Inter- and intra-L1 differences in L2 speech perception

Jeffrey J. Holliday

Department of Linguistics, The Ohio State University, USA
jeffh@ling.ohio-state.edu

Abstract

In a perception experiment, L1 Mandarin and L1 Japanese novice learners of Korean classified non-tense /s/- or tense /s*/-initial Korean CV tokens. A mixed effects logistic regression model with acoustic cues as predictor variables was built for each L1 group, and each individual's regression coefficients were interpreted to be the cue weighting used in identifying Korean /s/ and /s*/. We propose that the weighting of L2 perceptual cues is influenced by the weighting of the same cues in the L1 perception of acousticaly similar contrasts, but that intra-L1 individual variation is great enough that the expected inter-L1 differences may appear less well defined.

Index Terms: phonetics, fricatives, Korean, Mandarin, Japanese

1. Introduction

1.1. Perception by naive L2 listeners

According to Best [1], non-native (L2) speech sounds will sound most similar to native (L1) speech sounds that are perceived to be articulated in a similar way. When an L2 speech sound is particularly well-matched to an L1 speech sound in this respect, it is said to have "assimilated" to the L1 phonological category. L2 contrasts involving two speech sounds that can both be assimilated to an L1 phonological category are often the most difficult for L2 listeners (whether naive or not) to perceive. For example, native speakers of Japanese often have difficulty perceiving the difference between English /r/ and /l/ [2] [3]. The most likely explanation for this is that Japanese has only one phonological category that is acoustically similar to either English /r/ or /l/, and hence tokens of both English categories are perceived as belonging to, and thus assimilate to, a single Japanese category.

In [1], Best summarizes the possible L2 phonological assimilation patterns in her Perceptual Assimilation Model (PAM). L2 speech sounds can either assimilate to an L1 phonological category, be recognized as a speech sound but not assimilate to an L1 phonological category, or not be recognized as a speech sound at all. Among L2 contrasts, then, the most difficult challenge arises when both members of the L2 contrast assimilate as good exemplars of a single L1 phonological category. Anecdotal evidence suggests that the cases to be investigated in this study, the assimilation of Korean non-tense /s/ and tense /s*/ to Mandarin Chinese /s/ and Japanese /s/, are examples of this pattern.

1.2. Perception by L2 learners

PAM describes the assimilation patterns of L2 speech sounds by naive listeners, but it does not (nor was it intended to) describe how L2 learners perceive L2 speech sounds [4]. This is because the assimilation patterns outlined in PAM do not necessarily hold throughout the L2 acquisition process. As naive listeners are exposed to meaningful input they slowly build perceptual categories, and the incoming L2 speech sounds may not assimilate to the L1 categories to which they originally assimilated; rather, incoming L2 speech sounds may be correctly categorized into the newly formed L2 phonological categories. This raises the question of how an L2 learner can make the jump from simply assimilating L2 speech sounds into L1 categories to actually categorizing L2 speech sounds into L2 categories. A naive listener can by default only do the former, but an L2 learner should eventually be able to do the latter.

In this paper, we assume that there exist acoustic cues that can help the listener distinguish members of a phonological contrast. It follows that the task of learning to identify members of an L2 contrast amounts to learning to attend to the cue or set of cues that can reliably discriminate its members. Since it seems reasonable to assume that all humans share the same basic auditory function, it should not be the case that these acoustic cues are physically unavailable to L2 learners; instead, it is probably the case that L2 learners are not skilled at attending to them. Furthermore, because phonological contrasts in different L1s are realized in acoustically different ways, we might expect L2 learners from different L1 backgrounds to exhibit some degree of L1 influence on their weighting of acoustic cues in perception.

On the other hand, if the process of learning an L2 contrast involves learning the correct perceptual strategy (i.e. by attending to the correct acoustic cues) we might expect to find variation in terms of initial strategy selection. If learning occurs, over time these differences should disappear as learners eventually settle on the correct strategy. Work by Vihman [5] suggests that in the case of infants learning their first language, learning is a process of selecting an initial strategy and then fine-tuning it based on input. It seems reasonable to assume a similar model for L2 acquisition, except that the L1 may affect the initial strategy selection in ways not possible for infants learning their L1.

1.3. Korean, Mandarin, and Japanese sibilant fricatives

In this section we will briefly introduce the acoustic properties of the fricatives in each language that are relevant to the current study.

1.3.1. Korean /s/ and /s/*

Korean /s/ and /s*/ differ along several acoustic dimensions, including aspiration, $f0$, spectral tilt, intensity, and centroid freqency. Of these dimensions, however, only centroid frequency is a measure from within the fricative itself. $f0$, spectral tilt, and intensity are always measured in the vowel following the fricative and any intervening aspiration. In the remainder of this section we will discuss the findings of previous studies that have looked at these measurements in the Korean fricatives.

Korean /s/ has been claimed to be phonetically aspirated [6], which agrees with Kagaya's [7] observation that the glottal gestures of /s/ are quite similar to those of the aspirated stops. Cho et al. [8] point out that Korean /s/ frication is often followed by a period of aspiration, whereas /s*/ frication is sometimes followed by a short period of glottalization.

Cho et al. [8] report that the following vowel has a higher $f0$ for /s*/ relative to the value for /s/, but these results combine data from productions of both Seoul and Jeju dialect speakers. When looking at the Seoul dialect alone, however, /s*/ appears to have an only slightly higher $f0$ in the following vowel than /s/ does.

Spectral tilt describes the degree by which the amplitudes of progressively higher harmonics taper off in a spectrum. Differences in spectral tilt are easiest to measure near the fundamental frequency, where the amplitudes of the glottal wave components are highest relative to the amplitudes of the formants; hence a common method is to calculate the amplitude of the first harmonic relative to the second (H1-H2). Cho et al.'s [8] results show that H1-H2 in /s/ is significantly higher than in /s*/ at vowel onset, although the difference becomes non-significant at the vowel midpoint.

Chang [9] used two measurements of intensity for Korean fricatives: intensity buildup and average vowel intensity. Intensity buildup was intended to approximate the rate of change in intensity, and was calculated as the difference in intensity between consecutive glottal pulses. Chang measured this for the first four glottal pulses after vowel onset and found that intensity buildup is significantly more rapid in [s*a] than [sa], but was not significantly different between [s*u] and [su]. Average intensity was not found to differ significantly in vowels following /s*/ or /s/, in both /a/ and /u/ contexts.

Centroid frequency is calculated by treating the fricative spectrum as a normal distribution and finding its mean. Cho et al. [8] found that /s*/ has a significantly higher centroid frequency than /s/. This is most likely caused by the difference in constriction between /s*/ and /s/. First, /s*/ has a much tighter constriction than /s/ which makes the constriction tube longer and effectively decreases the size of the front cavity. Second, because the constriction of /s/ is much looser, there is a greater degree of back cavity coupling. The back cavity resonances are lower in frequency than the front cavity resonances and thus bring down the mean frequency of the spectrum. As the tongue constriction is released in the transition from /s/ to the following vowel the back cavity coupling gives way to aspiration and the centroid becomes even lower.

Previous published studies of /s/-/s*/ perception by L1 Korean listeners suggests that multiple cues may be at work. Yoon [10] determined that the duration of aspiration was the most important cue but that some other cues must also be at work because it had much less of an effect before /i/ than before /a/. Chang [9] found that a combined cue of F1 onset, intensity buildup, and spectral tilt correlated the most with native listeners' fricative classifications. Holliday [11], using the same stimuli and methodology as in the current study, found that aspiration duration was the most important cue, followed by centroid frequency. The discrepancies among these results suggest that vowel context and other stimulus factors may affect L2 assimilation patterns as well.

1.3.2. Mandarin /s/, /ɕ/, and /ʂ/

Mandarin Chinese has three sibilant fricatives, /s/, /ɕ/, and /ʂ/. Retroflex /ʂ/ contrasts with denti-alveolar /s/ and alveolopalatal /ɕ/ in terms of place of articulation. It can thus be easily distinguished from the other two sibilants by its lower centroid frequency [12]. The acoustic difference between /s/ and /ɕ/ is less well understood. Two measures suggested in [12], ampRatio (the difference in dB between the spectral peak and the F2, intended to measure palatalization) and centroid frequency in the Bark band above the F2, could classify Mandarin female fricative tokens with 92% accuracy using linear discriminant analysis. Their results also show that centroid frequency alone can separate /ʂ/ from /s/ and /ɕ/ quite well. In [13], results of a perception experiment indicate that centroid frequency can indeed distinguish /ʂ/ from /s/ and /ɕ/ in perception, whereas onset F2 is needed to distinguish /ɕ/ from /s/ and /ʂ/.

1.3.3. Japanese /s/ and /ɕ/

Japanese /s/ and /ɕ/, like their Mandarin analogs, are also not clearly distinguished by centroid frequency, probably because the difference in front cavity size is not robust [14]. The two measures suggested in [12] could also classify Japanese female fricative tokens with 92% accuracy using linear discriminant analysis; however, centroid frequency alone could not separate /s/ from /ɕ/ the way it could separate /ʂ/ from /s/ and /ɕ/ in Mandarin. Results of a separate perception experiment [13] showed that Japanese listeners likely rely on a combination of centroid frequency and onset F2 of the following vowel to classify Japanese fricative tokens. More studies are needed to verify these results and identify the perceptual cues with more certainty.

1.4. Research questions

In the early stages of L2 acquisition learners may not perform much differently from naive L2 listeners on any of a variety of tasks, but with enough experience they eventually reach a point that is statistically different from chance. It is obvious that there is some learning occuring during the time between these two points. To claim otherwise would be to assume that learning is instantaneous.

Thus, when comparing the performance of any two L2 learners we must disentangle at least three things: how much of their performance can be accounted for by L1 influence; how much of it can be attributed to extra-linguistic or "external" factors such as motivation and input; and how much of it must be relegated to random individual variation or some general language aptitude that we lack the means to quantify. The first of these is an inter-L1 difference, whereas the latter two are intra-L1 differences.

The purpose of the research presented in this paper is to observe the L1 influence by comparing the use of acoustic cues in classifying tokens of a difficult L2 contrast by novice learners from two different L1 backgrounds. Specifically, which acoustic cues do native Mandarin and Japanese listeners use when trying to identify Korean /s/ and /s*/? Previous studies have shown that centroid frequency by itself can separate one of the Mandarin sibilants from the two others, whereas no separation is possible in Japanese without the use of an additional cue. Thus, if there are effects of L1 influence, we might expect that Mandarin listeners are more skilled at paying attention to differences in centroid frequency and may rely on it more than Japanese listeners do in perceiving the Korean sibilants. In addition, because Mandarin has lexical tone we also might expect Mandarin listeners to pay more attention to differences in $f0$ than Japanese listeners. Therefore, we might expect to see evidence that L1 Mandarin listeners employ perceptual strategies that rely on centroid frequency and $f0$ more than L1 Japanese listeners. Furthermore, do listeners from the same L1 group look more like each other than listeners from the other L1 group, or are there listeners from different L1 groups who pattern similarly?

2. Methods

2.1. Participants

37 native Mandarin speakers and 22 native Japanese speakers were recruited from intensive Korean language programs at two universities in Seoul. The vast majority had been studying at these programs for approximately 6 to 20 weeks (in their first or second 10-week quarter), but a few had been studying for three or four quarters. By recruiting students who were all residing in Korea and receiving the same amount of classroom instruction time we hoped to control for level and type of input as much as possible.

2.2. Materials and procedure

Listeners were presented 324 word-initial fricative-vowel syllables excised from real-word productions of native speakers of Seoul Korean. That is, each token was either /sV/ or /s*V/. Fricative category and speaker gender were balanced across tokens. Vowel context was balanced across fricative category, but could not be fully balanced within fricative category.

The stimuli were presented one at a time using high-quality headphones and with the participant seated in front of a laptop. Participants responded by pressing a button to classify each token as either /s/ or /s*/.

The five acoustic parameters said to help differentiate Korean /s/ and /s*/ (aspiration duration, centroid frequency, intensity, and $f0$ and spectral tilt of the vowel) were measured for each token.

2.3. Data analysis

The results from each L1 group were analyzed separately. The five acoustic parameters were used as predictor variables in a mixed effects logistic regression model to predict individual listener choice, with /s*/ set as '1' and /s/ as '0'. We interpret the coefficients returned by the model for each individual listener to correspond to the listener's weighting of that acoustic cue in perception. A hierarchical clustering analysis was performed separately on each L1 group's listeners' coefficients to see whether there are groups of individual listeners that pattern similarly.

3. Results

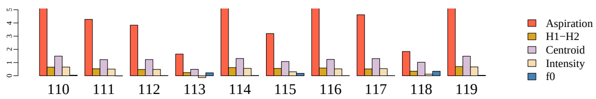

Figure 1: *L1 Korean cue weightings. 10 of 13 subjects are presented here. The y-axis scale has been zoomed out to show the full and relative magnitudes of each coefficient. Note that the f0 coefficient barely registers at all for most listeners.*

Adding spectral tilt did not contribute significantly to the L1 Mandarin group's results and resulted in an overfit in the L1 Japanese results, and was thus excluded from both models.

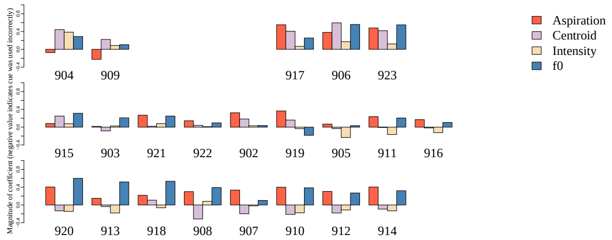

Figure 2: *L1 Japanese cue weightings. The first row contains two clusters and the second and third rows are separate clusters.*

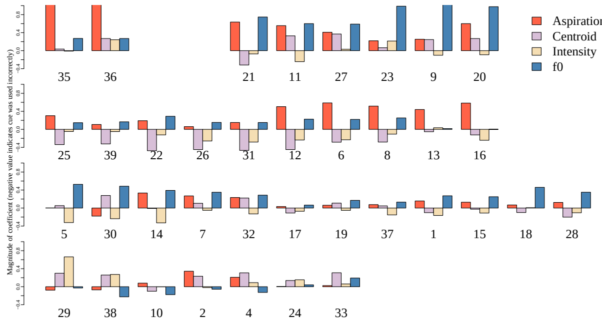

Figure 3: *L1 Mandarin cue weightings. The first row contains two clusters and the second through fourth rows are separate clusters.*

Figure 1 shows the results of 10 native Korean listeners presented in [11]. Figures 2 and 3 show the results from the current study. Each individual's cue weightings appear as a separate subplot, with four or five bars for the cues compared in each model. The number underneath each subplot is an identification code for the individual subject. The first rows in Figures 2 and 3 contain two separate clusters, whereas each of the remaining rows represents a separate cluster. The magnitude of each bar represents the coefficient value returned by the regression model, and the sign (positive or negative) represents whether the cue was used in a native-like way or not. Thus when considering how much a cue was attended to it is important to ignore the direction of the bar.

Figure 1 shows that there is clearly a set of "native" cue weightings that L1 Korean listeners use with only limited inter-individual variation. Korean listeners rely on aspiration the most, followed by centroid frequency. These results are presented in detail in [11].

Having established what a native-like perceptual strategy looks like, next we turn to the differences between the Japanese and Mandarin groups. Although the cluster analysis makes it easier to observe trends, the interpretation of Figures 2 and 3 is not obvious.

Some Mandarin listeners, most notably in the third cluster (subjects 25 through 31), appear to rely more on centroid (the second bar) than the other cues. Among the Japanese listeners, on the other hand, only subjects 904 and 906 rely the most on centroid frequency, and it competes closely with other cues.

Some other Mandarin listeners, such as 35, 36, 13, and 16, rely on aspiration (the first bar) much more than the other cues. Such listeners cannot be found in the Japanese group, with the possible exceptions of 902 and 919, although the magnitudes of their aspiration coefficients remain lower than that of many of the Mandarin listeners. This result was unexpected from the Mandarin listeners, but makes sense when one considers the fact that Mandarin has two affricates, /ts/ and /ts^h/, that are separated along the aspiration dimension. Mandarin listeners may be transfering the use of an aspiration cue to their fricative perception in Korean.

Although we predicted that the Mandarin listeners would be more sensitive to differences in $f0$ and hence rely on it more than Japanese listeners would, we found that $f0$ (the fourth bar) was used as a primary cue by several Japanese listeners as well. For a novice learner of Korean, attempting to distinguish /s/ from /s*/ based on $f0$ seems like a reasonable strategy because the plosive and affricate series all use $f0$ as a primary cue to distinguish the aspirated and lax categories. If a learner of Korean has picked up on the $f0$ difference in the plosives and affricates and then makes the incorrect assumption that the fricatives follow a similar pattern, we would expect to see the types of patterns found by some listeners in Figures 2 (e.g. subjects 913 and 918) and 3 (e.g. subjects 23, 9, and 20).

The use of the $f0$ as a perceptual cue is not very native-like. Notice in Figure 1 that the $f0$ bar is barely even present in any of the plots. Because no studies have found significant $f0$ differences between /s/ and /s*/ in Seoul Korean, especially in comparison to the differences found in aspiration and centroid, we would not expect L1 Korean listeners' responses to correlate at all with the $f0$ of the stimulus. However, for some of the L2 listeners we found that $f0$ predicts the listener's response better than any of the other cues. This is likely an example of an L2 listener using an ineffective cue to try to identify members of a new perceptual contrast.

4. Discussion

The goal of this research was to observe how much influence the L1 has on perceptual strategy selection by novice L2 learners. To the extent that there is no influence, we must consider non-L1 effects such as motivation, input, and random variation. Our results revealed some L1-specific trends, but also a high level of intra-L1 variation.

Although we assume that intra-L1 variation is due to the presence of other effects unrelated to the L1, it is also possible that we have not selected the correct parameters to put into the model. That is, the acoustic cues that the listeners from each L1 group attend to the most may not be among the five we selected. The five acoustic cues in the current model were selected based on the suggestions of previous studies that they can, either individually or as a set, reliably differentiate the Korean /s/ and /s*/ categories. If indeed onset F2 is what Japanese and Mandarin listeners rely on to differentiate /ɕ/ from the other sibilant categories, it would be worthwhile to see whether such learners of Korean try to use onset F2 in identifying Korean /s/ and /s*/. One could argue that non-native learners would not attend to onset F2 because it would be such an unreliable cue for identifying Korean sibilants, but that could also be precisely why the Korean /s/-/s*/ contrast is so difficult for native speakers of Japanese and Mandarin.

However, assuming we have indeed chosen the five most appropriate parameters to put in the model, the implication is that L2 perception of a new contrast is only partially shaped by the role of acoustic cues in the L1 and that there exists an element of individual strategy selection that may be random in nature. Our results here suggest that, at least for novice learners of Korean, non-L1 effects play an important role in L2 phonological acquisition. As we continue collecting data, including longitudinal data for some of the subjects presented here, we hope to see how the effect of L1 influence changes as proficiency increases.

5. Acknowledgements

We would like to thank Mary Beckman and Cynthia Clopper for their helpful advice, and also the Korean Language and Culture Center at Korea University and the Department of Korean Language & Culture at Kyunghee University for help in recruiting participants.

6. References

[1] Best, C. T., "A direct realist perspective on cross-language speech perception," In Strange, W. (Ed.) *Speech Perception and Linguistic Experience* (pp. 167-200). Timonium, MD: York, (1995).

[2] Sheldon, A., Strange, W., "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception," *Applied Psycholinguistics*, 3:243-61, (1982).

[3] Logan, J. S., Lively, S. E., Pisoni, D. B., "Training Japanese listeners to identify English /r/ and /l/: A first report," *JASA*, 89:874-886, (1991).

[4] Best, C. T., Tyler, M., "Nonnative and second-language speech perception: Commonalities and complementarities," In *Language experience in second language speech learning: In honor of James Emil Flege*, Edited by O.-S. Bohn and M. Munro. Amsterdam: John Benjamins, (2007).

[5] Vihman, M. M., "Variable paths to early word production," *Journal of Phonetics* 21:61-82, (1993).

[6] Park, H., "The phonetic nature of the phonological contrast between the lenis and fortis fricatives in Korean," In Proc. of ICPhS XIV, 1:424-427, (1999).

[7] Kagaya, R., "A fiberscopic and acoustic study of the Korean stops, affricates and fricatives," *Journal of Phonetics* 2:161-180, (1974).

[8] Cho, T., Jun, S.-A., Ladefoged, P., "Acoustic and aerodynamic correlates of Korean stops and fricatives," *Journal of Phonetics* 30:193-228, (2002).

[9] Chang, C. B., "Korean Fricatives: Production, Perception, and Laryngeal Typology," *UC Berkeley Phonology Lab Annual Report* 2007: 20-70, (2007).

[10] Yoon, K. "A study of Korean alveolar fricatives: An acoustic analysis, synthesis, and perception experiment," MA Thesis, University of Kansas, (1999).

[11] Holliday, J. J. "An acoustic study of L2 perceptual acquisition of Korean fricatives," In Ik-Hwan Lee (ed.), *Harvard Studies in Korean Linguistics XIII*. Cambridge, MA: Harvard University, (2010).

[12] Li, F., Edwards, J., Beckman, M. E., "Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese," In Proc. of ICPhS XVI, (2007).

[13] Li, F., "The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese," Ph.D. dissertation, The Ohio State University, (2008).

[14] Toda, M., Honda, K., "An MRI-based cross-linguistic study of sibilant fricatives", Proc. of the 6th Int'l Seminar on Speech Prod. (2003).