Attentional modulation and individual differences in explaining the changing role of f0 in the Korean laryngeal stop perception

Eun Jong Kong¹, Hyunjung Lee²
Korea Aerospace University¹, Kyungnam University²
ekong@kau.ac.kr, hyunjungeelee123@gmail.com

[Background] While the enhanced role of the f0 cue in distinguishing a three-way laryngeal contrast in the Korean stops (i.e., tense, lax and aspirated) has been a great interest for linguists due to the on-going sound change, there has not been strong consensus on whether the f0 cue is perceptually primary over VOT in the category distinction. The present study aimed to explore how the Korean listeners’ utilization of f0 (and VOT) interacts with attention resources in order to define the relative importance of the two acoustic cues. Attention given to speech sounds is known as a factor modulating the use of multiple acoustic cues in speech perception (e.g., Gordon et al., 1993, Francis et al., 2008). Gordon et al. (1993) conducted a perception experiment where the amount of attention allocated to speech perception was intentionally manipulated by introducing a distracting condition (arithmetic task) to the speech recognition task. Findings were that when the attention to the speech sounds was limited (i.e., in a distracted situation), listeners’ reliance on the primary cue is mitigated but that of the secondary cue is boosted. Given that an attention demand is associated with a primacy of the cue in the speech perception, we explore how each of the acoustic cues (f0 and VOT) is affected by the manipulation of attentional conditions in processing Korean stops. That is, if Korean listeners’ reliance on f0 decreases under the distracting listening condition, we could define f0 as a primary cue.

[Methods] The 3-alternative forced choice identification task (/tʰ/ / t/) was conducted over 28 adults speaking Seoul Korean aged from 20 to 27. A total of 35 audio stimuli (7-step VOT × 5-step f0) were created based on /ta/ (lax) and /tʰa/ (aspirated) base tokens produced by a 30-year-old male Korean speaker. The 3AFC identification task was blocked by the attentional condition (distractor vs. no-distractor). In the distractor session the listeners were asked to solve arithmetic questions while deciding the stop categories after the audio stimuli, whereas only the stop identification task was given in the no-distractor session. Each listener responded a total of 210 trials (35 audio stimuli × 2 distractor conditions × 4 repetitions). The mixed effects regression models assessed likelihood of listeners’ choosing one stop category over the other, yielding three different binomial logistic models: (1) tense-aspirated, (2) tense-lax, and (3) lax-aspirated responses. Fixed effects included two continuous variables of VOTs and f0s and a discrete variable of the distractor condition (distractor vs. no-distractor). The model formula in lme4 style in R was \( \text{stop_response} \sim (\text{VOT+f0}) \times \text{factor(attention)} + (\text{VOT+f0}|\text{subject}) \).

[Results] In the tense-aspirated model (Fig.1-leftmost), the rate of /tʰ/ (aspirated) responses increased with longer VOT (β = 6.75, p < 0.001), but the change in either f0 or the attentional condition did not affect the stop choice. A significant interaction between VOT and attention (β_{VOT,distractor} = 0.77, p < 0.01) was observed suggesting that the listeners’ reliance on VOT was weakened in the distractor condition of doing arithmetic task. However, there was no such attentional modulation with f0 in the perception of the tense-aspirated stop. In the tense-lax model (Fig.1-middle), the rate of /t/ (lax) responses increased with longer VOT (β = 1.47, p < 0.01) and decreased with higher f0 (β = -1.79, p < 0.001). Consistent with the tense-aspirated identification, the significant interaction with the attentional condition was found for VOT (β = 0.57, p = 0.03), but not for f0. In the aspirated-lax model (Fig.1-rightmost), the rate of /tʰ/ (aspirated) responses increases with longer VOT (β = 3.02, p < 0.001) and higher f0 (β = 1.62, p < 0.001). However, unlike the other two models, the attention condition did not interact with either VOT or f0: \( \beta_{\text{VOT,distractor}} = 0.32, \ p = 0.08; \ \beta_{\text{f0,distractor}} = 0.04, \ p = 0.79 \). While this model output might reflect a true lack of interaction between attention and the two acoustic cues, one might also suspect a situation where some listeners weighed f0 over VOT and others weighed VOT over f0, which might cancel out the interaction effect of attentional modulation in the model. To further explore the possibility, we selectively grouped the
listeners based on their relative cue-weighing patterns, separating them into $f0$-dependent (10 top listeners with greater $f0$ coefficients) and VOT-dependent groups (10 top listeners with greater VOT coefficients) (Fig. 2). When the regression models were made for each subgroup of listeners, the effect of the distractor on VOT was significant ($\beta_{\text{VOT:attention}} = 0.92, p < 0.05$), but the effect on $f0$ was not significant even for those listeners who relied on $f0$ more than VOT ($\beta_{\text{f0:attention}} = 0.37, p = 1.627$).

[Discussion] Our findings showed that the presence of the distractor did not affect the Korean listeners’ use of $f0$ in identifying the three-way laryngeal stop categories, indicating that the $f0$ cue for the Korean listeners does not operate with high level of attention. This result was borne out even when the effect of the distractor was examined over the confined sub-group of $f0$-dependent listeners for a lax and aspirated stop contrast. Considering that the VOT cue, which functions with relatively higher level of attention for the English listeners, is a primary acoustic cue in processing the English stops (Gordon et al., 1993; Whalen et al., 1993), the current results do not encourage us to argue that $f0$ is a solely primary cue in processing the Korean stops in the context of the sound change. Rather, it was the VOT cue that was consistently modulated with attentional resource, suggesting that it plays a primary role in cueing a Korean stop laryngeal contrast. A more accurate and precise picture of the sound change in the Korean stops needs to reflect how the speech community perceives as well as produces the sounds. The $f0$ as the primary cue in the Korean stops should be re-considered.

Figure 1. The estimated probability of stop responses for the pairs of tense-aspirated (leftmost), tense-lax (middle) and lax-aspirated (lax-aspirated) with respect to VOT and $f0$ between distractor and no-distractor conditions.

Figure 2. The estimated probability of aspirated responses over lax for VOT and $f0$ reliance groups as a function of VOT and $f0$ between distractor and no-distractor conditions.

[Reference]
