DEVELOPMENT OF A JAPANESE SPEAKER DISCRIMINATION TEST FOR EVALUATION OF HEARING ASSISTANCE DEVICES

Takayuki Kagomiya & Seiji Nakagawa

Health Research Institute,
National Institute of Advanced Industrial Science and Technology (AIST), Japan
t-kagomiya@aist.go.jp; s-nakagawa@aist.go.jp

ABSTRACT

For assessment of hearing assistance devices, a speaker discrimination test was developed. The speech material was drawn from a "Japanese phonetically balanced word speech database" which was developed by Electrotechnical Laboratory, Japan (ETL-WD corpus). The test consisted of ten words spoken by ten speakers. The words were selected considering familiarity, and the speakers were selected considering F0 and formant space size. To check the validity of the test, a series of hearing experiments was conducted. The stimuli of the experiments consisted of each sound of the test processed by a 1, 2, 4, 6, 8 and 12 channel cochlear implant simulator. The result showed that speaker discrimination score decreased according to degradation of the speech sounds. This result indicates the test score reflects the performance of hearing assistance devices and can thus be used to assess the devices.

Keywords: speaker discrimination, hearing aid, cochlear implant, F0, formant space size

1. INTRODUCTION

Over the past few decades, several kinds of hearing assistance devices have been developed. Hearingaids have been improved in their quality and portability, bone-conducted hearing-aids have been developed for severe conductive hearing loss patients, and cochlear implants (CI) have been developed for sensorineural hearing impaired patients who cannot perceive speech sounds through these hearing-aids. In addition, the bone-conducted ultrasonic hearing aid has recently been developed [7]. This device is also able to provide partial hearing for severe sensorineural hearing loss patients, but unlike CI, the device does not require a surgical operation [7].

As described above, nowadays hearing-impaired patients are able to choose an appropriate hearing assistance device from a range of options. Thus patients require more information or evaluation of these hearing assistance devices to select the most suitable one.

Evaluation of the performance of these hearing assistance devices has been mainly focused on transmission of speech sounds. Thus, monosyllable or word intelligibility tests have been widely adopted for the evaluation of the devices. In other words, evaluation has been focused on textual messages or linguistic messages. However, besides such linguistic information, speech sounds transmit additional messages. For example, we can recognize and discriminate a speaker's gender, age, emotion etc. even if the textual messages are the same. This is called paralinguistic information, and it enriches oral communication and makes it more expressive than written language. Thus patients want hearing assistance devices to transmit paralinguistic information.

However, as described above, little attention has been paid to the performance of hearing assistance devices with respect to transmission of paralinguistic information in speech signals. Such indifference to transmission of paralinguistic information causes having difficulty to perceive paralinguistic information and discriminate speakers for CI or other hearing assistance device users [6]. Therefore, methods for evaluation of the performance of transmission of such paralinguistic messages are strongly required.

Recently, several studies have begun that are focused on the transmission of these paralinguistic messages. An evaluation method of the hearing aids regarding speakers' intention or attitude was proposed [5]. Also a speaker discrimination test for German CI users was developed [6]. However, a Japanese version of a speaker discrimination test had not been developed. Therefore, we developed a Japanese version of a speaker discrimination test for evaluation of hearing assistance devices. The main concept of the test was based on the German speaker discrimination test [6]; however, we incorporated various attributes to adapt the test to

the Japanese language. In this article, we report on the development process of the test. The test was designed as speaker discrimination task. Examinees were requested to judge if the speakers of two sounds were the 'same' or 'different'.

2. DEVELOPMENT OF THE SPEAKERDISCRIMINATION TEST

2.1. Speech material

The speech material was extracted from the "Japanese phonetically-balanced word speech database" which was developed by Electrotechnical Laboratory, Japan (ETL-WD corpus). The corpus consists of 1542 phonetically balanced real words read by 10 male and 10 female native Japanese speakers. The speech sounds are recorded at 16 bit / 16 kHz sampling quality. Also F0 and F1-F4 formant information data are attached to each speech file.

2.2. Selection of speakers

Five male and five female speakers were selected, considering their F0 and formant space size.

F0 is one of the most salient speaker-specific features. As is well known, the F0 for an adult male voice is low, for a child it is high, and the adult female's is medium. These differences of F0 are derived from differences in the size of the speaker's vocal cords. The F0 value of each speaker was estimated from the F0 data attached to the corpus by calculating the average value of each speaker.

Also, the formant space size reflects the speakers' vocal tract length (VTL). If VTL is long, the formant space size is narrow. Thus, the adult male has a narrow formant space, children have a wide one, and females have a medium-sized one. According to source-filter theory, each formant (F_n) can be predicted using the following equation:

$$F_n = \frac{(2n-1)c}{4l} \quad (1)$$

where c is the speed of sound (34400 cm/sec.), and l is vocal tract length. Vocal tract length can be estimated from formant frequencies using inverse operations of the equation. Thus the vocal tract length was estimated from measured formant frequencies using the following procedure:

1. Vocal tract lengths l were estimated from each F_n using the following formula (reverse operation of equation (1).) :

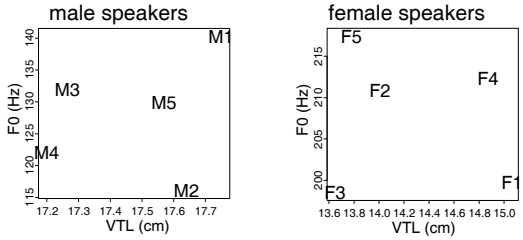
$$l = \frac{(2k+1)c}{4F_{k+1}} \quad (2)$$

where k = (0,1,2), F_{k+1} is the formant frequency of interest.

2. The mean value of l calculated from the three formants of each speech file was regarded as the estimated vocal tract length (VTL).

Considering these F0 and VTL values, 1) high F0 and short VTL speaker, 2) high F0 and long VTL, 3) low F0 and short VTL, 4) low F0 and long VTL, 5) middle F0 and middle VTL were selected respectively from males and females (Fig. 1).

Figure 1: F0 and estimated VTL of the selected speakers.



2.3. Selection of words

From the 1542 words stored in the corpus, 10 words were selected (Table 1) under the scheme described below.

Table 1: List of the selected words.

word ID	transcription	translation	familiarity
W0041	bon-pyaku	various	1.438
W0231	kokubyaku	black and white	1.812
W1533	yuiwata	a traditional woman's hear style	1.938
W0109	gen-un	vertigo	2.125
W1405	ron-kitsu	confute	2.281
W1069	bin-patsu	hairs at the sideburns	2.344
W1484	ten-ita	top board	2.344
W0005	an-utsu	melancholy	2.625
W1518	yaseyama	barren moutain	2.75
W0402	soeuma	carriage horses	2.844

2.3.1. Number of moras

The length of words was fixed. Japanese word length is usually counted in moras. In this research, word length was set at four moras. This number of moras is also applied in "Familiarity-Controlled Word Lists" [2], because four-mora words are the most frequently occurring type in Japanese [2].

2.3.2. Lexical accent

Japanese is a pitch-accented language. In this research, accent type was also fixed. The L-H-H-H type (unaccented type) was selected because this type is most common in Tokyo Japanese [2].

2.3.3. Familiarity

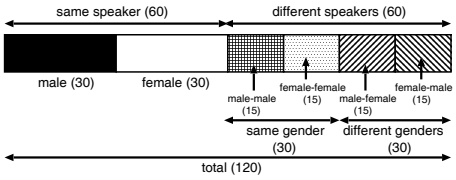
Word familiarity has a correlation with the speaker

identification score [3] therefore, to avoid this effect, low familiarity score words were selected. Familiarity scores was extracted from "Lexical Properties of Japanese" [1].

2.4. Stimuli lists

To control the difficulty and procedure of the test, randomized stimuli lists were generated (Fig. 2). Each list consisted of 120 word pairs. In each list, the number of pairs in the "same-speaker condition" and "different-speaker condition" were balanced (each had 60 pairs).

Figure 2: Component ratio of speaker-combination.



In the different-speaker condition, 30 pairs represented a "both speakers are different-gender condition", and the other 30 pairs represented a "both speakers are same-gender condition".

The same-gender condition was divided into two parts; 15 pairs with a "both speakers are male condition" and 15 pairs with a "both speakers are female condition".

Also the different-gender condition was divided into two parts; 15 pairs of "male-female order" and 15 pairs of "female-male order".

If the stimuli sentence was the same, the speaker discrimination score increased compared with the different sentence condition [4]. To avoid this effect, we ensured that each word pair in the lists contained different words.

2.5. Normalization of amplitude

The sound files were normalized for overall amplitude.

3. VALIDATION OF THE TEST

A series of experiments was conducted to validate whether the test which we developed was suitable for assessment of hearing assistance devices. If the development of our speaker discrimination test is succeeded, the accuracy of speaker discrimination reflects quality of speech sounds. Thus, a series of speech sounds which degraded in quality gradually was the stimuli of the experiments. Furthermore, to estimate the speaker discrimination scores of CI users, degraded speech sounds were created by using a CI simulator (CIsim).

3.1. Cochlear implant simulator

The Tiger CIS (*http://www.tigerspeech.com/tst_tigercis.html*) was adopted as the CIsim. This software converts WAV format sound files into various qualities of CI simulated sound files. To investigate various levels of degraded sound, the number of channels of CIsim were controlled and set to 12, 8, 6, 4, 2, and 1 respectively. The number of channels of CI represents frequency resolution. Because limited number of CI channels, transmition of F0 and formant information by CI is mediocre and CI users have difficulity to discriminate speakers.

In order to assess the performance of transmission of linguistic messages by using the CIsim, a monosyllable intelligibility test was carried out prior to the main experiment. The preparatory experiment was conducted by using the "20 monosyllable intelligibility test" contained in the "67-S Word List" which is one of the standard intelligibility tests used in Japan.

3.2. Participants

Nine native Japanese speakers (5 males and 4 females), with no reported hearing or speaking defects, participated in the experiments. Their ages were in the range 18-41 years.

3.3. Apparatus

Both the presentation of the stimuli and the recordings of the responses were executed using a personal computer. Also, the stimuli were played using a FireWire-based audio interface (Echo Audiofire12) attached to the personal computer. The participants listened to the stimuli on headphones (Sennheiser HDA200) in a soundproof chamber. The sound levels of the stimuli were adjusted to the most comfortable levels for each participant.

4. RESULTS AND DISCUSSIONS

The results of the monosyllable intelligibility tests are shown in Fig. 3. In the 12ch condition, the correct perception score reached 0.879. On the other hand, the intelligibility score (correct perceived ratio) was limited to 0.061 in the 1ch condition. In this way, the scores decreased as a function of the number of channels of CIsim. A repeated one-way ANOVA showed the effect of the number of CIsim channels was significant ($p < 0.001$). A post-hoc test (Tukey's HSD) revealed that there was no significant difference between the 12, 8 and 6 channel conditions. Significant

differences were observed between these three conditions and the other three conditions, i.e. 4, 2, and 1 channel conditions, and between the three conditions. These results indicate that perception of monosyllable is robust against a decrease of CIsim channels while CIsim has more than 6 channels; however, if there are fewer than 4 channels, intelligibility scores decrease steeply.

Figure 3: Monosyllable intelligibility scores as a function of numbers of channels of the CIsim.

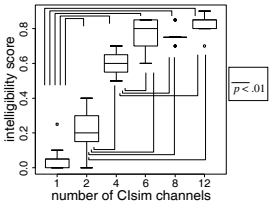
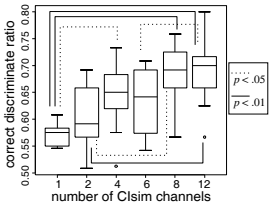


Figure 4: Correct speaker discrimination ratio as a function of numbers of channels of the Cisim.



On the other hand, although the accuracy of speaker discrimination decreased according to the reduction in the number of channels of CIsim (Fig. 4), and the speaker discrimination score was almost a chance level of 0.572 in the 1ch condition, the score was limited to 0.690 even in the 12ch condition. A repeated one-way ANOVA revealed the effect of the number of CIsim channels on speaker discrimination score was also significant ($p < 0.001$). However, the result of Tukey's HSD showed a significant difference was observed if the difference in the number of channels was more than 4, and no significant difference was observed if the difference in the number of channels was 2. This result indicates the following points: the score of the test can reflect the efficiency of the CIsim; and the effect of degradation of sound quality on speaker discrimination is gradual.

Comparing the result of the speaker discrimination test with that of the monosyllable intelligibility test, correlation between the results of the tests was observed (r=0.563). However, as described above, a difference between the response tendency of the speaker discrimination test and the monosyllable intelligibility test was also observed. Thus, the score of the speaker discrimination test could not be predicted directly from the intelligibility scores, and to assess the performance of speaker discrimination, the intelligibility test is insufficient and the test developed in this research is required.

5. CONCLUSION

These findings can be summarized as follows: 1) The score of the speaker discrimination test can reflect the sound quality of hearing-assistance devices as can monosyllable intelligibility test. 2) However, a difference between the response tendency of the speaker discrimination test and the monosyllable intelligibility test was observed. These results indicate that the score of the speaker discrimination test cannot be predicted directly from intelligibility scores. Therefore, to assess the performance of speaker discrimination by using hearing assistance devices, the intelligibility test is insufficient and the test developed in this research is required.

In conclusion, the test can serve to evaluate the performance of hearing assistance devices.

6. ACKNOWLEDGEMENTS

This research was supported by the Grants-in-Aid for Scientific Research of Japan Society for the Promotion of Science (21700592, 22680038) and the Funding Program for Next-Generation World-Leading Researchers provided by the Cabinet Office, Government of Japan.

7. REFERENCES

[1] Amano, S., Kondo, T. 1999. *Lexical Properties of Japanese*. Tokyo: Sanseido.

[2] Amano, S., Sakamoto, S., Kondo, T., Suzuki Y. 2009. Development of familiarity-controlled word lists 2003 (FW03) to assess spoken word intelligibility in Japanese. *Speech Communication* 51, 76-82.

[3] Amino, K., Arai, T., 2009. Effects of linguistic contents on perceptual speaker identification: Comparison of familiar and unknown speaker identifications. *Acoust. Sci. Tech.* 30, 89-99.

[4] Gonzalez, J., Oliver, J.C. 2005. Gender and speaker identification as a function of the number of channels in spectrally reduced speech. *JASA* 118, 461-470.

[5] Kagomiya, T., Nakagawa, S. 2010. An evaluation of bone-conducted ultrasonic hearing aid regarding perception of paralinguistic information. *Proc. Speech Prosody 2010* Chicago, 100867:1-4.

[6] Műler, R., Ziese, M., Rostalski, R. 2009. Development of a speaker discrimination test for cochlear implant users based on the oldenburg logatome corpus. *J. Oto-Rhino-Laryngology, Head and Neck Surgery* 71, 14-20.

[7] Nakagawa, S., Okamoto, Y., Fujisaka, Y. 2006. Development of a bone-conducted ultrasonic hearing aid for the profoundly sensorineural deaf. *Trans. of the Japanese Soc. Med. Biol. Eng.: BME* 44, 184-189.