception

• **Jeewon Yoo** corresponding author
Department of Communication Disorders
Ewha Womans University
52 Ewhayeodae-gil, Seodaemun-gu,
Seoul 120-750, Korea
Tel: 02-3277-2120
Email: jeewonyoo@ewhain.net
Fields of interest: Executive Function, language processing in L1 & L2