LATERAL RELEASE IN THE ARTICULATORY CLOSURE OF THE VELAR STOP SOUNDS [k, g]: AN ACOUSTIC STUDY

ZYUN'ICI SIMADA — Dept. of Physiology, School of Medicine, Kitasato University

MINAKO KOIKE — Dept. of Rehabilitation, Kitasato University, East Hospital

AKIKO OKAMOTO — Dept. of ENT, Kitasato University Hospital

SEIJI NIIMI — Research Institute of Logopedics and Phoniatrics, University of Tokyo

ABSTRACT

Lateral lisps involving the velar stop sounds [k, g] were examined acoustically. Lateralized and non-lateralized tokens consisting of the mora [ki], [ke] or [gi] were recorded from the same two adult females who had previously shown distorted pronunciations, but had been corrected to normal articulatory gestures with speech training. It was found that the r.m.s. amplitude of the burst of lateralized tokens was larger than that of non-lateralized tokens. In spectra, a local peak was noted consistently in the vicinity of 5.0 kHz for the lateralized tokens [ki] across both subjects.

INTRODUCTION

In speech clinics, therapists have noted that consonants such as [s], [k] and [tʃ] sometimes sound deviant before front vowels. This sound distortion has been observed typically among children with cleft palate, and is often referred to as a lateral lisp or, more specifically, as being lateralized. It is partly the turbulent outflow of air from the lateral paths of the consonantal constriction that causes the acoustic distortion detectable to speech therapists. However, the acoustic properties corresponding to lateral lisps are less clear, except in a very few reports [1]. In this study, we will focus on the acoustic differences between the normal and the lateral release of the closure of the velar stops [k, g] produced by speakers having previously shown lateral lisping in spite of no organic defects of the speech organs.

Theoretical considerations permit us to expect experimental results as follows. When a consonantal stoppage is released at the lateral side(s) of the tongue contact, the air needs to flow out through the narrow, constricted path(s). This probably results in turbulence noises with larger amplitudes compared to more open paths if we assume the air volume flowing into the constriction to be constant. Moreover, voicing may be delayed, since it takes a long time for the articulatory stoppage to achieve a complete opening. Accordingly, we will employ two measures to characterize velar stop bursts, namely, their duration and root-mean-square (r.m.s.) amplitude. Finally, a linear predictive coding (LPC) analysis will estimate spectral properties for the bursts due to distorted articulatory configurations.

METHODS

Subjects

Subjects were two adult females in their early twenties, S.A. and T.M., who had no apparent pathological defects of the speech organs. In their childhood, they had displayed the deviant type of articulation outlined in the introduction, but had been corrected to normal articulatory gestures with speech training. Subject S.A. was from Kobe, a port in the Kansai area of the Japanese mainland where a dialect different from Tokyo Japanese is spoken. Subject T.M. grew up in Sagamihara, a city on the outskirts of Tokyo.

Test Moras and Procedure

In both subjects' speech, distorted pronunciations had been salient in particular whenever velar stops occurred before the front vowel [i] or [e] and in palatalized moras like [kja]. The flap [r] of subject S.A. had also been deviant in the same phonetic environments. A total of 38 moras were chosen as test moras. These were the non-lateralized moras [ka, ko, ke, ku, ki, kja, kjo, kju, ga, go, ge, gu, gi, gja, gjo, gju, ra, ri, ru, re, ro, rja, rjo, rju] and the lateralized moras [Ke, Ki, Kja, Kjo, Kju, Ge, Gi, Gja, Gjo, Gju, Ri, Rja, Rjo, Rju] (for simplicity, lateralized moras will be represented by capital letters). The lateralized moras [Ke, Ge] and those containing the flap [r] were eliminated from the productions of subject T.M., since these were never lateralized.

The test moras were read in a sound-proof room and recorded with an audio tape recorder standing outside it and running at a speed of 38 cm/s. Both subjects were told to produce three repetitions of every mora in isolation with a short pause between them. Recordings were made of the non-lateralized series of tokens in the earlier stage of the recording sessions, and the lateralized series in the later stage. In the present study, we chose the moras [ki, ke, gi, Ki, Ke, Gi] of subject S.A. and [ki, Ki] of subject T.M. from the samples obtained in the recordings. Among the voiced tokens, only one [gi] and two samples of [Gi] could be analyzed, since the remaining tokens showed voicing simultaneously with the burst.

Acoustic Segmentation

The tokens of interest were transferred to a Hewlett-Packard 1000 computer for acoustic analysis. In digitization, the audio signals recorded

were passed through a low-pass filter, with roll-off beginning at 8.5 kHz, and then sampled with a quantization level of 12 bits and at an interval of 50 μs. The waveform of every token was displayed on a graphics display terminal for the visual demarcation of the acoustic segments which were necessary in the acoustic analysis. The boundaries were determined by taking the coordinate values of the waveform in the vicinity of the point of interest with the aid of a movable cursor.

RESULTS

Figure 1 compares the waveforms of the non-lateralized token [ke] and the lateralized token [Ke] as produced by subject S.A. By inspection, we specified three landmarks on each waveform. The first was the instant of stop release--the point at which the signal exceeded the noise level of the baseline. The second was the onset of voicing in the vowel--the point at which the first glottal pulse emerged as a periodic pitch. Finally, the beginning of every pitch period that followed the first pulse was defined as the point at which the waveform curve crossed the baseline in an upward direction. The phrase "consonantal burst" in this study refers to the interval between the first and the second point.

Duration of the Burst

As is apparent in Figure 1, lateralized tokens tended to have a longer burst duration than non-lateralized tokens. The lateralized class had an average duration of 37.2 ms for [Ki], 58.9 ms for [Ke] and 13.6 ms for [Gi] in subject S.A., and 51.8 ms for [Ki] in subject T.M. The non-lateralized class had an average duration of 35.1 ms for [ki], 31.6 ms for [ke] and 30.3 ms for [gi] in S.A., and 44.0 ms for [ki] in T.M. Here, it is noteworthy that subject S.A. showed a shorter duration for the lateralized voiced tokens [Gi]. (One sample of [ki] and one sample of [Ke] produced by S.A. were eliminated from the averaging of the durations because of large deviations.)

R.M.S. Amplitude of the Burst

Figure 1 also illustrates the larger amplitudes of the lateralized token [Ke]. What time course did the r.m.s. amplitude take during the tokens? To answer this question, we made preliminary measurements of the short-term r.m.s. amplitudes. Figure 2 shows the curves obtained for the samples of Figure 1 over a period of 200 ms from the onset of speech, where a rectangular window was moved successively in steps of 5 ms during the burst, and placed pitch-synchronously over every glottal pulse during the vowel. As can be seen in Figure 2, the lateralized [Ke] had a temporal course at a higher level of the r.m.s. amplitude during the burst. Thus, again, we first measured the r.m.s. amplitude for every token of the same mora over the total length of the burst, and we then computed a ratio by dividing the average value of the lateralized tokens by the average value of the non-lateralized tokens. The ratios derived from subject S.A. were 2.42 (3.8 dB) for [Ki/ki], 1.68 (2.3 dB) for [Ke/ke] and 2.62 (4.2 dB) for [Gi/gi], and 2.66 (4.2 dB) for [Ki/ki] from subject T.M.

Above, we pointed out that the lateralized tokens had a larger r.m.s. amplitude value in the burst than the non-lateralized tokens. However, we could not directly evaluate the amplitude of the burst, since any change in the subjects' overall loudness would have had an influence on the absolute level. Accordingly, in order to evaluate the r.m.s. amplitude of the burst relative to the vowel portion, we adopted a slightly modified version of the energy ratio proposed by Jongman et al. [2]. Incidentally, in examining Figure 2 closer, we can note that the curves there show a rapid rise in amplitude and slow down some five glottal pulses from the onset of voicing, though the r.m.s. amplitude varies at a higher level for the token [Ke] than for the token [ke]. Therefore, we decided to average the amplitudes at the onset of the vowel over the period of its four complete pitches. The first glottal pulse was further added to the averaging step whenever it did not form a complete period. The size of the window for the vowel varied from 12.3 to 14.9 ms for S.A., and from 13.6 to 16.0 ms for T.M. In the last step, a ratio was computed between the r.m.s. value at the onset of the vowel and the r.m.s. value at the consonantal burst. The bigger the amplitude of the burst, the smaller the ratio value.

To sum up the energy ratios, we obtained the following average values for the lateralized tokens: 2.90 (4.6 dB) for [Ki], 4.10 (6.1 dB) for [Ke] and 1.69 (2.3 dB) for [Gi] from S.A., and 1.25 (1.0 dB) for [Ki] from T.M. For the non-lateralized tokens: 5.50 (7.4 dB) for [ki], 5.40 (7.3 dB) for [ke] and 2.90 (4.6 dB) for [gi] from subject S.A., and 4.63 (6.7 dB) for [ki] from subject T.M.

Spectral Properties

In the previous sections, we argued that the lateralized tokens were characterized by a relatively larger r.m.s. amplitude. If this property were ascribed to the peculiar configuration of the tongue, then we would expect, for example, divergences in the location of local peaks to be detectable in their spectral envelopes. A variety of research efforts so far have provided evidence that spectral peaks of short-term acoustic representation are important for the phonemic distinction of stop sounds [3, 4]. It is natural that we should want to take enough care in accurately estimating the spectral shapes of non-stationary waveforms, for instance, stop releases. Thus, in this report, we used an LPC analysis similar to the method of Stevens and Blumstein, despite its some shortcomings [3]. The mean value was first calculated for a truncated waveform and subtracted from the original signal. Then, the signal, after being multiplied by a half-Hamming window without high-frequency pre-emphasis, was submitted to a 24-term LPC computation using the autocorrelation method. The window was 25 ms long and placed at the start of the burst.

Samples of the spectra of the [ke] and [Ke] tokens, which correspond to the waveforms of Figure 1, are shown in Figure 3. In examining these spectra, we can note that both tokens have concentrations of energy in the regions below 0.5 kHz and from the mid-frequency 2.0 to 3.0 kHz. However, the component in the mid-frequency region was not very regular for the tokens [ke]; the concentration of energy was high and flat for one

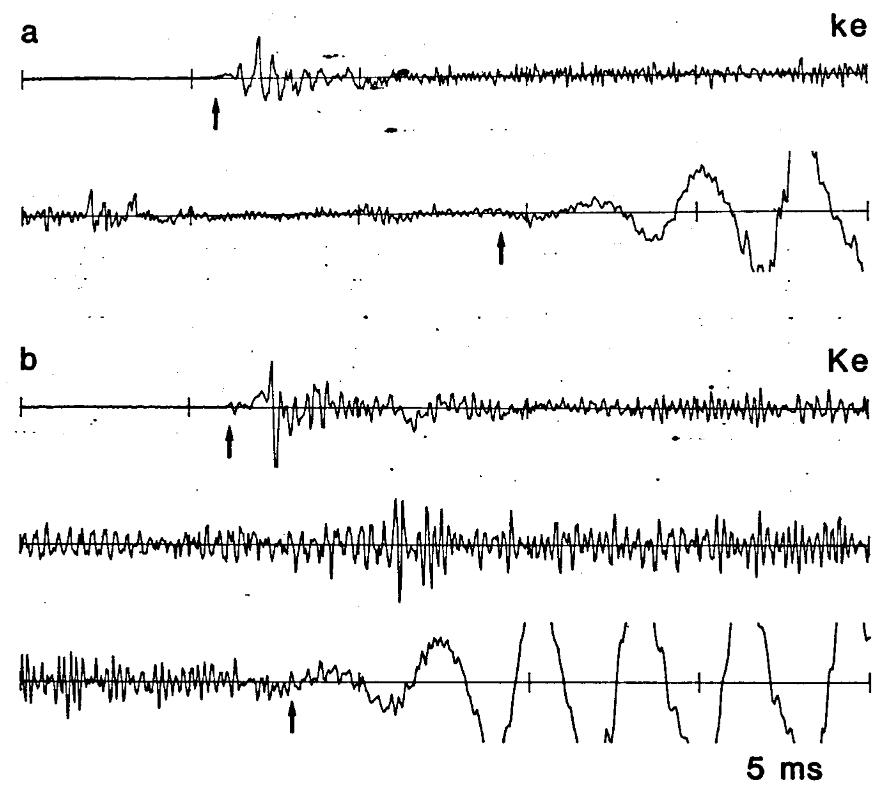

Figure 1. Waveform displays for the non-lateralized mora [ke] and lateralized mora [Ke] as produced by subject S.A. The arrows indicate the instant of the stop release and the onset of voicing in the vowel, respectively.

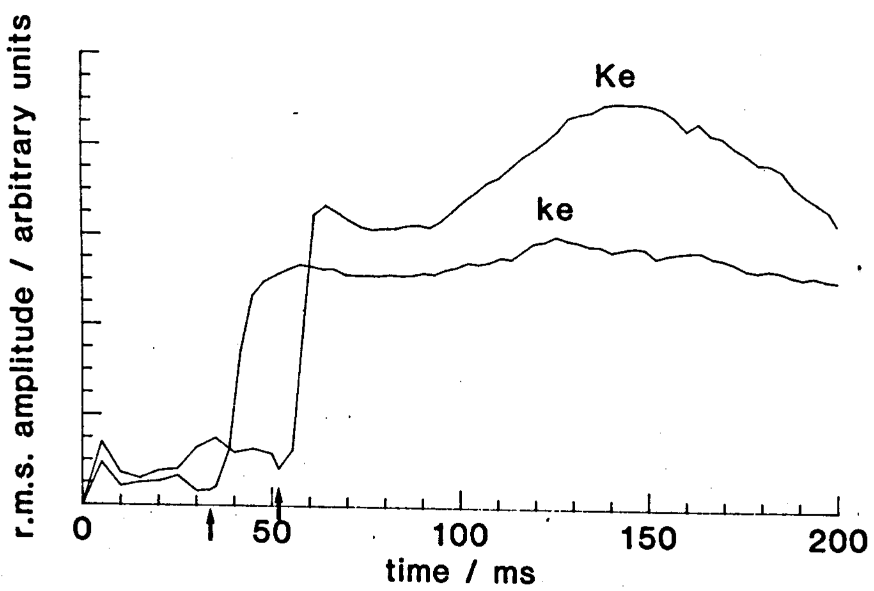

Figure 2. Time tracings for the short-term r.m.s. amplitudes corresponding to the samples in Figure 1. The arrows point to the onset of voicing in the vowels.

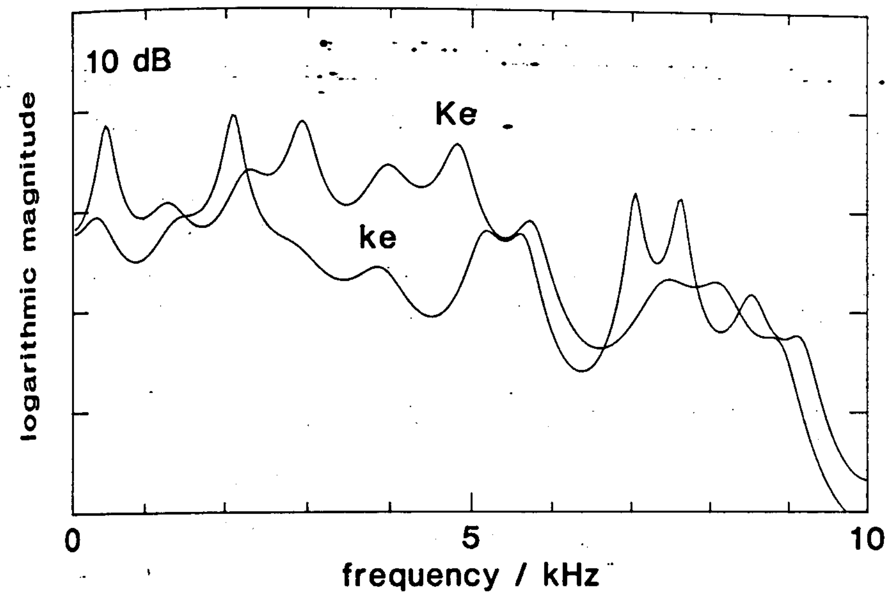

Figure 3. Spectra for the bursts obtained in the same samples as in Figure 1. A 25-ms half-Hamming window was used for the analysis.

sample, and had two separate peaks between 2.0 and 3.0 kHz in the other sample. On the other hand, for the lateralized [Ke] tokens, a single prominent peak was found consistently in the vicinity of 3.0 kHz. The non-lateralized [ke] tokens showed a significant drop in level in the region of 6.0 kHz.

In the spectra of [ki] and [Ki] from S.A., two local peaks were prominent in the vicinities of 3.0 kHz and 5.0 kHz. In particular, peaks at 5.0 kHz for the tokens [Ki] occurred rather consistently and showed as almost equal an amplitude as those at 3.0 kHz. In contrast, the corresponding peaks were irregular in shape and level for the non-lateralized tokens [ki].

In the spectra of the [gi] and [Gi] tokens, some samples showed a higher level between the first peak below 0.5 kHz and the second peak around 2.9 kHz. The spectra of [Gi] were similar in their envelopes to those of [Ki] in the range from 0 to 5.0 kHz.

For subject T.M., on the other hand, the spectra obtained showed consistent patterns in their envelopes for both classes of tokens. In the spectra of the non-lateralized [ki] tokens, a prominent peak was consistently found between 3.0 and 3.5 kHz. In the spectra of the lateralized [Ki] tokens, a peak was also seen in the same region, but it gradually shifted to a higher, sharper peak emerging in the region of 5.0 kHz. This was in contrast to the [ki] tokens, for which the spectra were rather flat in this region.

DISCUSSION

It was the aim of this study to examine how well the acoustic characteristics that seem to play a phonemically contrastive role distinguish such allophonic processes as non-lateralized versus lateralized articulatory gestures. According to Jongman et al. [2], the energy ratio of the burst is kept appreciably contrastive for alveolar and dental stops even in English, a language that has only alveolar stops, or Dutch, a language that has only dental stops. Our results also suggest that this metric of energy ratio is effective for separating non-lateralized and lateralized sounds. The ratio value of four, which indicates a difference of about 6.0 dB, seems to be adequate for this separation.

Finally, does the local peak found consistently around 5.0 kHz for the lateralized [Ki] tokens across both subjects contribute to the perceptual impression characteristic of these tokens? This question is interesting, since the role that local peaks play in the perception of phonetic variants of stops is not clear, and it needs to be resolved by perceptual experiments and other studies.

ACKNOWLEDGEMENTS

We are grateful to Y. Takeuchi for help with the audio recordings, to N. Takizawa for discussion on the issues of acoustic analysis, and to T. Katakura for technical assistance.

REFERENCES

[1] K. Takahashi, K. Michi, H. Hamada and T. Miura, "Acoustic characteristics of Japanese lateral misarticulation," Transactions of the Committee on Speech Research, The Acoustical Society of Japan, S85-92, pp. 719-725, 1986.

[2] A. Jongman, S. E. Blumstein and A. Lahiri, "Acoustic properties for dental and alveolar stop consonants: a cross-language study," J. Phonet., vol. 13, pp. 235-251, 1985.

[3] K. N. Stevens and S. E. Blumstein, "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am., vol. 64, pp. 1358-1368, 1978.

[4] D. Kewley-Port, "Time-varying features as correlates of place and articulation in stop consonants," J. Acoust. Soc. Am., vol. 73, pp. 322-335, 1983.