Weighting of acoustic cues shifts to frication duration in identification of fricatives/affricates when auditory properties are degraded due to aging

Keiichi Yasu[1], Takayuki Arai[1], Kei Kobayashi[2], Mitsuko Shindo[3]

[1]Faculty of Science and Technology, Sophia University, Japan
[2]Faculty of Medical and Health Sciences, The University of Auckland, New Zealand
[3]Faculty of Foreign Studies, Sophia University, Japan
k-yasu@sophia.ac.jp, arai@sophia.ac.jp

Abstract

In previous studies, we conducted several experiments, including identification tests for young and elderly listeners using /shi/–/chi/ (CV) and /ishi/–/ichi/ (VCV) continua. For the CV stimuli, confusion of /shi/ as /chi/ increased when the frication had a long rise time, and /chi/ was confused with /shi/ when the frication had a short rise time. This was true for the group with the following auditory property degradation: 1) elevation of absolute threshold, 2) presence of loudness recruitment, and 3) deficit of auditory temporal resolution. When auditory property degradation was observed, the weighting of acoustic cues shifted to frication duration rather than the gradient of the amplitude of frication. The latter was calculated by dividing frication amplitude by rise time. In the VCV stimuli, confusion of /ichi/ as /ishi/ occurred for a long silent interval between the first V and C with auditory property degradation, and the weighting of acoustic cues shifted from the silent interval to frication duration. In the present study, we unified these findings into a single framework and found that degradation of auditory properties causes listeners to prefer duration of frication as a cue for identifying fricatives and affricates.

Index Terms: elderly listeners, hearing impairments, aging, weighting shift of acoustic cues, trading relations, speech perception, fricatives/affricates

1. Introduction

Elderly people often complain that they struggle with consonant identification when listening to spoken language. The degradation of auditory properties, including (a) an elevation in the threshold of hearing levels [1, 2], (b) a deficit of temporal resolution [3, 4], (c) loudness recruitment [5, 6], and so on, is one of the causes of the confusion in the consonant identification.

There are many cues used for consonant identification: temporal cues [7], spectral cues [8], voice onset time (VOT) [9], formant transition [10], and other cues. It is reported that temporal cues are especially important for identification of fricatives and affricates. Several studies have found that the duration of frication is the primary cue for fricative and affricate identification [11, 12, 13, 14, 15]. When the duration of frication is longer, identification shifts to fricatives from affricates. When preceded by a vowel V_1, the interval of silence (SI) between V_1 and a fricative become a cue for identification [15, 16, 17, 18]. Longer SI tends toward the perception of an affricate.

If there are multiple cues for identification of a consonant, cue-trading may occur. Repp *et al.* conducted an identification of the shop–chop continuum when stimuli were preceded by a vowel [16]. As described earlier, there are two cues used to discriminate between fricatives and affricates: frication duration and the interval of silence between a preceding vowel and fricative. Repp *et al.* identified that listeners perceived the stimuli as shop when the consonant was long and the silent interval was short, and they perceived the stimuli as chop when the consonant was short and the silent interval was long [16]. When the silent interval and the consonant were both short, the identification was affected by weight of two acoustic cues. This is a cue-trading relation between frication duration and silent interval. Erickson *et al.* conducted identification tests on the slit–split continuum by expanding the silent interval between s and l for younger listeners. The results demonstrated that the interval of silence as well as spectral differences were cues for identification, and there was a trading relation between the two cues [20]. In other trading patterns, Best *et al.* found that trading occurs when a silent interval exists between s and a for the say–stay continuum [19] and indicated a trading relation between silent interval and Fl (first formant) onset frequency as cues for the say–stay distinction. Nittrouer and Studdert-Kennedy conducted identification tests of [s] and [ʃ] with children and adults [21]. Results showed that weighting shifts from formant transition to the spectrum of the fricative according to development. Furthermore, Hirai *et al.* conducted an experiment concerning children with a developmental articulation disorder and a language learning impairment and found that weighting shifts to formant transitions with these disorders [22].

In this paper, the relationship between the degradation of auditory properties and the weighting of acoustic cues in identification of affricates and fricatives is discussed. Three auditory properties of elderly listeners were measured: 1) the elevation of the absolute threshold by pure tones and fricatives, 2) a deficit of auditory temporal resolution, and 3) the presence of loudness recruitment by loudness scaling [23]. CV stimuli (with word initial frication) and VCV stimuli (with medial frication) are used for identification tests. We investigate how a degradation of auditory properties causes listeners to shift from reliance on one set of auditory cues to another.

2. Identification tests by young and elderly listeners

The identification tests adopted in this paper were done in our previous studies [15, 24, 25, 26]. In this section, we summarize the two experiments for CV and VCV stimuli, further analyzed those results, and discussed on the directions where the weighting of acoustic cues shifted.

2.1. Participants

Nineteen younger listeners, described as Y, participated in the experiments. All were native speakers of Japanese, aged 21 to 24 years old, with an average age of 22.3. None of the listeners have ever had hearing problems.

Fifty-five elderly listeners participated in the experiments. All were native speakers of Japanese, aged 62 to 83 years old, with an average age of 72.2. They are all from Chiyoda City, Tokyo. None of them have ever had a history of wearing hearing aids. We measured three auditory properties: the deficit of auditory temporal resolution (subscript t); the elevation of the absolute threshold (subscript f); the presence of loudness recruitment (subscript r). From our measurements, we categorized the elderly listeners into five groups: E (25 people), E_t (six people), E_f (seven people), E_{ft} (ten people), and E_{ftr} (ten people).

2.2. Stimuli

We define the duration of rising portions of fricatives as R, steady portions as S, falling portions as F, and total duration of fricatives to $T = R + S + F$. The CV continuum stimuli were made by making alternations to R (rising portion) and S (steady-state portion). Consonants were followed by the vowel [i]. R was set to 0–90 ms, S was set to 0–180 ms in 10 ms steps, respectively. In addition, some of the combinations of R and S were excluded to meet the following condition: $R + S \leq 180$. F was fixed to 22 ms. The amplitude of C was set in two different conditions, C^1 and C^{10}, for ten times the amplitude. There were 292 stimuli.

The VCV stimuli were made by changing the interval of silence (SI) between the preceding vowel V_1 and the consonant C. There were three lengths for C: The short length was R = 20 ms, S = 0 ms, C_{short}. The mid-length C was R = 40 ms, S = 80 ms, C_{middle}. The long C was R = 60 ms, S = 120 ms, C_{long}. The silent interval SI between V_1 and C was set to the range of 0–100 ms in 10 ms steps (11 steps). The amplitude of V_1 and C was set to one and ten times larger than the original amplitude. If the amplitude of V_1 was 1, it was indicated as V_1^1, and V_1^{10} for ten times larger. There were four combinations such as $V_1^1C^1$, $V_1^1C^{10}$, $V_1^{10}C^{10}$, and $V_1^{10}C^1$. The total number of stimuli was 132.

2.3. Procedure

All experiments were conducted in a sound-treated room. The 2-AFC (alternative forced choice) tests were used in experiments. Listeners were asked to selected [shi] or [chi] for the CV stimuli, and [ishi] or [ichi] for the VCV stimuli. A total of 292 stimuli for CV stimuli and 132 stimuli for VCV stimuli were used in the experiments.

2.4. Results

Figure 1 shows the results of the identification tests for the CV stimuli in group E [15]. The horizontal axes show frication duration T and the gradient of amplitude of frication (amplitude/ms), which is obtained by dividing amplitude of frication by R. The vertical axis shows the response rate of [ʃ]. The amplitude of the steady portions of fricatives in C^1 stimuli were normalized to 0.1. The normalized amplitude of the C^{10} stimuli are 1 because the steady portions of the C^{10} stimuli were ten times larger than the C^1 stimuli. The normalized amplitudes were divided by R and placed on the vertical axis in Figure 3. The gradient from 0 to 0.1 corresponds to C^1, and the gradient from 0.1 to 1 corresponds to C^{10}. The condition of R = 0 ms was omitted because its slope is infinite. We defined the response rate of [ʃ] for the CV stimuli as p_1 and fitted with the sigmoidal function

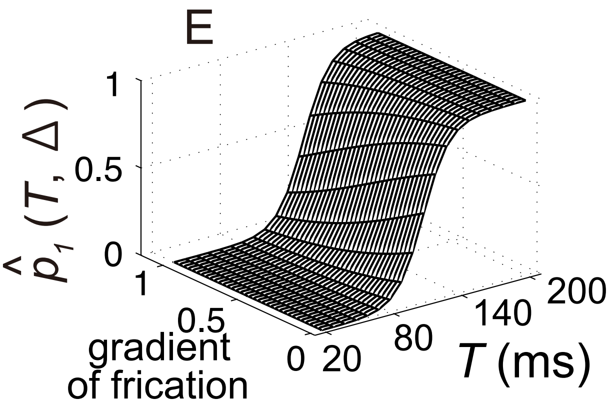

Figure 1: *Results for the identification test for the CV stimuli in group E. The horizontal axes show T and the slope of the amplitude of frication. The vertical axis $\widehat{p_1}(T,\Delta)$ shows the response rate of [ʃ].*

$$\widehat{p_1}(T,\Delta) = f(a_1T + b_1\Delta + k_1) \quad (1)$$

$$f(x) = \frac{1}{1+\exp(-x)} \quad (2)$$

where $f(x)$ is a basic sigmoidal function and Δ is the slope of the amplitude of frication, T is a frication duration, and $\widehat{p_1}(T,\Delta)$ is a function fitted by a sigmoidal function. a_1, b_1, and k_1 are parameters calculated with a minimum MSE (mean square error) between $\widehat{p_1}(T,\Delta)$ and $p_1(T,\Delta)$. When the frication duration is shorter, listeners identified the stimuli as affricates. Longer frication duration tended toward fricative identification.

Figure 2 shows the results of the identification tests for the VCV stimuli in group E [26]. The horizontal axes show SI and frication duration T. The vertical axis shows the response rate of [ʃ]. The response rate of [ʃ] was fitted with a sigmoidal function. As with equation 1, the response rate of [ʃ] was set to p_2 and obtained by the following function:

$$\widehat{p_2}(T,SI) = f(a_2T + b_2SI + k_2) \quad (3)$$

When the silent interval is shorter and frication is longer, listeners identified the stimuli as fricatives. Longer intervals and shorter frication tended toward affricate identification. We obtained the response rate of [ʃ] in the same way as in equations 2 and 1.

3. Discussion

3.1. Cue trading between the slope of R and frication duration in the CV stimuli

Figure 3 shows the phonetic boundary between fricatives and affricates. The horizontal axis shows the frication duration T. The vertical axis shows the gradient of the amplitude of frication (amplitude/ms). When $\widehat{p_1}(T,\Delta) = 0.5$, the phonetic boundary was obtained from equation 1 as a linear function as in the following equation.

$$a_1T + b_1\Delta = -k_1 \quad (4)$$

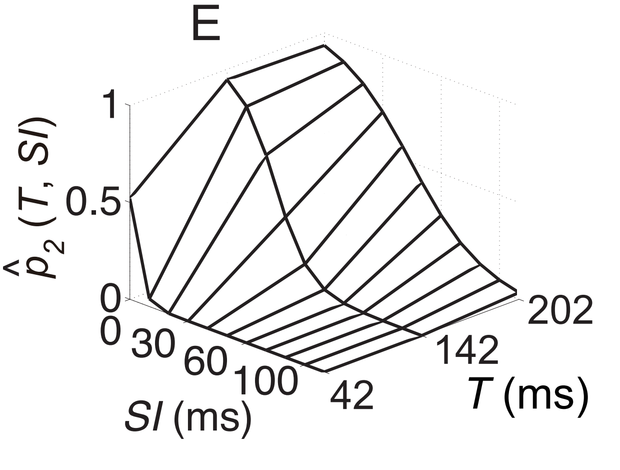

Figure 2: *Results for the identification test for the VCV stimuli in group E. The horizontal axes show SI and frication duration T. The vertical axis $\widehat{p_2}(T, SI)$ shows the response rate of [ʃ] [26].*

Table 1: *Angle θ [deg] at the phonetic boundary in the identification tests for the CV stimuli.*

angle θ [deg] at phonetic boundary					
E	E_t	E_f	E_{ft}	E_{ftr}	Y
90.8	93.0	94.8	88.7	89.5	85.2

Angle θ is obtained by the following equation

$$\theta = \arctan(-a_1/b_1) \quad (5)$$

Table 1 shows the angle of the phonetic boundary line in each participant's group.

The group of young listeners perceived the stimuli as the affricate [ʧ] when frication length T was shorter. They perceived the stimuli as the fricative [ʃ] when T was longer. The perception shifts from [ʃ] to [ʧ] as the slope increases. Those results show a cue trading relationship between frication duration and the slope of the amplitude of frication in younger listeners.

θ in Table 1 indicates the degree of the angle of phonetic boundary line and represents the weighting of acoustic cues. The angle is introduced by Hirai *et al.* [22]. The elderly listeners' θ is larger than that of the younger listeners. Compared with the younger listeners, elderly listeners, who have a degradation in auditory properties, are less influenced by the gradient of the amplitude of frication. These results showed that aging and degradation of auditory properties affect cue trading, in that the duration of the frication was more important than the gradient of the slope of the amplitude for fricative and affricate identification for elderly listeners.

3.2. Cue trading between the silent interval and frication duration in VCV stimuli

Figure 4 shows the relations between the fricative duration T and the interval of silence SI at the phonetic boundary [25]. Figure 4 (a) contains the results for the younger listeners, and Figure 4 (b) contains the results for the elderly listeners. The horizontal axis shows frication duration. The vertical axis shows SI. We obtain the phonetic boundary from equation 1

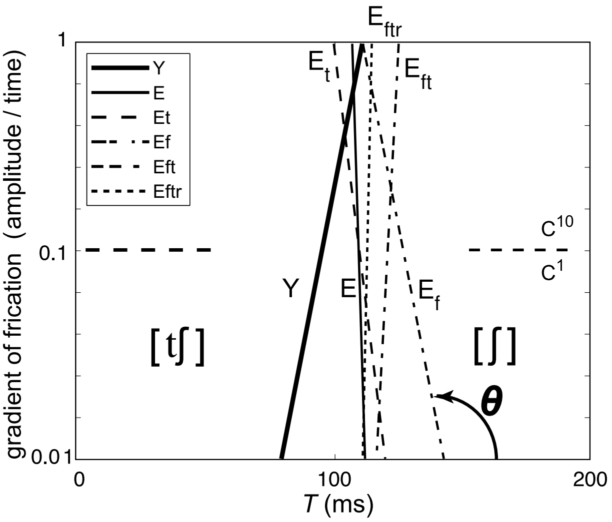

Figure 3: *The relation between frication duration and the gradient of the amplitude of frication at the phonetic boundary. The horizontal axis shows the frication duration T. The vertical axis shows the gradient of the amplitude of the frication (amplitude/ms). θ indicates the degree of the angle of the phonetic boundary line.*

Table 2: *Angle θ [deg] at the phonetic boundary in identification tests for VCV stimuli. The results are arranged in the order of $V_1^1C^1$, $V_1^1C^{10}$, $V_1^{10}C^1$, $V_1^{10}C^{10}$.*

	angle θ [deg] at phonetic boundary						
	E	E_t	E_f	E_{ft}	E_{ftr}	Y	Y'
$V_1^1C^1$	25.3	54.3	42.1	50.7	80.5	30.1	38.6
$V_1^1C^{10}$	10.3	17.8	26.8	18.6	25.8	31.6	45.6
$V_1^{10}C^1$	27.7	45.9	43.6	42.7	68.8	34.6	43.1
$V_1^{10}C^{10}$	18.9	21.3	24.9	16.3	31.3	40.0	45.1

as a linear function.

$$a_2T + b_2SI = -k_2 \quad (6)$$

Angle θ is obtained by following equation:

$$\theta = \arctan(-a_2/b_2) \quad (7)$$

Table 2 shows the angle of the phonetic boundary line in each participant's group.

In Figure 4, all listeners perceived the stimuli as [ʃ] when the consonant C was long and the silent interval SI was short, and they perceived the stimuli as [ʧ] when the consonant C was short and the silent interval SI was long. These two perception are what is expected for fricative/affricate identification. When the silent interval SI and the consonant C were both short, the phonetic boundary exists. Therefore, there is a cue-trading relation between frication duration and SI. This result supports the findings of Repp *et al.* [16] which also describes the cue-trading between frication duration and SI.

In Figure 4 (b), there are also trading relations between frication duration and SI for elderly listeners. In each line of Figure 4 (b), the slope is increasingly steep as we arrange the listeners according to their auditory properties in the following order: E, E_f, E_t, E_{ft}, E_{ftr}. This finding suggests that auditory property degradation affects the perceptual apparatus such

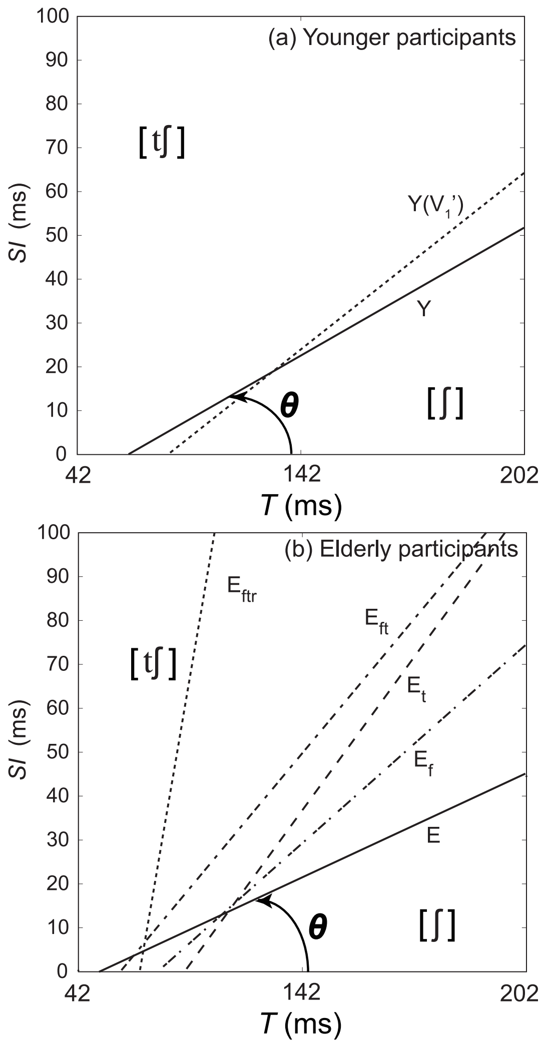

Figure 4: *The phonetic boundary of fricative and affricates was described as a relationship between frication duration T and silent interval SI in $V_1^1C^1$ condition. The horizontal axis shows frication duration. The vertical axis shows SI. (a): younger listeners, (b): elderly listeners, modified from [25]. θ indicates the degree of the angle of the phonetic boundary line.*

that listeners shift from relying on the silent interval for identification, and choose instead frication duration as an auditory cue. In other words, elderly people with auditory degradations tend to clue into frication duration, increasingly as the auditory degradations complete, rather than the silent interval to distinguish between affricates and fricatives.

In Figure 4 (a), in the condition of V_1' (shorter V_1, shortened to 140 ms from 229 ms), the slope of the phonetic boundary line is steeper in elderly listeners than in younger listeners. These results showed that a shorter V_1 leads to cue-trading, where frication duration is preferred as a cue for discrimination over SI. Only the younger listeners participated in the identification test using shorter V_1, if elderly listeners perceived V_1' stimuli, the weighting-shift from SI to friction duration may occur in groups with auditory degradation.

3.3. General discussion

For the CV stimuli, the weighting of cue-trading shifts from the gradient of the amplitude of frication to frication duration for elderly listeners having a degradation in auditory properties. The same is true for the VCV stimuli, where weighting shifts were observed from SI to frication duration. These results showed that auditory degradation affects for weighting shift of cue-trading to frication duration in these two different stimuli, CV and VCV. It is possible to unify these findings into a single framework for identification of fricatives and affricates where different cues are preferred in the presence of degradation of auditory properties due to aging.

4. Conclusions

Fricative and affricate identification tests were conducted in CV and VCV stimuli for young and elderly listeners in our previous studies. By minute investigating in these results, we observed that the angle calculated from phonetic boundary got larger, and the weighting of acoustic cues shifted from the gradient of the amplitude of the frication to frication duration with consonant initial stimuli. In the condition of a longer SI, elderly listeners with the degradation of auditory properties perceived [ʃ], because they clued into frication length rather than the presence of an interval of silence for discrimination. Thus we observed that the weighting of acoustic cues had shifted from SI to frication duration in elderly listeners especially for the listeners with degradation of auditory properties. As a conclusion, it is possible to unify these findings into a single framework for identification of fricatives and affricates where different cues are preferred in the presence of degradation of auditory properties due to aging.

5. Acknowledgements

This research was partly supported by Sophia University Open Research Center from MEXT. We would like to thank the CHIYODA CITY Silver Human Resources Center for their participation. We would like to thank Terri Lander and anonymous reviewers on their helpful comments.

6. References

[1] ISO 226, Acoustics — Normal Equal-Loudness-Level Contours, International Organization for Standardization, Geneva, 2003.

[2] C. Bunch, "Further observations on age variations in auditory acuity," Arch. Otolaryngol Head Neck Surg., 13(2):170–180, 1931.

[3] A. Strouse, D. H. Ashmead, R. N. Ohde and D. W. Grantham, "Temporal processing in the aging auditory system," J. Acoust. Soc. Am., 104(4):2385–2399, 1998.

[4] P. J. Fitzgibbons and S. Gordon-Salant, "Auditory temporal processing in elderly listeners," J. Am. Acad. Audiol., 7(3):183–189, 1996.

[5] E. Villchur, "Simulation of the effect of recruitment on loudness relationships in speech," J. Acoust. Soc. Am., 56(5):1601–1611, 1974.

[6] M. Florentine, S. Buus and T. Poulsen, "Temporal integration of loudness as a function of level," J. Acoust. Soc. Am., 99(3):1633–1644, 1996.

[7] R. V. Shannon, F. Zeng, V. Kamath, J. Wygonski and M. Ekelid, "Speech recognition with primarily temporal cues," Science, 270(5234):303–304, 1995.

[8] S. E. Blumstein, E. Isaacs and J. Mertus, "The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants," J. Acoust. Soc. Am., 72(1):43–50, 1982.

[9] B. H. Repp, "Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants," Language and Speech, 22(2):173–189, 1979.

[10] A. M. Liberman, F. S. Cooper, D. P. Shankweiler and M. Studdert-Kennedy, "Perception of the speech code," Psychological Review, 74(6):431–461, 1967.

[11] P. Howell and S. Rosen, "Production and perception of rise time in the voiceless affricate/fricative distinction," J. Acoust. Soc. Am., 73(3):976–984, 1983.

[12] K. R. Kluender and M. A. Walsh, "Amplitude rise time and the perception of the voiceless affricate/fricative distinction," Percept. Psychophys., 51(4):328–333, 1992.

[13] S. Mitani, T. Kitama and Y. Sato, "Voiceless affricate/fricative distinction by frication duration and amplitude rise slope," J. Acoust. Soc. Am., 120(3):1600–1607, 2006.

[14] S. Amano and K. Yamakawa, "Perception and production boundaries between fricative [s] and affricate [ts] in Japanese," Proc. of ICPhS XVII, Hong Kong, 228–231, 2011.

[15] K. Yasu, T. Arai, K. Kobayashi, M. Shindo, "Identification of voiceless fricatives/affricates by elderly listeners: Effect of degradation of auditory properties," J. Acoust. Soc. Jpn., 68(10):501–512, 2012 (in Japanese).

[16] B. H. Repp, A. M. Liberman, T. Eccardt and D. Pesetsky, "Perceptual integration of acoustic cues for stop, fricative, and affricate manner," J. Exp. Psychol. Hum. Percept. Perform., 4(4):621–637, 1978.

[17] M. F. Dorman, L. J. Raphael and D. Isenberg, "Acoustic cues for a fricative-affricate contrast in word-final position," J. Phonetics, 8:394–405, 1980.

[18] S. Gordon-Salant, G. H. Yeni-Komshian, P. J. Fitzgibbons and J. Barrett, "Age-related differences in identification and discrimination of temporal cues in speech segments," J. Acoust. Soc. Am., 119(4):2455–2466, 2006.

[19] C. Best, B. Morrongiello and R. Robson, "Perceptual equivalence of acoustic cues in speech and nonspeech perception," Percept. Psychophys., 29(3):191–211, 1981.

[20] D. Erickson, H. Fitch, T. Halwes, and A. Liberman, "Trading relation in perception between silence and spectrum," J. Acoust. Soc. Am., 61:46, 1977.

[21] S. Nittrouer and M. Studdert-Kennedy "The role of coarticulatory effects in the perception of fricatives by children and adults," J. Speech Hear. Res., 30:319–329, 1987.

[22] S. Hirai, K. Yasu, T. Arai and K. Iitaka, "Perceptual Weighting of Syllable-Initial Fricatives for Native Japanese Adults and for Children with Persistent Developmental Articulation Disorders," Sophia Linguistica, 53:49–76, 2005.

[23] T. Brand and V. Hohmann, "An adaptive procedure for categorical loudness scaling," J. Acoust. Soc. Am., 112(4):1597–1604, 2002.

[24] K. Yasu, T. Arai, K. Kobayashi, M. Shindo, "Cue trading in fricatives and affricates: Focusing on rise time of frication and preceding silent interval," Autumn Meet. Acoust. Soc. Jpn., 549–552, Sep., 2011 (in Japanese).

[25] K. Yasu, T. Arai, K. Kobayashi, M. Shindo, "Cue trading in the perception of initial fricatives/affricates: Focusing on duration and amplitude of frication," Proc. of the General Meeting of the Phonetic Society of Japan, 109–114, Sep., 2011 (in Japanese).

[26] K. Yasu, T. Arai, K. Kobayashi, M. Shindo, "Identifying specific characteristics of hearing impairments to predict miss-identification of voiceless fricatives/affricates by elderly listeners," Autumn Meet. Acoust. Soc. Jpn., 503–506, Sep., 2012 (in Japanese).