euronoise

The role of tongue articulation for /s/ and /z/ production in whispered speech

Hirohide Yoshioka

University of Tsukuba, 1-4-5 Shinmei, Urawa, 336-0023 Saitama, Japan
hirohide-y@mtc.biglobe.ne.jp

Although the timing of the initiation and cessation of vocal fold vibrations is crucial to characterize the voiced and voiceless cognates, other cues, such as the duration of preceding vowels, the patterns of the formant transitions in the following vowels, and the period of stop closure, may also play important roles in performing these distinctions. The present study is to further investigate the role of tongue articulatory movements during voiced and unvoiced consonant production, specifically when the vocal folds do not vibrate during the production of fricative consonants, /s/ and /z/ in whispered speech. A normal Japanese speaker served as the subject. The palato-lingual contact patterns during the intervocalic consonant /s/ and /z/ in whispered speech were recoded using dynamic electro-palatography. The results show that the area of palato-lingual contact is clearly wider during /z/ production than /s/ production; the opening for fricative turbulent noise production is narrower and longer for /z/ than for /s/. In addition, the contact pattern is unstable for /s/ production.

1 Introduction

Voiced sounds can be identified as vertical striations in the narrow band sound spectrograms. These vertical striations indicate that the vocal folds do vibrate during the pertinent periods.

The presence or absence of vocal fold vibrations, more specifically the timing of the initiation and cessation of vocal fold vibrations is crucial to differentiate the voiced and voiceless sounds (Lisker and Abramson, 1964).

According to traditional phonetics, manner and placement of articulation are considered to be identical within a set of voiced and voiceless cognates. However, the perceptional experiments by use of synthesized speech sound reveal that the duration of preceding vowels (Denes, 1955), the patterns of the formant transitions in the following vowels (Liberman, Delattre and Cooper, 1958), and the period of stop closure (Lisker, 1978) may also play important role in performing these distinctions.

These acoustic parameters, such as the duration of vowels, the closure period of consonants, and the formant transitions of vowels are originated via the temporal patterns of the articulatory movements of speech organs, such as lips, mouth, jaw, pharynx and so on.

Among others, the tongue movement with special reference to the hard palate is highly believed to perform the differentiation of these acoustic parameters.

The present study was conducted to further investigate the role of tongue articulatory movements during the voiced and unvoiced consonant production, specifically when the vocal folds do not vibrate during production of fricative consonants, /s/ and /z/ in whispered speech.

The author believes that speaker will differentiate a set of voiced and unvoiced cognates more clearly using other articulatory organs under such condition as the absence of vocal fold vibration during whispered speech.

2 Method

A female normal Japanese speaker served as the subject. The palato-lingual contact patterns were recorded by use of the elecro-palatography (RION DP-01). The artificial palate contains 63 electrodes, which deliver the on-and-off signals of the contact of tongue at each point. These signals were recorded at a sampling rate of 64 frames per second.

The linguistic materials were meaningless words /kaCaka/ (C=s, z), which contain /s/ or /z/ in the word-medial position. The subject was instructed to pronounce these words in whispered speech more than 50 times, respectively, at a random order.

The acoustic signals were also recorded by use of digital audio tape recorder (SONY TCD-03). The audio tape was played back to be used as the speech materials for the perceptual experiments.

For this perceptual experiments, four normal Japanese speakers were requested to judge each sound as /s/ or /z/, after listening audio signals on playback.

3 Results

3.1 Perceptual judgments

Fig. 1 contains the results of the perceptual experiments. Each bar chart shows that the percentage of the correct judgments for /s/ and /z/ production.

As for /z/ production, the correct percentage of subject #1 is 94%, subject #2, 100%, subject #3, 94%, and subject #4, 98%, respectively. The average score, 96.5% is relatively high.

In contrast, the correct percentage for /s/ production is 82% for subject #1, 78% for subject #2, 84% for subject #3, and 66% for subject #4. Thus, the average score, 77.5% for /s/ production is lower than for /z/ production.

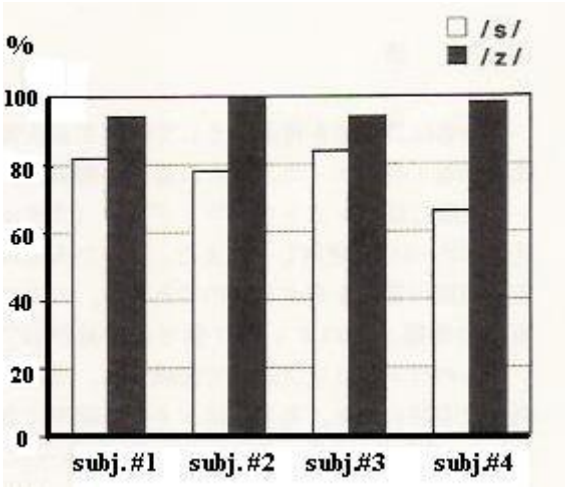

Fig.1 Percentage of correct judgments for /s/ and /z/ production during whispered speech.

These data show that listeners are able to differentiate the voiced and voiceless cognates with substantial accuracy even under the condition that the vocal folds do not vibrate

during whispered speech. Furthermore, misjudgments may occur, in case that the voiceless /s/ production is labeled by native listeners somehow as voiced /z/ under the whispered condition. In contrast, the voiced /z/ production could seldom be misjudged as voiceless.

3.2 Palato-lingual contact patterns

Among the obtained data with regard to palato-lingual contact patterns, 4 samples of /s/ production, misjudged by listeners, were excluded for the averaging. Fig.2 shows the averaged temporal patterns of palato-lingual contact area during /s/ and /z/ production in whispered speech, respectively.

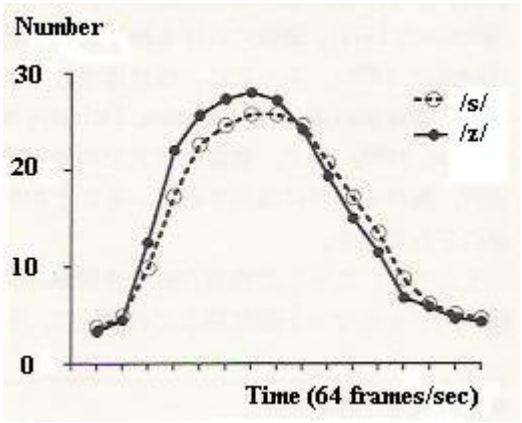

Fig.2 Averaged temporal patterns of palato-lingual contact area during /s/ and /z/ production in whispered speech

The vertical axis corresponds to the number of on-electrodes, which mean that the portions on the tongue contact with artificial hard palate; the horizontal axis represents the time course of ongoing speech production.

The overall shapes of these two curves resemble each other. The initiation of contact, followed by gradual increase of contact area, will be replaced by the cessation of contact after an approximately same interval.

The peak values of the maximum contact area, however, are different; higher for voiced /z/ than for voiceless /s/. Thus, the area of palato-lingual contact is wider for /z/ production than for /s/ production.

In order to clarify how the difference of maximum contact area between /s/ and /z/ production is attained, the most lateral front points and the second lateral front points among the 63 palato-lingual contact points were selected to draw Fig.3 and Fig.4. Here, "the front" points mean the points from the front to the 4th row.

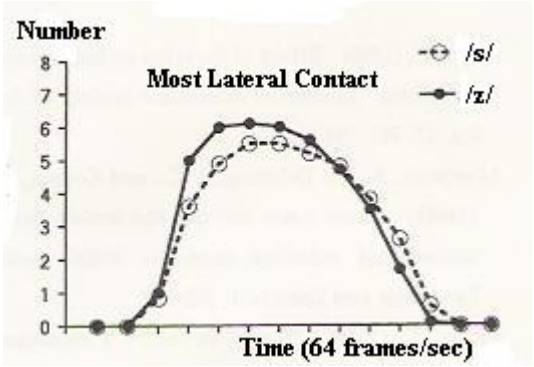

Fig.3 Averaged temporal patterns of most lateral front contact area during /s/ and /z/ production

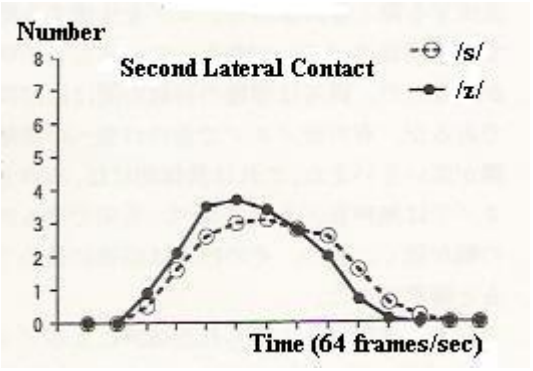

Fig.4 Averaged temporal patterns of second lateral front contact area during /s/ and /z/ production

These figures demonstrate that the difference of peak maximum contact area between /s/ and /z/ production, illustrated in Fig.2, is due to the most lateral and second lateral points. In other words, as for /z/ production, the tongue groove, shaped by the apical tongue movement, is narrower and longer than for /s/.

Incidentally, observing the palato-lingual contact patterns for the /s/ production more carefully, the peak values of contact area, specifically misjudged as voiced /z/ , are all narrow and wide.

3.3 Palato-lingual contact stability

The stability of the palato-lingual contact patters for /s/ and /z/ production in whispered speech was examined. In the middle columns of Fig.5, the palatogram at the maximum contact timing is illustrated, together with the preceding and proceeding frames.

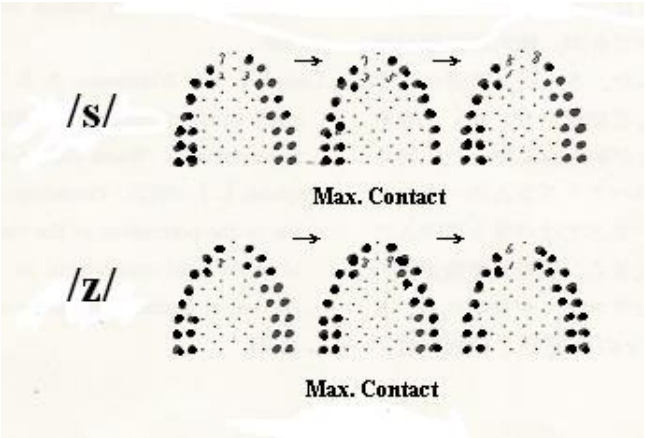

Fig.5 Frequency of contact at each point

Filled circle: always contacted

Dotted point: never contacted

Numerical point: such times out of 10 trials

Each filled circle corresponds to the point, where all the samples demonstrated its contact. On the other hand, the point, written in number, means that its contact may occur such times out of 10 trials.

According to Fig. 5, numerical points for /s/ production are numerous in comparison with those for /z/ production. Therefore, it may be concluded that the tongue articulatory movement for /s/ production is rather unstable. Along this

line, the misjudgments for whispered /s/ production may be attributed to this instability.

4 Discussion

The present study is to further investigate the role of tongue articulatory movement with regard to voiced and voiceless distinctions, specifically when the vocal folds do not vibrate in whispered speech. The palato-lingual contact patterns were recorded using dynamic palatography during Japanese fricative consonants /s/ and /z/ production.

The results clearly demonstrate that the temporal palato-lingual contact patterns are substantially different between /s/ and /z/. The maximum contact area is wider for voiced /z/ than for voiceless /s/, although the duration of the frication noise production is rather similar. In other words, the tongue groove, shaped by the apical portion, is narrower and longer for voiced /z/ production than for voiceless /s/ production, at least in whispered speech.

The perceptual experiments using whispered speech materials were also conducted. Under such circumstances as whispered speech, the listeners are able to label voiced distinction with substantial accuracy.

In addition, when misjudgments do occur with respect to voicing distinction, the speech materials are frequently voiceless. Voiced /z/ production is seldom misjudged, despite lack of vocal fold vibrations in whispered speech..

These acoustic data may be accounted for by the fact that the palatogram for /s/ production is unstable in comparison with /z/ production

The author believes that the vocal fold vibrations play the essential role in performing voicing distinction in natural speech. The current data dealing with whispered speech, in which the vocal folds do not vibrate, may reveal that exaggerated articulatory movement does occur.

Acknowledgments

The original data of this experiment were collected in collaboration with Shinobu Murase (Gifu University, Gifu, Japan). Her special contribution to this study is highly acknowledged.

References

[1] L. Lisker, A. S. Abramson, “Cross-language study of voicing in initial stop: Acoustical measurements”, *Word* 20, 384-422 (1964)

[2] P. Denes, "Effect of duration on the perception of voicing”, *J. Acoust. Soc. Am.* 27, 761-764 (1955)

[3] L. Lisker, "Rapid vs rabit: A catalogue of acoustic features that may cue the distinction”, *Haskins Laboratories Status Report* SR-65, 251-259 (1978)

[4] A. M. Liberman, P. C. Delattre, F. S. Cooper, "Some rules for the distinction between voiced and voiceless stops in initial position” , *Language and Speech* 1, 153-167 (1958)