Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

An Evaluation of Bone-conducted Ultrasonic Hearing-aid regarding Transmission of Japanese Prosodic Phonemes

Takayuki Kagomiya (1) and Seiji Nakagawa (1)

(1) Health Research Institute, National Institute of Advanced Industrial Science and Technology (AIST) Japan, 1-8-31, Midorigaoka, Ikeda, Osaka, 563-8577, Japan

PACS: 43.66.Ts, 43.35.Yb, 43.71.Es

ABSTRACT

Human listeners can perceive speech signals from a voice-modulated ultrasonic carrier which is presented through a bone-conduction stimulator (bone-conducted ultrasound, BCU), even if they are sensorineural hearing loss patients. As an application of this phenomenon, we have been developing a bone-conducted ultrasonic hearing aid (BCUHA). The performance of BCUHA has been evaluated considering syllable articulation and word intelligibility. These studies showed that the syllable articulation scores when using BCU were over than 60%, and the word intelligibility scores for words with high familiarity were over than 85%. The patterns of confusion in speech perception in the case of BCU have many points of similarities with those for air-conduction (AC). Although performance of BCUHA regarding perception of segmental units were evaluated, performance of perception of suprasegmental or prosodic units have not been evaluated. Japanese is a pitch accent language. Many words are contrasted by its tonal features, for example, “a’ka” (red) vs. “aka” (dirt). Also Japanese is a mora-timed language. All vowels and some consonants are contrasted by its lengths, like “obasan” (aunt) vs. “oba:san” (grandma), “kita” (came) vs. “kit:a” (cut). Since prosodic units of Japanese function as phonemes as described above, thus the perception of prosodic elements plays important role for Japanese. The purpose of this study is to evaluate the performance of perception of prosodic phoneme through BCUHA. A series of experiments which consist of minimal pairs judgement task was conducted. The minimal pairs were differentiated by prosodic elements. A pair of logatomes “etete” and “ete:te” was selected for long/short vowel discrimination task, “etete” and “etette” were for single/geminate consonant, and a pair of real words “a’ka” and “aka” was for pitch accent. Stimuli were speech sounds continua which manipulated its prosodic element. Experiments were designed as single-stimulus two-alternative forced-choice identification tasks. Ten native Japanese listeners participated in the experiments. To examine whether difference between normal air-conducted hearing (AC) and BCU hearing regarding perception pattern of prosodic elements was observed, the same tasks were conducted in AC and BCU condition. From results of the experiments, the position and the sharpness of categorical boundaries were computed by using logistic regression analyses, and then examined whether significant differences were observed by using t-tests. From analysis of t-tests, significant difference between AC and BCU conditions was not observed in all tasks. This result indicates that the BCUHA can effectively transmit prosodic phoneme as well as segmental element.

INTRODUCTION

For patients with acute sensorineural hearing loss who are not able to hear using a normal hearing aid, we have been developing a bone-conducted ultrasonic hearing aid (BCUHA) (Nakagawa et al. 2006).

Ultrasound is defined as sound waves which travel at such a high frequency that they cannot be heard by humans. However, if the ultrasound is presented through a bone-conducted stimulator (bone-conducted ultrasound, BCU) the ultrasound is perceived by human listeners (Lenhardt et al. 1991). In addition, if BCU signals are amplitude-modulated by speech signals, listeners can perceive the original speech signals Lenhardt et al. (1991). These voice-modulated BCU signals enable patients with acute sensorineural hearing loss perceive speech signals Nakagawa et al. (2006). The BCUHA being developed is based on these observations.

Performance of the BCUHA have been evaluated by using syllable articulation and word intelligibility. Syllable articulation scores were over 60% (Okamoto et al. 2005) and word intelligibility scores for words with high familiarity were over 85% (Nakagawa et al. 2006). The patterns of confusion in speech perception in the case of BCU have many points of similarities with those for air conduction (AC) (Okamoto et al. 2005).

Although the usability of the BCUHA has been evaluated as mentioned above, the evaluations have been restricted to the transmission of segmental information, or textural messages. In other words, little attention has been paid to the transmission of prosody.

For Japanese speaker, perception of suprasegmental, especially lengths of segments and pitch patterns, are very important because those features can be distinctive for Japanese language. For example, the lengths of vowels /a/ contrast in the words /obasan/ (“aunt”) and /oba:san/ (“grandmother”), the lengths of consonants /t/ contrast in /kita/ (“come”) and /kit:a/ (“cut”). Likewise, /ka’ki/ (“oyster”) and /ka‘ki’/ (“persimmon”) are discriminated by their pitch patterns.

The purpose of this study is to evaluate the performance of perceptions of those suprasegmental phoneme via BCUHA. It is well known that the discrimination of all minimal pairs as mentioned above are categorical. Thus, the research questions of the present study are as follows: 1) Are the categorize thresholds for identifications of the minimal pairs same with normal hearing and BCU? 2) Are categorization sharpness of these discriminations decreased? To solve these questions, several series of psychological experiments were conducted.

METHOD

The experiments were identification tasks of each minimal pair which is differentiated by suprasegmental element as mentioned above. To examine whether differences were observed in the identification between AC and BCU conditions, the same tasks were conducted under both conditions.

Stimuli

All experiments were identification tasks of the stimulus continua. The stimulus were speech sounds that manipulated the target segment lengths or pitch pattern from original sounds.

Design of minimal pairs

Short/long vowel discrimination task To examine ability of discrimination between short and long vowels, a minimal pair of non-sense words /etete/ vs. /ete:te/ was selected. The words were set as non-accented words.

The reasons for selecting this minimal pair were that; First, as stated above, the words were set as non-sense words. Sakamoto et al. (2004) reported that the intelligibility of a word is influenced by the familiarity (Amano and Kondo 1999) of the word. i.e. the familiarity of a word is higher, the intelligibility score becomes greater. Therefore, to avoid this influence, non-sense words were selected.

Second, the segmental construction of the words is three vowels /e/ separated by two silence plosives /t/ (Figure 1). Openness of the vowel /e/ is middle. Openness affects sonority of the vowel. Open vowels have high sonority while close vowels have low sonority. Since openness of /e/ is middle, the sonority of /e/ is regarded as middle (Kubozono 1999). Therefore effect of sonority to intelligibility can be small. Moreover, in Tokyo Japanese, a close vowel which is placed between two voiceless consonants tends to be devoiced (Kubozono 1999, McCawley 1968). Since /e/ is a middle vowel, /e/ is not devoiced (Kubozono 1999, McCawley 1968). Furthermore, The target vowel /e/ is placed between two silent plosives /t/, thus manipulation of the vowel can be comparatively easy.

Third, the words were designed to avoid influences of Japanese tonal rules. Maekawa (1997) and Oguma (2000) reported that if there are sharp pitch movements inside a vowel, the vowel perceived as a long vowel. In Tokyo Japanese, a word accent is realized as a sharp pitch fall (Pierrehumbert and Beckman 1988). This is the reason of the words of the minimal pair set as non-accented words. In addition, in Tokyo Japanese, if a word pronounced as isolated word, a sharp pitch raise appears from initial mora to second one unless the initial mora is accented (Pierrehumbert and Beckman 1988). For example, as for Figure 1, pitch of second mora /e/ is much higher than that of first mora. To avoid this effect of initial rising, target vowel /e/ placed in the second mora of three mora word. Moreover, mora in the end of the utterance suffers final lengthening which makes segmental durations at the word final position longer than other ones (Klatt 1975, Takeda et al. 1989). To avoid this effect, the target mora placed in the second from the end.

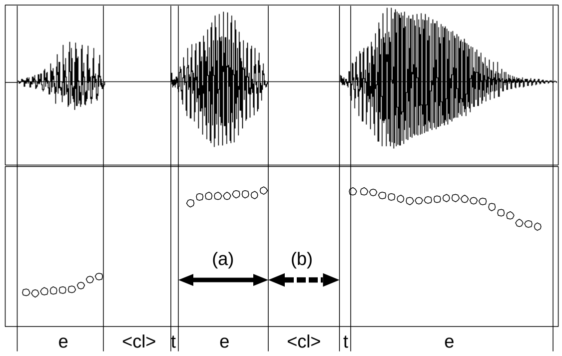

Figure 1: Modification of the target segment of the long vowel discrimination task and the geminate discrimination task. (a) Segment length of /e/ was manipulated for long vowel task. (b) Closure length of /t/ is manipulated for geminate task.

Single/geminate consonant discrimination task For the task of distinction of geminate and singleton consonant, a minimal pair of non-sense words /etete/ vs. /etet:e/ was selected. Most of reasons to select this minimal pair are same as the long vowel task except for the selection of the target consonant. The consonant /t/ is silent plosive, so extending and shortening of the consonant can be relatively easy (Figure 1). Thus, this is the reason to select /t/ as the target consonant.

Pitch accent discrimination task A minimal pair of real words /a'ka/ ("red") vs. /a'ka'/ ("dirt") was selected.

The reasons for selecting this minimal pair were that; First, since it is reported that non-linguistically trained people have great difficulty to point out which mora is accented (Mizutani 1990), thus this task was set to distinction of a minimal pair of real words.

Second, this words consist of two open vowels /a/ and one voiceless plosive /k/. Since open vowel have good sonority, it is suitable for hearing experiments. In addition, two vowels separated by voiceless plosive, thus the manipulation of the pitch of the vowels can be easy (Figure 2).

Third, both words have familiarity scores of more than 5.5 (Amano and Kondo 1999). This value regarded as very high (Sakamoto et al. 2004). Thus, perception of each word is hard to be affected by its familiarity.

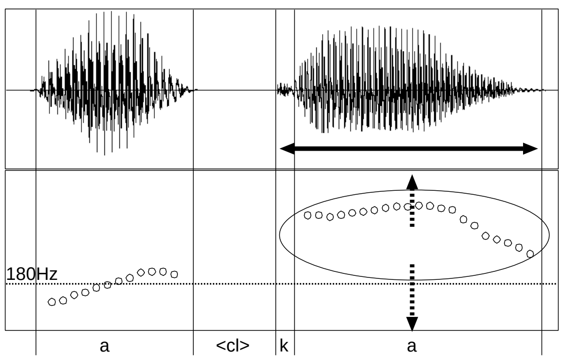

Figure 2: Modification of the stimuli of the pitch accent discrimination task. Original sound was /a'ka'/. Horizontal arrow indicates time range of /ka/. Average fundamental frequency of /ka/ (circled) was ascent or descent across 180Hz (dotted line).

Recording of original speech materials

Speaker A native Japanese male in his thirties whose native dialect is Tokyo (standard) Japanese was participated in sound recordings.

Recording Environment Each word as mentioned above was spoken by the speaker in an anechoic room. The speaker repeated each word 10 times. The speech samples were recorded at 16 kHz / 16 bit resolution and stored in a personal computer.

Sound Manipulation

Short/long vowel discrimination task The stimuli used for the short/long vowel discrimination task were generated in the following manner. An utterance /etete/ which has average length of the target phoneme /e/ in 10 times repeats was selected as the original sound. The length of the target phoneme /e/ in the utterance was 97 msec. The vowel was extended to +200 msec and shortened to -20 msec from the original length by 20 msec step. This range included the minimal length of the vowel in /etete/ and the maximum length in /ete:te/ in each 10 times repeat. STRAIGHT sound synthesizer (Kawahara 2006) was applied for this operation. STRAIGHT is a tool for manipulating voice quality, timbre, speech length and pitch respectively with keeping natural sound quality. Therefore, STRAIGHT was suitable for the sound manipulation of this study. As a sequel of this procedure, total of 13 stimuli were generated.

Single/geminate consonant discriminate task An utterance /etete/ was selected applying similar criteria to that of short/long vowel task. The length of the target closure in the consonant /t/ was 80 msec. This closure length was manipulated from -20 to +200 msec from the original length by 20 msec step. This range was selected also applying the same criterion as that of the short/long vowel discrimination task. Total of 13 stimuli were generated.

Pitch accent discrimination task An utterance /a‘ka’/ which has average f0 values in both vowels /a/ in 10 times repeats was selected as original sound. The average f0 value in the first /a/ was 155 Hz, and in the second /a/ was 200 Hz. For this task, the sound modification was raising or downing of f0 value of /ka/ compared with /a/. Therefor, f0 value of /a/ must be not so high and must be not so low. Thus, average f0 value of /a/ was ascent to 180 Hz. This was about middle f0 value of original /a/ and /ka/. Then, f0 value of /ka/ was shifted within the range of 130 Hz to 230 Hz by 10 Hz step . This range was $\pm$50 Hz across 180Hz (Figure 2). This range also included all 10 times repeats of /a‘ka’/. STRAIGHT sound synthesizer was also used in this operation. As a sequel of this procedure, total of 11 stimuli were generated.

Participants

Ten native speakers of Japanese with no reported speech or hearing defects. Their ages were in the range 22-45 years.

Sound Presentation

The sound stimuli mentioned above were presented through a headphone (Sennheiser HD650) under AC conditions.

The stimuli presented under BCU conditions were 30 kHz ultrasounds that were amplitude-modulated by speech signals. The amplitude modulation method applied in this study was the double sideband-transmitted carrier (DSB-TC) method since previous studies had found this method to be capable of speech modulation for BCU Nakagawa et al. (2006), Okamoto et al. (2005) . With the DSB-TC method, the modulated speech signals $U(t)$ are given by the following expression:

$$U(t) = (S(t) - S_{\min}) \times \sin(2\pi f_c t) \quad (1)$$

where $S(t)$ is the speech signal, $S_{\min}$ is the minimum amplitude of $S(t)$, and f_c is the carrier frequency (30 kHz, Figure 3).

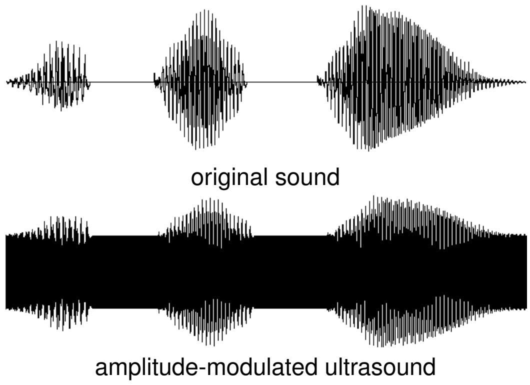

Figure 3: Ultrasonic amplitude-modulation of a stimulus /etete/. Original sound (upper) was transformed into amplitude-modulated ultrasonic carrier (lower) by applying DSB-TC method.

Figure 4: Ceramic vibrator of the BCUHA attached to the mastoid with a hair-band-like device

The stimuli under BCU conditions were presented using a custom-made ceramic vibrator (Figure 4). Bone-conducted ultrasound can be perceived when it is applied to various parts of our body, and the mastoids are among the locations where the perception is high. Therefore, we applied the vibrator to the left or right mastoid of the subject using a hair-band-like supporter (Figure 4).

Procedure

The experiments carried out under both AC and BCU conditions were conducted in a soundproof chamber. Both the presentation of the stimuli and the recording of the responses were executed using a personal computer. Further, the stimuli were provided using a FireWire-based audio interface (Echo Audiofire 12) attached to the personal computer. In both AC and BCU conditions, the sound levels of the stimuli were adjusted to the most comfortable levels for each participant.

The participants were requested to judge and answer that the presented stimulus was which one of the minimal pairs as two-alternative forced choice.

For each participant, two sessions were conducted with more than a week-long interval. The AC condition experiments were administrated at the first session. More than one week after this AC condition session, BCU condition experiments were conducted.

ANALYSIS

Based on the categorical perception theory, the boundary threshold and the sharpness of response were calculated. The proportions of responses could be approximated to logistic function (formula (2)).

$$P(\theta) = \frac{1}{1+\exp(-k(\theta-\theta_c))} \tag{2}$$

Where θ is the controlled value of the stimuli, and $P(\theta)$ is the proportion of response to the stimuli which have the value θ. θ_c is the θ value at $P(\theta)$ is 50%, k is the slope of the logistic curve in θ_c, i.e. max value of the slope (Baker 2001).

All parameters were calculated applying least square method. Although k is the parameter of categorization sharpness, it is difficult to understand immediately how slope value indicate how sharp categorization, thus each θ that $P(\theta)$ becomes 25% and 75% were calculated, the range between this two θ were defined as the parameter of categorization sharpness (Figure 5)

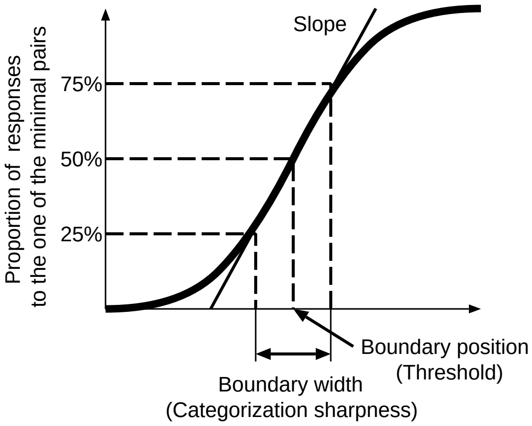

Figure 5: Modeling of categorical identification of minimal pairs by logistic functions.

Proportion of responses to each alternative of minimal pairs were calculated for each participants. Let manipulated segment lengths or pitch heights be θ, let the proportion of responses to the stimulus θ be $P(\theta)$, the parameter of the logistic function above were calculated.

As a sequel of this procedure, thresholds of categorization and sharpness of categorization to each minimal pair by each participant were calculated.

RESULTS AND DISCUSSION

Figure 6 shows the distribution of participants for the threshold and categorization range. Ordinates in graphs of threshold indicate the distance from the original sounds to the categorical boundaries.

In each task, a series of paired t-test revealed that both thresholds and ranges of AC condition and BCU condition did not differ significantly at the 0.05 level. This result suggests that each suprasegmental phoneme can be discriminated even through

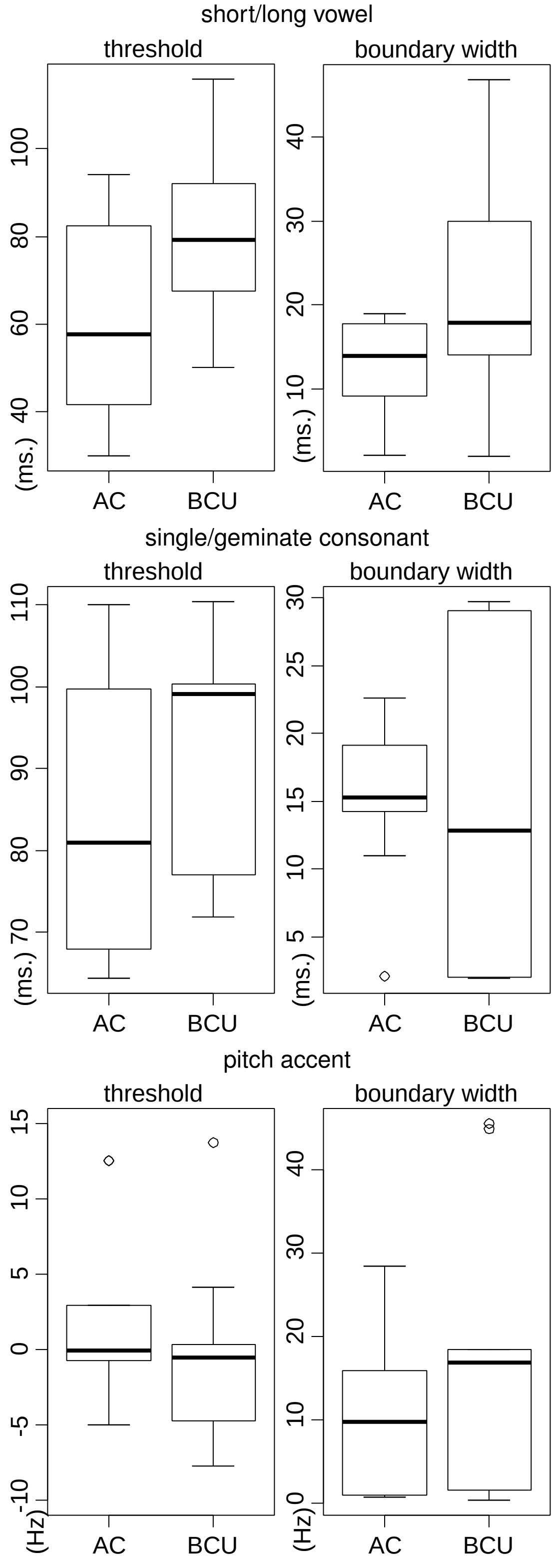

Figure 6: Distribution of participants for threshold and boundary width of categorization for each minimal pair.

BCU.

At the same time, although there were no significant differences, there were tendencies that the thresholds of single/geminate consonant discrimination task and short/long vowel discrimination task shifted toward the later and longer side (short/long vowel discrimination task: $p = 0.074$, single/geminate consonant discrimination task: $p = 0.127$). This tendencies suggest that BCU discrimination of long segment from short one requires more difference between them than AC. What is more, there were tendencies that the individual difference of the ranges became larger in single/geminate consonant discrimination task and short/long vowel discrimination task. To examine that difference of threshold between AC and BCU have any relation with difference of range between AC and BCU, correlation coefficient between the two differences were calculated. The result was revealed that there was high correlation coefficient ($r = 0.745$, $p < 0.05$) in single/geminate discrimination task. This correlation suggests that the participants had less discrimination ability of gemimate in BCU were requires to more time differences between geminate and single stops to recognize the stop is geminate.

CONCLUSION

This research examined the performance of discrimination in suprasegmental phoneme. The following were the main findings; 1) The threshold of discrimination to categorize suprasegmental phoneme in Japanese were not shifted with BCU. 2) Also the sharpness of discrimination were not reduced with BCU. Overall, these findings suggest that BCUHA will serve in hearing the languages even if suprasegmental features were distinctive in the languages.

However, there were individual differences among participants in those discrimination abilities with BCU. To understand from what were this individual differences derived should be studied further.

ACKNOWLEDGMENTS

This research was supported by the 2006FY Industrial Technology Research Grant Program of the New Energy and Industrial Technology Development Organization (NEDO) Japan, 2009FY Grants-in-Aid for Scientific Research of Japan Society for the Promotion of Science, and the 2009FY Research Grant of the Sound Technology Promotion Foundation of Japan.

REFERENCES

Shigeaki Amano and Tadahisa Kondo. *Lexical Properties of Japanese*, volume 1. Sanseido, Tokyo, 1999.

Frank B. Baker. *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD, 2 edition, 2001.

Hideki Kawahara. STRAIGHT, exploration of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. *Acoustic Science and Technology*, 27(6): 349–353, 2006.

Dennis H. Klatt. Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3:129–140, 1975.

Haruo Kubozono. *Nihongo no Onsei [Phonetics of Japanese]*. Iwanami Shoten, Tokyo, 1999.

Martin L. Lenhardt, Ruth Skellett, Peter Wang, and Alex M. Clarke. Human ultrasonic speech perception. *Science*, 253: 82–85, 1991.

Kikuo Maekawa. A layman's view on korean-japanese phonetics. In The National Institute for Japanese Language, editor, *Japanese and Korean : research articles*, pages 173–190. Kuoroshio Publishers, Tokyo, 1997.

James D. McCawley. *The Phonological Component of a Grammar of Japanese*. Mouton, The Hague, 1968.

Osamu Mizutani. Akusento to intoneshon no shutoku-hou [learning method for accent and intonation]. In Miyoko Sugito, editor, *Nihongo no onsei, on'in (Ge) [Japanese Phonetics and Phonology II]*, Kouza Nihongo to Nihongo kyouiku [Japanese and Japanese Education cource]. Meiji Shoin, Tokyo, 1990.

Seiji Nakagawa, Yosuke Okamoto, and Yoh-ichi Fujisaka. Development of a bone-conducted ultrasonic hearing aid for the profoundly sensorineural deaf. *Transactions of the Japanese Society for Medical and Biological Engineering : BME*, 44(1):184–189, 2006.

Rie Oguma. Perception of Japanese long vowels and short vowels by English speaking learners. *Japanese language education around the globe*, 10:43–55, 2000.

Yosuke Okamoto, Seiji Nakagawa, Kiyoshi Fujimoto, and Mitsuo Tonoike. Intelligibility of bone-conducted ultrasonic speech. *Hearing Research*, 208:107–113, 2005.

Janet B. Pierrehumbert and Mary E. Beckman. *Japanese Tone Structure*. The MIT Press, Cambridge, MA and London, 1988.

Shuichi Sakamoto, Naoki Iwaoka, Yôiti Suzuki, Shigeaki Amano, and Tadahisa Kondo. Complementary relationship between familiarity and SNR in word intelligibility test. *Acoustical Science and Technology*, 25(4):290–292, 2004.

Kazuya Takeda, Yoshinori Sagisaka, and Hisao Kuwabara. On sentence-level factors governing segmental duration in Japanese. *The Journal of the Acoustical Society of America*, 86(5):2081–2087, 1989.