Cue Integration in the Perception of Korean Fricatives

Sarah Lee

Undergraduate Linguistics Honors Thesis

April 22, 2013

Abstract

Recent literature on Korean fricatives has shown that the perception of the plain /s/ over the fortis /s*/ correlates to low f0 onset of the following vowel, despite the fact that there is no such correlation in the production of /s/ over /s*/. The integration of f0 with other low frequency cues that covary more reliably with /s/ presents a possible explanation for this disparity between production and perception. In a fixed classification experiment, native Korean speakers distinguished between stimuli that differed in intervocalic voicing and f0 onset following /s/. Results showed the integration of these two cues into a single low-frequency intermediate perceptual property.

1. Introduction

1.1 Goals of this paper

The distinction of a contrast in a single phonological feature often involves several correlates in the acoustic signal that contribute to a particular percept. The laryngeal contrast in Korean fricatives is especially interesting because of the difficulty in classifying the two kinds of /s/ and understanding how they are perceptually distinguished. In studying the numerous acoustic properties that influence the perception of this contrast, an interesting fact has been repeated in the literature regarding the f0 onset of the vowel following a fricative. Although there does not seem to be any significant correlation between f0 onset and the production of either /s/ or /s*/, f0 onset has been shown to contribute to the perception of one or the other (Chang 2013). An intriguing explanation for this mismatch stems from the possibility that multiple cues interact with one another, physically similar cues in particular, to lend themselves to a single percept. Kingston et al (2008) suggests that rather than paying attention to every individual property in the acoustic signal, the auditory system might perceive acoustically related cues as a single "intermediate perceptual property" (see section 1.3). Specifically, Kingston et al (2008) found evidence for the perceptual integration of multiple properties associated with low frequency spectral continuity into a single "low frequency" property.

The concept of cue integration offers a way to think about the surprising correlation between low f0 onset and perceiving the plain /s/, suggesting a potential integration effect between f0 onset and some acoustically similar property that is associated with the plain /s/ in both production and perception. Such a physically similar property could be the presence of voicing, which has been found to be associated with the plain /s/, but not /s*/, in the intervocalic context. The two low frequency properties of low f0 onset and voicing could affect the likelihood of perceiving the plain /s/ in an integrated fashion. In fact, if a "low frequency" intermediate perceptual property is detectable by the auditory system in a low-level psychoacoustic fashion, as suggested by Kingston et al (2008), the integration of low f0 onset and voicing would be expected cross-linguistically. The goal of this study is to determine the perceptual relationship between these two low frequency acoustic cues, voicing and low f0, in Korean and to consider what that relationship would mean for the perception of the plain versus fortis fricative contrast. Furthermore, the results from this study should add to the discussion of whether or not the integration of these acoustic cues is indeed a property of general auditory perception as opposed to language-specific perception.

The perception experiment presented in this paper provides evidence for perceptual integration of intervocalic voicing and low f0 onset within the category of the plain /s/, suggesting an explanation for the unexpected effect of f0 onset on the perception of the fortis vs. plain contrast. The relationship between these two cues also offers further evidence for the notion that cue integration is a property of the auditory system.

1.2 Previous literature on Korean fricatives

The laryngeal contrast in Korean obstruents has been studied extensively from both an acoustic and a perceptual standpoint. Korean is especially recognized for its typologically unusual three-way laryngeal contrast in stops and affricates, generally categorized as lenis, fortis, and aspirated, which are all voiceless in word-initial position. Lenis stops are characterized by a breathy quality, slight aspiration, and intervocalic voicing, fortis stops are tense and notably unaspirated, and aspirated stops are heavily aspirated (Cho et al. 2002). The literature has typically focused more on stops because of this typologically unusual property, but Korean also differentiates between two forms of the denti-alveolar fricative, s, which are not as easily classified. One has been identified as fortis, typically notated as /s*/, corresponding to the characteristics of the fortis stop category, including lack of aspiration and a constricted glottal opening (Cho et al. 2002). However, the other, referred to henceforth as the plain /s/, shares characteristics with both the lenis and aspirated stops and thus has been more controversial in the literature.

Though the plain /s/ displays a greater degree of aspiration than lenis stops, it is treated phonologically as lenis. The plain /s/ undergoes the post-obstruent tensing that affects lenis stops, as in the word 박사 /paksa/ for "Ph.D" being realized as [paks*a]. The plain /s/ is realized as the fortis, or tense, /s*/ after the lenis stop /k/ (Cho et al. 2002). Unlike the lenis stops, the plain /s/ was originally thought not to undergo intervocalic voicing associated with lenis stops and still described as such in recent literature (Chang 2013). However, Cho (2002) found that in addition to being unaspirated word-medially, the plain /s/ can undergo partial and even full voicing, with 46% of tokens of intervocalic /s/ being fully voiced in their production experiment.

In the literature, the plain /s/ has been analyzed as aspirated (Yoon 2002), lenis (Cho et al. 2002), or a hybrid category distinct from those used to classify the stops (Chang 2013). However, in each of these cases, the analyses are primarily based on articulatory grounds, particularly with regard to the configuration of the glottis, and do not shed much light on how the plain /s/ is perceptually distinguished from the fortis /s*/.

Perception experiments have not yielded conclusive results as to how Korean speakers are able to differentiate between the plain /s/ and the fortis /s*/, particularly due to the variability of many acoustic properties by vowel context. Yoon (2002) did not find any feature associated with the oral tract that lent itself to distinguishing between the two. Chang (2013) similarly ascertained in his perception experiment that the most significant parameter used in labeling stimuli was whether or not the vowel following the fricative had originally followed /s/ or /s*/. Chang also noted that the quality of the following vowel also had an effect on perception, where high vowels were more likely to be labeled as plain.

Interestingly, in the same experiment, Chang (2013) found that f0 had some perceptual effect on this contrast – lower f0 onsets made fortis identification less likely. These results accord with previous studies demonstrating an effect of f0 on the perception of Korean fricatives (Holliday 2010). However, production experiments did not show any acoustic correlation between f0 onset and producing /s/ versus /s*/ overall nor in any vowel context (Chang 2013). In comparing results from speakers of the Cheju and Seoul dialects in a production experiment, Cho (2002) found that in the Cheju dialect, vowels following /s*/ have a significantly higher f0 onset, but that there is no such distinction in the Seoul dialect. The fact that speakers of the Seoul dialect are not reliably producing any difference between the f0 onset after /s/ and /s*/ and yet are listening for a difference in perceiving one or the other is very surprising. One possible explanation for this phenomenon is the perceptual integration of multiple acoustic cues, of which f0 onset is one. It may be that another cue that is in fact reliably produced with /s/ and not /s*/ contributes to the perception of this contrast and integrates with f0 onset. Thus, it is not that speakers are paying attention specifically to f0, which would be irrelevant in the case of /s/ and /s*/, but rather that f0 onset integrates with another cue and the combination of the two contributes to the likelihood of perceiving /s/.

1.3 Perceptual integration – Kingston et al (2008)

The unexpected correlation between low f0 and perceiving /s/ over /s*/ coincides with the previously denied but now substantiated possibility of intervocalic voicing of /s/. Together, these two facts suggest that the detection of low frequency properties is used in the perception of /s/. Specifically, the integration of low f0 onset and intervocalic voicing may allow for low f0 to influence the perception of voicing, thereby enhancing the perception of the plain /s/ over the fortis /s*/. In such a case, the acoustically similar properties of voicing and low f0 might form an "intermediate perceptual property," or IPP, through which these two properties associated with low frequency are integrated perceptually (Kingston et al. 2008).

Kingston et al (2008) presents evidence for a low-frequency property in which voicing integrates with both low f0 and low F1 in the perception of the voice contrast in English stops. In a fixed classification task, as presented in this paper as well, native English speakers discriminated between stimuli differing by f0, F1, and presence of voicing in a way that was not true for f0 with closure duration, although all of these properties covary with the voice contrast. Low f0-voiced stimuli were easily discriminated from high f0-voiceless stimuli, while low f0-voiceless stimuli were difficult to distinguish from high f0-voiced stimuli, although these stimulus pairs differed by the same amount in each of the two dimensions. This suggests that the sum of these two cues influenced perception, rather than each cue independently influencing perception. The same results were found in the combination of F1 and voicing, but in varying f0 and closure duration, the two parameters seemed to affect perception independent of one another. Also, f0 and F1 did not seem to integrate perceptually, suggesting that the low frequency IPP relates to spectral continuity in low frequencies, rather than low frequency energy. Strikingly, these effects were found to be true not only for English stops but also in non-speech analogues (Kingston et al. 2008). The argument that cue integration is belongs to general, low level auditory perception suggests that the same integration would take place cross-linguistically.

2. Methods

2.0 The Garner paradigm

The Garner paradigm, as adapted by Kingston et al (2008), provides an approach to determining the perceptual independence or dependence of two acoustic cues. A 2x2 stimulus array varying along the dimensions of voicing and f0 onset leads to the construction of four stimuli. Presence of voicing and low f0 onset are associated with the hypothesized integrative low-frequency property, so each of the four stimuli can be defined by whether its value for these two dimensions add to the low-frequency percept (+) or subtract from it (-), as shown in Figure 2-1 below.

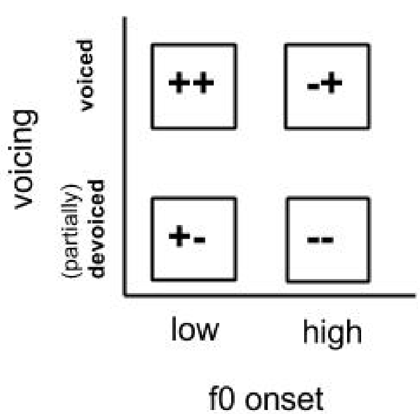

Figure 2-1 A 2x2 stimulus array showing the physical or acoustic distances between stimuli with respect to the dimensions of voicing and f0 onset, after Kingston et al (2008).

Ideally, each stimulus would be separated by a just noticeable difference in each dimension, but the important point is simply to have a difference allowing for subjects to perform a classification task above chance but below ceiling. The perceptual distance between the + and – values chosen for each dimension need not be identical for the integration effect to be observable (meaning the diagram above does not need to be an exact square; it could be a rectangle). Furthermore, the two properties would only be modified within the boundaries of natural variability found in the /s/ category in the Seoul dialect of Korean (note that voicing has not been reported in the /s*/ category). The task is not to label the fricative in the stimuli as belonging to one phonemic category or the other, so phonological knowledge does not play a role in this experiment.

In a fixed classification task, subjects would be presented with one of two stimuli and would be asked to classify them. A single-dimension task would only vary in one of the two dimensions, presenting for example the ++/+- (LV/LD) pair, which differs only in voicing, in order to investigate the perceptual distance along one side of the square. A two-dimensional task would vary in both dimensions, investigating the perceptual distance between diagonal ends of the square, as in the +-/-+ (LD/HV) pair. Perceptually independent cues would yield two diagonals of equal length, while perceptually dependent cues would yield diagonals of differing length - where the longer diagonal would correspond to the ++/-- (LV/HD) pair. This would be expected because the reinforcement of both cues adding to the same intermediate perceptual property in each stimulus would allow for much clearer perceptual distinction between the two.

2.1 Stimuli

As much as possible, we attempted to adhere to the experimental design presented in Kingston et al (2008). According to the hypothesis that a low frequency property is involved in the perception of the plain /s/ in Korean, f0 should integrate perceptually with voicing. In order to test this, all six comparisons of the Garner paradigm were tested in a fixed classification experiment.

A native speaker of Seoul Korean was asked to record several tokens of words with an intervocalic, plain /s/ in a sentence-medial context. From these tokens, we chose one that could be easily manipulated in terms of f0 and voicing. For this speaker, most tokens included a very short voice bar at the beginning of each intervocalic, plain /s/. However, in one particular word, 기숙사 /kisuksa/ (pronounced [kisuks*a] by the post-obstruent tensing rule discussed in the introduction), meaning 'dormitory,' the /s/ was consistently fully voiced. Also, this female speaker had an average value of f0 onset following /s/ of roughly 260 Hz, but ranging from 160 Hz to 340 Hz. A token of the aforementioned word [kisuks*a] featured both the presence of voicing in the plain /s/ of the second syllable and an intermediate f0 value of 250 Hz in the onset of the following vowel, facilitating the creation of multiple stimuli from this token.

This token was isolated and manipulated in Praat to construct the four stimuli compared in the Garner paradigm: low f0-fully voiced (LV), high f0-fully voiced (HV), low f0-partially devoiced (LD), and high f0-partially devoiced (HD). In creating these stimuli, the hope was to reach a just noticeable difference in each dimension, as indicated by Kingston et al (2008), but this was not rigorously pursued. As mentioned above, the goal was to achieve a perceptual distance recognized above chance but below ceiling.

First, the token had to be excised from running speech. It appeared in the context of the sentence: 친구가 <u>기숙사</u>로 갔다 [tʃinguga <u>kisuks*a</u>ro katd*a] "My friend went to the <u>dormitory.</u>" The target word was easily distinguished from the preceding vowel due to silence before initial stop [k]. The liquid of the following particle [ro] slightly obscured the right edge of the word, but the boundary was selected before the formant structure shows the transition to the liquid. Both edges were selected at zero crossings and the resulting fragment of 0.424 s was taken as the original token from which stimuli were made.

This token was 100% voiced and left intact in the fully voiced stimuli, while 75% of the voicing was removed from the right edge of the fricative inward for the partially devoiced stimuli. It was found that many of the recorded tokens of this and other words featured some voicing in the beginning of the intervocalic plain /s/ fricative, so the left edge of voicing was preserved while subsequent voicing was removed. To do so, a pass band filter was used to remove all energy under 1000 Hz in Praat from one quarter through the duration of the fricative to the end of the frication.

Taking the fully voiced and newly made partially devoiced tokens, the f0 onset dimension was then manipulated: the f0 onset in the original token (251 Hz) was raised or lowered by 30 Hz to make the high f0 (281 Hz) and low f0 (221 Hz) stimuli, respectively, for a difference of 60 Hz. This was done in Praat by creating a manipulation object with a time step of .005 s and altering the pitch tier. We either raised or lowered the first pitch point in the vowel by 30 Hz and moved the three subsequent pitch points in such a way as to allow for a smooth transition into the rest of the vowel within 20 milliseconds (taking up about 45% of the duration of the vowel). The manipulation object was then resynthesized with the new pitch tier by the resynthesis (overlap-add) function in Praat.

2.2 Subjects

Fourteen adult native Korean speakers participated in the experiment, seven male and seven female. These subjects reported no speech or hearing disorders. Their ages ranged from 18 to 35, with the average age being 22.3. They were recruited from the community at and around the University of California, Berkeley and thus all spoke English as a second language. They all spent the majority of their childhood in South Korea, and all but one spoke the Seoul dialect of Korean (one subject's reported hometown was Ulsan, where the Gyeongsang dialect is found). The speaker who produced the tokens used for making stimuli did not participate in the experiment. All subjects were paid for their participation.

Of the fourteen subjects who participated in the experiment, one subject's data was excluded because of an extreme irregularity in the rate of 'l' responses (this subject hit 'l' in all but two trials for two full blocks).

2.3 Procedures

The experiment was run in a quiet room and no more than two subjects were run at a time. Stimuli were output from a PC and presented to subjects through AKG K240 semi-open studio headphones. Responses were recorded on a standard computer keyboard. The experiment was designed in E-Prime 2 and was run in E-Run 2.

Each subject performed six blocks of a fixed classification task. The pairs of stimuli in each block correspond to the six comparisons in the Garner paradigm: four single-dimension tasks (i.e. the LVHV pair only tests along the f0 dimension) and two "diagonal" or two-dimensional tasks (i.e. the LDHV pair tests along both the f0 and voicing dimensions). Each block consisted of 20 randomized training trials with feedback and 60 randomized test trials with feedback (30 responses per stimulus per block), without a gap in between. In each trial, the subject had an unlimited time to respond, after which a 2s feedback screen would appear. The blocks were presented in random order, and in between blocks, subjects were given the option to take a break and press a button when they were ready to continue.

Subjects were instructed through onscreen directions to listen to one of two stimuli, for each of six blocks, and identify the stimulus as either “A” or “B” by pressing one of two buttons (‘a’ on the keyboard corresponding to “A” and ‘l’ corresponding to “B”). These directions were also explicitly stated by the experimenter. Additionally, subjects were told that they would learn which stimulus was which through feedback in the training trials at the beginning of each block, and that the distinction between stimuli would change after each block. Lastly, subjects were informed that the stimuli would be difficult to tell apart.

2.3 Statistical analysis

In analyzing the results, rather than the calculation of d’ values to measure sensitivity as performed by Kingston et al (2008), a logit mixed effects model was generated with the lme4 package for R. Logit models express how the likelihood, in the form of log odds ratios, of a binary response varies according to the stimuli presented, e.g. the likelihood of responding ‘l’ to the LDHD stimulus pair. Mixed models takes into account random variables, or variables with levels sampled from a population of possible levels (Jaeger 2008). In this case, it allows us to generalize across subjects while modelling the random effects of variability by subject in bias and sensitivity to the different blocks, as well as to the LVHD “diagonal” stimulus pair as compared to the LDHV “diagonal.” The model presents fixed effects for sensitivity to the f0 contrast, the devoicing contrast, and the two “diagonal” stimulus pair conditions, as well as the effect of block number on sensitivities, and the effects of responding after being incorrect in the previous trial or having a different stimulus from the previous trial. It should be noted that rather than comparing the sensitivities of each of the six conditions, we collapsed the two f0-only contrast stimulus pairs into one, and did the same to the voicing contrast, after having found no significant difference in the sensitivities between the two sides of the same contrast.

3. Results

3.1 Description of Data

The data collected from the experiment were the responses subjects gave to stimuli in each trial of each block. We examined the sensitivity of subjects to the differences in the pairs of stimuli by taking into account both correct responses (hits) and incorrect responses (false alarms). When a subject responds 'l' to a stimulus, the log odds that the response was a false alarm are subtracted from the log odds that the response was a hit (i.e. logit('l' stimulus | 'l' response) – logit('a' stimulus | 'l' response)). This data is shown in Figure 3-1, with error bars of 95% confidence intervals.

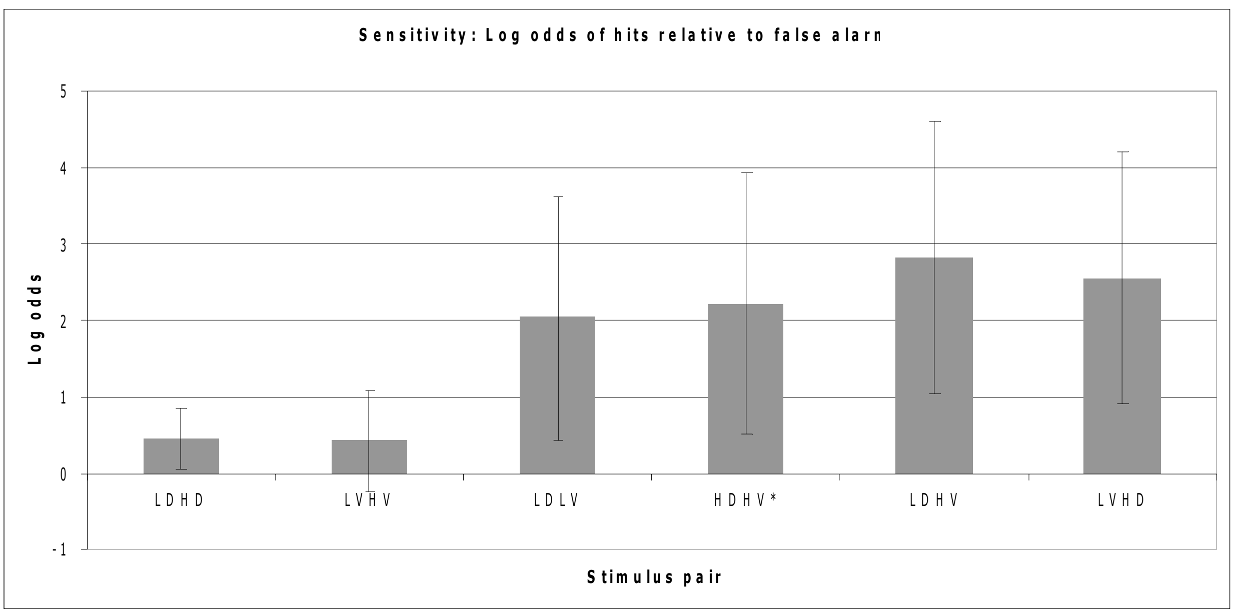

Figure 3-1 The sensitivity as the log odds of hits relative to false alarms, by the stimulus pair presented. *One subject responded correctly to all the test trials of the HDHV block. Because the odds are undefined (60/0), this value was changed to 119/1 (by convention, such values are adjusted by 1/2n, where n is the number of trials) in the calculation of the log odds presented above as well as in subsequent graphs.

The general pattern observed here is that the sensitivity to the differences between stimuli was lowest in the f0-only contrast (LDHD and LVHV), significantly higher in the voicing only contrast (LDLV and HDHV) and highest when both of the parameters varied (LDHV and LVHD). It is also important to note that the odds are roughly the same between the two pairs of stimuli that differed only in f0 and the two pairs of stimuli that differed only in voicing, which is expected because the parameters were altered by the same amount in each case. We do observe a small difference in sensitivity in the two-dimensional pairs of stimuli, where the odds appear higher in the LDHV condition than in the LVHD condition. With respect to the hypothesized IPP, the LVHD pair would positively correlate, mutually reinforcing a "low frequency" property in one direction or the other. On the other hand, the LDHV condition shows a negative correlation with respect to this IPP. Although this difference in sensitivity is slight, the trend that subjects showed greater sensitivity to the LDHV condition runs contrary to the initial hypothesis that low f0 and voicing would perceptually integrate, allowing for the greatest perceptual distinction between the LV and HD stimuli.

3.2 Confound in the results

Although the data above seems to contradict the hypothesis that low f0 and voicing integrate to form an intermediate perceptual percept, upon closer inspection of the results, we found that the pairs of stimuli were not distributed evenly throughout the blocks across subjects. We randomized the order of the blocks in which the different pairs of stimuli would be presented, but with a relatively small number of subjects, certain stimulus pairs occurred in some blocks more often than others. In particular, the stimulus pairs were the least evenly distributed are LVHD, which appeared most frequently in blocks 4 and 5, and LVHV, which appeared most frequently in blocks 1 and 2. This is significant, as we found that block number had a great effect on how subjects did on a given block. Figure 3-2 shows the accuracy of subjects as a function of block number, demonstrating an increase and decrease in overall accuracy due to acclimation and fatigue.

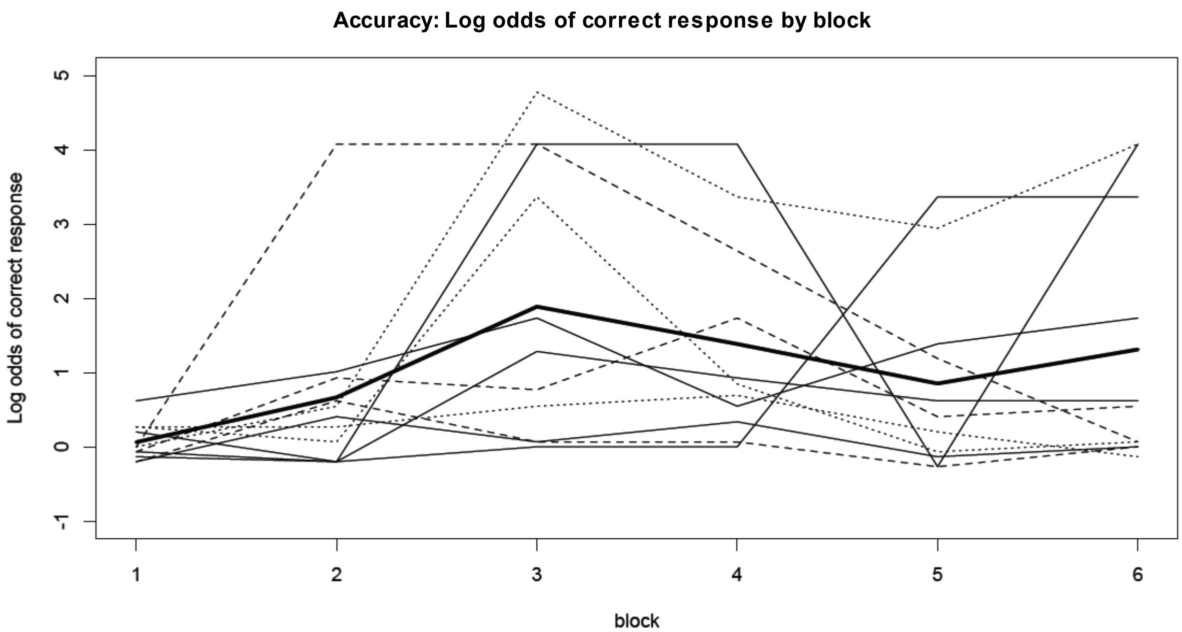

Figure 3-2 The accuracy of subjects by block number. Each thin line represents a single subject, and the thick line represents the average of all subjects.

Despite considerable variability in terms of the exact timing of these effects, the general trend is that subject accuracy, as measured by the log odds of a correct response, increased moderately in block 2 and then more significantly in block 3 (suggesting acclimation), but dropped off in blocks 4 and 5 (fatigue), with a slight increase again in block 6 (possible 'second wind'). The decrease in accuracy in blocks 4 and 5 corresponds to the blocks in which the LVHD stimulus pair most often occurred. The actual sensitivity of subjects to the LVHD pair is obscured in the raw data, but revealed by the statistical model.

3.3 Statistical model

In order to get at the real sensitivity of subjects to the differences between the various stimuli, we used a statistical model with the capacity to take into account the significant effect of block number on the results, as well as capture the effects of a number of other parameters. To streamline the model, the trajectory of the blocks was captured by a 4-term orthogonal polynomial contrast, instead of comparing all six blocks to one another. Also, rather than having each stimulus pair be represented and compared to the one before, we decided to consider the LDHD and LVHV pairs together as the f0 contrast, and the LDLV and HDHV pairs as the voicing contrast, having found no significant difference between parallel sides of the "square" of the stimulus array we considered earlier. The model takes into account the random effects of subject variability of sensitivity and bias in comparing blocks and in responding to the LVHD pair. Figure 3-3 shows the fixed effects of the logit mixed effects model.

Fixed effects of logit mixed effects model

Sensitivity to	Compared to	β	Z	p-value	Significance
f0 contrast	Chance	1.248339	3.352	0.000803	***
Voicing contrast	f0 contrast	0.451460	2.196	0.028063	*
LDHV pair	Voicing contrast	-0.172502	-0.537	0.591577	
LVHD pair	LDHV pair	2.094982	2.649	0.008078	**
Trial after error	Non-'l' stimulus	-0.182326	-1.311	0.189943	
Trial after switch	Non-'l' stimulus	-0.473180	-3.535	0.000408	***

Key for significance: p-value of: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure 3-3 Table of estimated regression coefficient, z-score, and p-value for each fixed effect given in the model.

The model gives the estimated regression coefficient as a measure of the effect of a particular parameter on the sensitivity of subjects to the given task. We find that the sensitivity of subjects to the f0-only contrast is significantly above chance. As expected from the description of the raw data, there is also a marked increase in sensitivity to the voicing contrast as compared to the f0 contrast. The model did not show a significant difference in sensitivity between the LDHV pair and the voicing contrast, once various effects of block and subject variability were factored out, although it is interesting to note that the β given is negative. However, we do find in this model that subjects are significantly more sensitive to the LVHD pair than to the LDHV pair, a result which was completely obscured in the raw data. We also found that sensitivity in each trial was affected by the immediately preceding trial, decreasing after an error or a switch in the stimuli. Finally, we looked at how sensitivity related to block. Using the 4-term orthogonal polynomial as a proxy for block number, the linear and quadratic terms were found to be significant (linear: $\beta = 0.93$, $z = 1.963$, $p < 0.05$; quadratic: $\beta = -1.10$, $z = -2.928$, $p < 0.01$). These two terms together correspond to the 'hat' shape of acclimation and fatigue found in Figure 3-2, showing that block number did have a significant effect on the results.

Figure 3-4 shows the sensitivity as determined by the statistical model we used. Once the effects of acclimation and fatigue by block number and trials following errors are regressed out, we find that subjects are significantly more sensitive to the difference between the LV and HD stimuli, affirming the hypothesized integration effect.

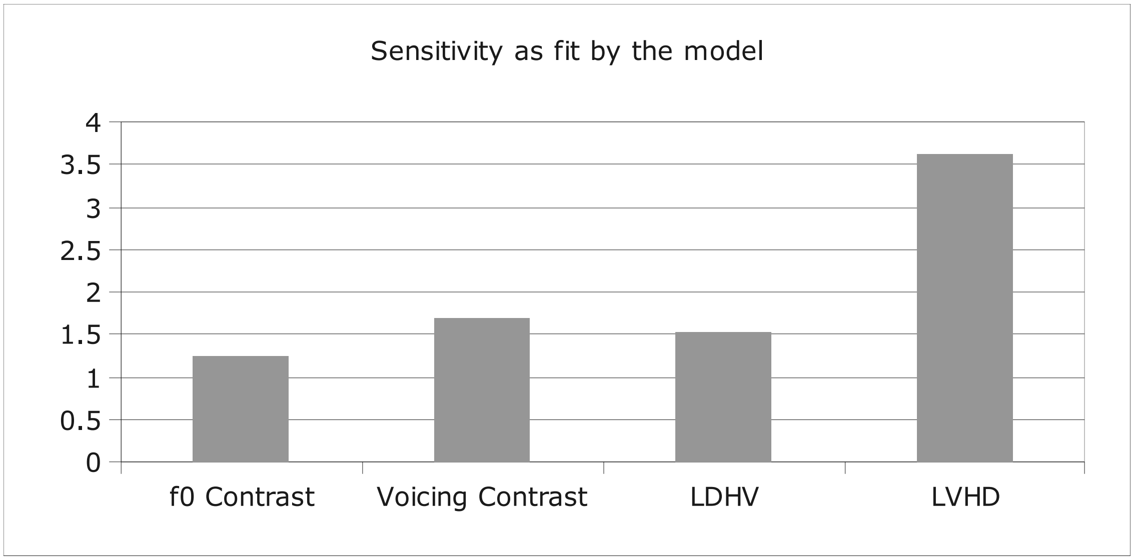

Figure 3-4 The sensitivity to each contrast as the log odds of hits relative to false alarms, by the condition tested. The two f0-only pairs and the two voicing-only pairs were collapsed into a single condition.

4. Discussion

4.1 Summary of results

The results of the experiment suggest that low f0 and voicing do indeed integrate perceptually as a “low frequency” intermediate perceptual property. When these two acoustic cues positively correlate with respect to this property, stimuli are easier to discriminate, as suggested by the higher sensitivity to the difference in the LVHD stimulus pair than that of the LDHV pair.

In discussing the Garner paradigm, we used a 2x2 stimulus array, reproduced below, to establish what stimuli would be used and how they would be constructed. If the stimuli were perceptually separable, meaning that they did not integrate in any way, this diagram could also represent the perceptual distance between stimuli.

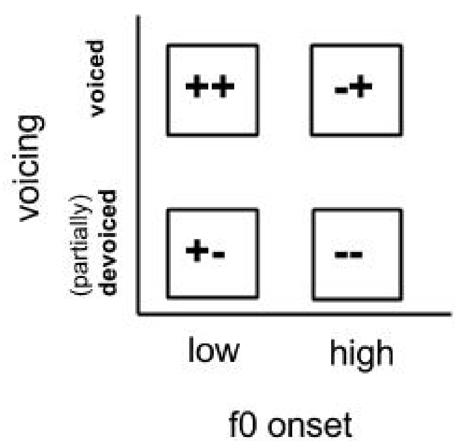

Figure 4-1 A 2x2 stimulus array showing the acoustic distances between stimuli with respect to the dimensions of voicing and f0 onset, after Kingston et al (2008). If these dimensions were perceptually independent, this figure would also represent the perceptual distance between stimuli.

However, we found that these two dimensions integrated perceptually in the context of the plain /s/, when the dimensions were varied within the natural variability of /s/. Thus, the perceptual distance might be better represented by a rhombus than a square, as shown in Figure 4-2.

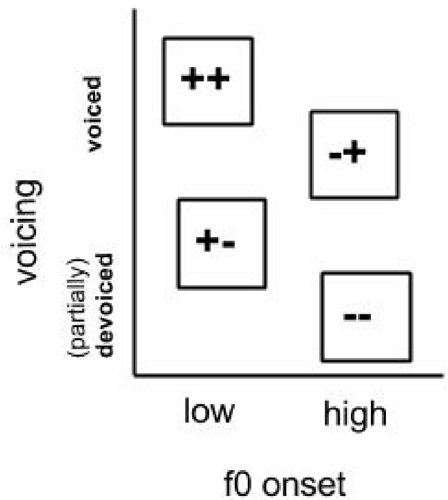

Figure 4-2 A 2x2 stimulus array mapped onto perceptual space in which the two dimensions, voicing and f0 onset, are perceptually integral.

Figure 4-2 shows that that parallel sides are equal in length, corresponding to the fact that each single-dimension contrast, whether f0-only or voicing-only, was found to be equally perceptually distinguishable in both stimulus pairs addressing that contrast. That is, subjects showed roughly the same average sensitivity to the LDHD and LVHV stimulus pairs, both testing the f0 contrast in differing voicing backgrounds. It should be noted that this sensitivity to the f0 contrast was fairly low, meaning subjects found it very difficult to differentiate between stimuli solely on the basis of the 60 Hz difference in f0 onset of the vowel following the plain /s/. The sensitivity to the voicing contrast was significantly higher, suggesting that the orthogonal sides of the figure are not equal in length, simply meaning that the 75% decrease in voicing did not affect perception to the same degree that the 60 Hz difference in f0 onset did. However, this does not have any bearing on the observed integration effect.

The most relevant aspect of Figure 4-2 is the clear difference in perceptual distance between the two diagonals; the integration of voicing and f0 onset allows for greater perceptual distance, and therefore ease of discrimination, between the LV and HD stimuli. On the other hand, the results suggested that the short diagonal between the LD and HV stimuli may be even shorter than the distance between the single dimension tasks. This is consistent with the IPP hypothesis because the perceptual distance allotted by varying one parameter is subtracted from when the other parameter contributes to the IPP in the opposite direction.

4.2 The Korean fricative contrast

This experiment tested perception within the plain /s/ category in order to examine what was happening on a purely psychoacoustic level, without the effect of phonological knowledge. However, these results address our initial question of why f0 onset plays a role in distinguishing between the plain /s/ and the fortis /s*/. Due to the fact that f0 onset and voicing were found to integrate perceptually, we can assume that the "low frequency" IPP is associated with the plain /s/. Thus, although f0 onset is not reliably produced lower in the context of the plain /s/ versus the fortis /s*/, speakers of Korean pay attention to the contribution of f0 onset to the low frequency IPP as one way of distinguishing between these two fricative categories.

4.3 Conclusion

This experiment investigated the relationship between voicing and f0 onset in the following vowel of the Korean plain fricative /s/. The results of this study reproduced the integration effect proposed by Kingston et al (2008) with respect to the laryngeal contrast in Korean fricatives. The integration of f0 onset and voicing in English stops, non-speech analogues, and now Korean fricatives strongly supports the notion that perceptual integration takes place as a low level auditory process, rather than being learned as significant for a particular contrast in a particular language. Although this study investigated the same low frequency IPP as discovered in Kingston et al (2008), the concept of cue integration could go a long way in explaining general auditory processes. The acoustic signal includes a plethora of distinct physical properties that could be grouped by the types of auditory effects they produce. The perception of any single phonemic contrast may be understood as the interaction of intermediate level properties that sum up multiple auditorily similar cues, rather than the notion that the auditory system is paying attention to every individual acoustic cue that is known to have a perceptual effect on that contrast.

REFERENCES

Chang, Charles B. "The Production and Perception of Coronal Fricatives in Seoul Korean" (2013).

Cho, Taehong, Sun-Ah Jun, and Peter Ladefoged. "Acoustic and aerodynamic correlates of Korean stops and fricatives." *Journal of Phonetics* 30.2 (2002): 193-228.

Holliday, Jeffrey. "An Acoustic Study of L2 Perceptual Acquisition of Korean Fricatives." *Harvard Studies in Korean Linguistics* 13 (2010): 17–32.

Jaeger, T. Florian. "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models." *Journal of memory and language* 59.4 (2008): 434- 446.

Kingston, John, et al. "On the internal perceptual structure of distinctive features: The [voice] contrast." *Journal of phonetics* 36.1 (2008): 28-54.

Yoon, Kyuchul. "A production and perception experiment of Korean alveolar fricatives." *음성과학* 9.3 (2002): 169-184.