Using spectral measures to differentiate Mandarin and Korean sibilant fricatives

Jeffrey Kallay, Jeffrey J. Holliday

Department of Linguistics, The Ohio State University, Columbus, OH, USA
kallay.11@osu.edu, jeffh@ling.ohio-state.edu

Abstract

We explore the use of two spectral measures calculated in ERB space for differentiating between the friction noise of sibilants in Mandarin Chinese and Korean. The peak frequency (peakERB) of the spectral representation was used to capture differences in front cavity size and a diffuseness index (DI) was used to capture the bandwidth of the peak. In both [a] and a high vowel context, the peakERB measure differentiated between Mandarin [s], [c], and [j], and also between Korean [s*], [s*], and [s*] which, although considered to be articulated at the same place, differ in front cavity size due to the tighter lingual constriction of [s*]. The DI measure helped further differentiate the fricatives, with Mandarin [c] having a broader peak (higher DI) than [s] or [j], and Korean [s*] having a broader peak than [s*]. When applied to the L2 Korean productions of L1 Mandarin speakers, we found evidence for both Korean fricatives assimilating to Mandarin [s] before [a], and to Mandarin [c] before [i].

Index Terms: sibilant fricatives, Mandarin Chinese, Korean

1. Introduction

Mandarin Chinese has voiceless sibilant fricatives at three places of articulation, although the alveolopalatal [c] is sometimes analyzed as being in complementary distribution with the other two, occurring before [i], [y], [e], [j], and [q], where the alveolar [s] and postalveolar [j] do not occur [1]. Korean has two alveolar fricatives, which contrast primarily in terms of their laryngeal settings, so that [s*] is associated with breathy voice on the following vowel, and an interval of aspiration before non-high vowels [2], whereas [s*] is associated with pressed voiced on the following vowel, as well as a more forceful (“tense”) articulation overall [3], so that spectral energy during the fricative is more intense and more concentrated at higher frequencies. Both Korean fricatives become palatalized to [c*] and [c*] in the same [i] and [j] environments in which the Mandarin alveolopalatal fricative occurs, and in these cases their differences in aspiration and voice quality are greatly reduced. The palatalized variants of the Korean sibilants do not occur in the other environments in which the Mandarin alveolopalatal does, however, and so their distribution is not as broad. Furthermore, this also means that the Korean [s*] and [s*] have a less circumscribed distribution than the Mandarin alveolar and postalveolar sibilants.

Finally, it should be noted that while we use the notation [i] to refer to the high-front vowel in this paper regardless of the preceding sibilant, the actual quality of this vowel varies in Mandarin, to be more like the central vowel [i] following [s], and a rhotic variant of that vowel following [j]. Given this, the Mandarin high vowel is not always parallel to the high-front [i] of Korean. In fact, Korean also has the centralized [i], and it contrasts with [i] following other non-sibilant consonants in a way which does not occur in Mandarin. Thus, the variation between the Korean sibilants and their palatalized counterparts is easily analyzed as being conditioned by the contrasts in the following vowels (an analysis reinforced by Korean orthography), whereas in Mandarin the analysis of the conditioning could go in either direction. That is, rather than the variation in the sibilant being conditioned by the following vowel, it could be the variation of the quality of the following vowel being conditioned by the preceding sibilant (an analysis which is also reinforced by Chinese pinyin). The purpose of this paper is to explore the use of spectral measures of friction to quantify the phonemic and allophonic relationships between the fricatives of Mandarin and Korean. To this end, we analyzed the fricative productions of both L1 Korean speakers and L1 Mandarin speakers who were also novice L2 learners of Korean.

2. Methods

2.1. Participants, materials, and recordings

Recordings were collected from 10 female native speakers of Mandarin (L1M) and 7 female native speakers of Korean (L1K). The native Mandarin speakers were all enrolled in their first quarter of intensive Korean class at universities in Seoul. The native Korean speakers, in their 20s or early 30s, were either born in Seoul or moved to Seoul at a young age.

The materials were randomized lists of Mandarin
words (for the L1M group) and Korean words (for both groups), each of which included 18 target sibilant-initial words. For Mandarin, there were three words for each of [sa], [ja], [eja], [si], [ji], and [ci], and for Korean there were five words for [s*a] and [s*a] and four for [e*i] and [e*i].

Participants were seated at a desk or table in front of a University Sounds US658H microphone mounted on a table-top mic stand approximately 15 to 20 cm from the participant’s mouth. Participants were instructed to read the lists at a comfortable pace, inserting at least 1 second between words, and to avoid using list intonation. The L1M recordings were done with the microphone connected to a laptop computer through a Roland UA-30 USB interface. The L1K recordings were done with the same microphone connected directly to a Roland Edirol digital recorder. All recordings were sampled at 44,100 Hz.

2.2. Spectral measures

The spectral measures were derived from a representation that reflects certain sensitivities of the human auditory system. To compute the values of the spectral representation, we took a 40 ms rectangular window from the midpoint of each word-initial fricative and computed its power spectrum using the FFT function in R. It has been shown that the bias properties of a rectangular window are much smaller at frequencies closer to the peak, having little effect on the peak itself, and so its use was appropriate for the purposes of this study [4, p. 227]. The power spectrum was then transformed through a 361-channel 4th order gammatone filterbank with the bandwidth set to 1.019 times the ERB number of the center frequency [5]. The center frequencies of the channels ranged from 3 to 39 ERB, spaced at 0.1 ERB intervals. The spectral representation is then the total power of the output of each channel plotted against the ERB number of the center frequency of that channel. The two spectral measures we used were the peak ERB frequency of the excitation pattern and a measure of the diffuseness of the peak.

2.2.1. Peak ERB frequency (peakERB)

Previous studies [6] have shown the peak spectral frequency to be useful in capturing place distinctions in sibilant fricatives. Our peakERB measure is the ERB frequency at which the total output power of the filter is the highest.

2.2.2. Diffuseness index (DI)

Our second measure was a diffuseness index (DI), which is a measure of the relative spread or bandwidth of the output power. We centered a band at the peakERB with the upper and lower limits set to the points at which there was a 9 dB drop on either side of the peak. The DI value was then obtained by subtracting the ERB value at the lower limit of the band from that at the upper limit. Thus, a higher DI value corresponds to a wider peak.

3. Results

3.1. peakERB: L1 productions

For the native productions of both the Mandarin and Korean speakers, peakERB appears to be a reliable measure in distinguishing between sibilant fricatives at different places of articulation. The first three rows of Figure 1 show the excitation patterns of the native Mandarin productions (labeled L1M), with the /a/ and /i/ tokens in the left and right columns, respectively. The gray lines represent individual productions and the dark lines represent the mean of all productions from all speakers for each CV target. The mean peakERB is indicated at the top of each plot.

The first three rows of Figure 1 show that [s] has the highest peakERB in both vowel contexts, with mean values of 34.3 for /a/ and 34.9 for /i/. [j] has the lowest peak in each context, with mean values of 27.7 for /a/ and 26.9 for /i/, while [c] is in between with mean values of 31.1 for /a/ and 32.1 for /i/. The peakERB values for each fricative vary minimally across vowel contexts.

The bottom two rows of Figure 1 show the plots of the native Korean productions. [s*a] has a mean peakERB of 34.7 and 31.5 for /a/ and /i/, respectively, whereas [s*i] has lower mean peaks of 32.4 and 29.9 in the same contexts. Thus, the mean peakERB of the Korean fricatives were lower when they preceded /i/, although this difference between the two vowel contexts was not as big as the difference between Mandarin [s] and [c]. However, this difference between vowel contexts was bigger than the difference between the two fricatives, particularly in the /i/ context.

3.2. peakERB: L2 productions

The fourth and fifth rows of Figure 1 show the native Mandarin speakers’ L2 Korean productions. Their overall spectral shape and peakERB values suggest that native Mandarin speakers assimilate both Korean [s*a] and [s*i] to Mandarin [s] before /a/ and to Mandarin [c] before /i/. The peakERB values of L1M [sa], L2K [s*a], and L2K [s*i] are all similar, as are the peakERB values of L1M [ci], L2K [c*i], and L2K [c*i].

Overall, this data seems to suggest that the native Mandarin speakers do not distinguish between Korean [s*i] and [s*a] as separate phonemes in production. They do appear to be distinguishing between the allophonic variants [s*i]-[c*i] and [s*a]-[c*i], but they assimilate these Korean allophones to different Mandarin categories based on acoustic similarity.
3.3 DI: L1 productions

Figure 2 shows the DI values plotted against peakERB for both L1 groups’ native productions. In the Korean

Table 1: Median DI values for each fricative-vowel context for each L1 and L2 group.

Looking first at the L1M productions, there appears to be a good deal of variation between the spectral shapes of the postalveolar [] depending upon the vowel context. However, for both the alveolar [s] and alveopalatal [c] there is less of this variation. In either context, the alveolar has the lower median DI, with values of 6.4 before /a/ and 6.0 before /i/. The DI values of the alveopalatals are about 2 to 3 ERB higher, meaning that the spectral peaks of this sound are wider, which is what would be expected given the way in which these sounds are articulated. In the plots of Figure 2, these differences can be seen in the locations of the cross and circle clusters along the y-axis.
For the L1K productions, the differences in DI values between the sibilants appear to be much more dependent upon vowel context. For the ones before /a/, the tense \([s^*]\) is fairly tightly clustered near the median of 6.4. The values for the aspirated \([s^h]\) appear to be much more variable, but they can be seen to be generally higher in this context, which is reflected in the median value of 11.0. This is the highest value of any group of sounds, and is reflective of the high degree of aspiration of \([s^h]\) in this context, which brings down the peak. By contrast, the Korean sibilants appear to be nearly indistinguishable by this measure before /i/ context, with the same rounded value of 8.1. In Figure 2, the plots of \([c^*i]\) and \([c^h]\) are much more tightly clustered. These values suggest that there is a noticeable contrast in phonation between the two phonemes before a low vowel, but that this contrast is greatly reduced when the following vowel is high.

### 3.4. DI: L2 productions

In comparison to their native productions, the DI values of the Mandarin speakers’ L2 productions follow a very similar pattern as the peakERB values did. In Figure 3 the DI values of the L2 sibilants in each context are shown plotted against the peakERB along with the native phonemes to which they most closely assimilate in terms of this measure.

In the /a/ context, the spectral shape of the L2K \([s^*]\) is similar to the L1K productions of this sound, with a median DI of 6.9 compared to 6.4. Unlike the native productions, however, there is no evidence of a phonation contrast between \([s^*]\) and \([s^h]\) in this context by the L2K speakers. In fact, the DI values for the two phonemes are nearly identical, suggesting that they are not distinguishing between them at all in this regard. In the left panel of Figure 3, it can be also be seen that both of these L2 sounds are generally clustered around the same DI values of their L1 alveolar sibilant, which suggests that these speakers have a tendency to map both Korean sibilants to this native sound, and ignore the contrast that the native Korean speakers distinguish.

In the /i/ context, the L2K productions have median DI values of 8.3 for \([c^*]\) and 8.2 for \([c^h]\). These are very similar to the corresponding DI values of the L1K speakers. Contrary to the /a/ context, however, these values differ in terms of their relationship to their native sibilants. While both L2 sounds patterned with the alveolar before /a/, they appear to be more closely associated with the alveolo-palatal before /i/. In the right panel of Figure 3, it can be seen that the most noticeable cluster of these three sounds is higher than in the left panel, and the median DI values for all three are very similar.

### 4. Conclusion

The peakERB and DI measures captured differences in spectral peak and bandwidth that correspond to the sibilant fricative contrasts in Mandarin and Korean before /a/ and /i/. These spectral measures also provided some evidence that L1 Mandarin learners of Korean assimilate Korean sibilants to the most acoustically similar Mandarin sibilant categories, eschewing the analysis reinforced by Korean orthography that \([s^h]\] and \([c^h]\] (and \([s^*]\) and \([c^*]\)) should be thought of as the “same” sound. This finding agrees with other evidence suggesting that native Mandarin learners of Korean assimilate Korean sibilants in perception as well, with Korean \([s^a]\) and \([s^h]\) assimilating to Mandarin \([sa]\) and Korean \([c^h]\) and \([c^*i]\) assimilating to Mandarin \([ci]\) [7].

### 5. Acknowledgements

We would like to thank Pat Reidy for help in implementing the transform in Section 2.2. This work was funded by an NSF grant (#1024286) awarded to Mary Beckman and the second author.

### 6. References


