社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
SP2005-168 (2006-03)

Improving Speech Intelligibility for Elderly Listeners by Steady-State Suppression

Kei KOBAYASHI† Yukari HATTA† Keiichi YASU† Shinji MINAMIHATA†
Nao HODOSHIMA† Takayuki ARAI† and Mitsuko SHINDO‡

† Department of Electrical and Electronics Eng., Sophia University

‡ Research Center for Communication Disorders, Sophia University

7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan

E-mail: †kei-koba@ba2.so-net.ne.jp

Abstract Many individuals experience some degree of hearing loss as they age. In previous studies, Arai et al. (2001, 2002) reported that steady-state suppression of speech improves speech intelligibility in reverberant environments. Steady-state portions are defined as those having more energy, but which are less crucial for speech perception. Kobayashi et al. (2005) confirmed the possibility of consonant enhancement for improved intelligibility when using a hearing aid and the results indicated that intelligibility of a monosyllable was improved with significant difference for 50 elderly listeners. In the present study, we also investigated whether intelligibility of words and monosyllables in speech and intelligibility of monosyllables co-occurring with stationary noise was improved for 23 elderly listeners by consonant enhancement using the steady-state suppression. The results indicated that the intelligibility of a monosyllable in speech was improved in individuals with hearing impairment and the intelligibility of a word might be relatively effective to the degradation of their hearing levels.

Keywords Speech Intelligibility, Consonant enhancement, Hearing impairment, Steady-state suppression, Elderly people, Hearing aid

1. Introduction

Many individuals experience some degree of hearing loss as they age. Sensory hearing loss involves degradations of frequency selectivity, temporal resolution and temporal masking, and is characterized on an audiogram as a gradual bilateral decrease in high frequency hearing. Much less is known about the decline in temporal resolution and masking in the elderly, although Gehr and Sommers [1] showed experimentally that the elderly were influenced more by masking than young people. On the other hand, speech has a complex structure, being comprised of compound sounds with various frequency characteristics. A consonant, for example, is characterized by short temporal length and includes many noise components [2]. The formant transitions and voice onset time (VOT) in the consonant section of a speech sound such as a voiced stop consonant are also used as cues of speech. Therefore, it appears reasonable that a consonant will not be well heard when there is degradation of high frequency discrimination or when there is degradation of temporal processing.

Many consonant enhancement techniques have been applied to hearing aids so as to improve listening capability in sensorineural hearing loss. Gordon-Salant [3] and Kennedy et al. [4] reported consonant enhancement via processing that changed the level ratio of the amplitude of a monosyllable (CV) effectively increased intelligibility in the elderly and individuals with sensorineural hearing loss. Yoshizumi et al. [5] suggested a method of consonant enhancement that has a lateral inhibition mechanism which takes account of temporal masking. Yonemoto and Kurauchi [6] suggested a temporal shift in processing whereby the consonant is placed far from both the initial and final vowels. This technique was found to effectively compensate for sensory hearing loss associated with degradation of temporal resolution. These methods represent just a few of the various consonant enhancement techniques, which also includes compressive amplification [7] and VOT processing [8].

Kobayashi et al. [9] and Arai et al. [10] suggested a new consonant enhancement technique for suppressing the

stationary portion of a vowel, which is not so important for speech discrimination [11, 12]. Steady-state suppression [13, 14, 15] was used to suppress the stationary portion of the vowel for the pre-processing in the reverberant environment but also led to effective enhancement of its consonant in a monosyllable. Moreover, steady-state suppression does not suppress the end of the vowel in a monosyllable. The end of the vowel, called the off-glide, is important for speech perception because it is associated with a consciousness of the following consonant [16]. Also it does not suppress the formant transition of a consonant because these transitions are not stationary. In a previous experiment by Kobayashi et al. [9] the technique indicated significant improvement in intelligibility of a monosyllable by fifty elderly people. In the present study, for an investigation of the effect of a continuously syllables, the effect against speech, and the effect in a noise environment, we report the intelligibility of a monosyllable in speech, the intelligibility of a word and the intelligibility of a monosyllable in a white noise which is representative of stationary noise.

2. Implementation of consonant enhancement using the steady-state suppression method

A principal method employed here was the same as that used in our previous work [9]. Briefly, steady-state suppression as proposed by Arai et al. [13] was used to enhance consonants, but in the present study, the algorithm was modified such that the suppression parameter changed from 40% to 50% since this could be heard more naturally for young adult (normal hearing), and a consonant discriminative processing with spectral moment (mean) analysis [17, 18] was added in order to give an information to judge a suppression in a speech with spectral transition (*D*) [11]. The threshold of the moment was taken as 3750 Hz.

3. Experiments

3.1. Stimuli

The stimuli for investigating intelligibility of a monosyllable in speech were created by inserting a monosyllable as the target into a Japanese carrier sentence, “daimoku to shite wa the target to iimasu” (“As a title, it is called the target”). The monosyllables were obtained from the ATR Japanese speech database (speaker: MAU, 40 years-old male). Consonants /p/, /t/, /k/, /b/, /d/, /g/, /s/, /ə/, /h/, /tə/, /dz/, /dŋ/, /m/, /n/ were used as Consonant (C) and vowels of /a/, /i/ was used as the vowel (V) in each CV. The loudness of monosyllables and the carrier sentence was pre-set for equal loudness, and the start position of the vowel of each target was inserted into the career sentence so that it was consistently 150 ms from the final portion of a vowel that was in front of the target. Thus, we created what we expected listeners to perceive as original speech. Processed speech was produced by the consonant enhancement technique using the steady-state suppression toward the original speech sample.

With regard to the creation of stimuli for investigating the intelligibility of a word with 4-mora phonemes, a Japanese speech database for the word intelligibility test based on word familiarity and phonetic balance [19] was used. We used 48 words categorized into four different familiarity groups (12 words for each group). The word list is shown in TABLE I. Because presentation of the same word twice to a subject—once without processing (original signal) and then later with processing, or vice versa—increases the correct rate for the stimulus presented later, one familiarity group was divided into two subgroups (6 words for each subgroup) which are indicated by an asterisk in Table I. Thus, we prepared two sets of 24 words, and each subject heard each word only once, one set without processing and the other set with processing.

In order to create the stimuli for investigating the intelligibility of a monosyllable in a white noise, the same Japanese monosyllables /pa/, /ba/, /sa/, /təú/, /dza/ and /ma/ were used. The white noise was added to the each monosyllable with seven different signal-to-noise ratios (S/Ns): -10, -5, 0, 5, 10, 15, 20 dB. Thus, we created what we expected listeners to perceive as original speech. Processed speech was produced by the consonant enhancement technique using the steady-state suppression toward the original speech samples. There were 84 stimuli in total.

All stimuli as stated above were processed with 16 kHz sampling.

TABLE I. Word list for investigating the intelligibility of a word with 4-mora phonemes (male speaker） .The asterisk (*) indicates the divide of subgroups for presentation to a subject.

Groups (familiarity)	Words
Fami1 (2.5~1.0)	a.i.kya.ku i.chi.ha.tsu u.ra.jya.ku e.ra.bu.tsu o.ku.de.N ga.ra.yu.ki a.i.ba.N * i.ri.ga.ta * u.bu.su.na * e.ki.yu.: * ga.ku.sa.N * ka.yu.ba.ra *
Fami2 (4.0~2.5)	a.to.zu.ke * i.chi.yu.: * u.chi.wa.ta * o.shi.wa.ri * ka.tsu.gi.ya * ga.N.ku.tsu * i.to.wa.ku u.su.be.ri o.shi.ku.ra ka.ri.o.ya kyu.:.se.tsu gi.N.se.N
Fami3 (5.5~4.0)	a.i.a.i i.chi.bu.N u.ra.ga.ne o.ha.gu.ro ga.i.yu.: ka.za.a.na a.re.ha.da * i.ri.fu.ne * u.wa.ba.ri * e.su.pu.ri * o.shi.no.bi * ga.ku.da.N *
Fami4 (7.0~5.5)	a.ma.gu.mo * i.ma.fu.: * u.chi.ga.wa * o.shi.da.shi * o.ya.mo.to * ga.ni.ma.ta * a.shi.ba.ya u.wa.ba.ki e.:yu.: o.ya.yu.bi ka.ta.ma.ri ga.bu.no.mi

3.2. Subjects

Subjects were 9 male and 14 female Japanese. They all lived in Chiyoda City, Tokyo. The average age of subjects was 72.7 years (range: 64 - 91 years) and the average hearing level was 23.8 dBHL. None of them had been assessed to suffer from mental disorders such as dementia nor had a history of wearing a hearing aid.

In terms of hearing ability, subjects were classified as having normal hearing (n=9; Group A), gradual hearing loss (n=4; Group B) and other types of hearing loss (n=10, Group C). The number, mean age, and mean hearing level (air conduction threshold in good ear) of each group are shown in TABLE II and audiograms are shown in FIG. 1, FIG. 2 and FIG. 3. Group A subjects had hearing levels above 30dBHL and no degradation in the audiogram as seen in the subjects in the other groups. Group B subjects had gradual hearing loss. Group C subjects had other hearing loss due to degradation in lower frequency or a specific frequency or hearing level or right-left difference, etc.

TABLE II. Subjects

Groups	Number	Mean age (years-old)	Mean hearing level (dbHL)
A	9	71.3	15.0
B	4	77.5	19.4
C	10	72.0	33.4

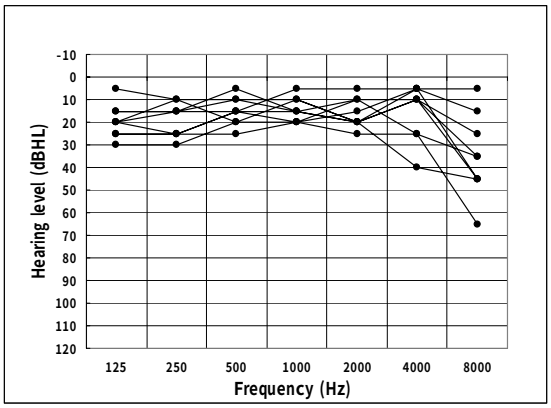

FIG. 1. Audiograms of group A (normal hearing)

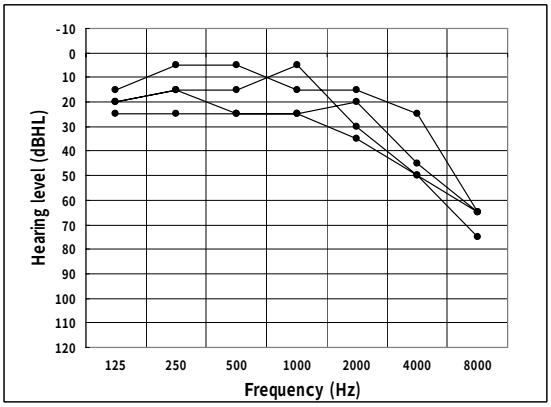

FIG. 2. Audiograms of group B (gradual hearing loss)

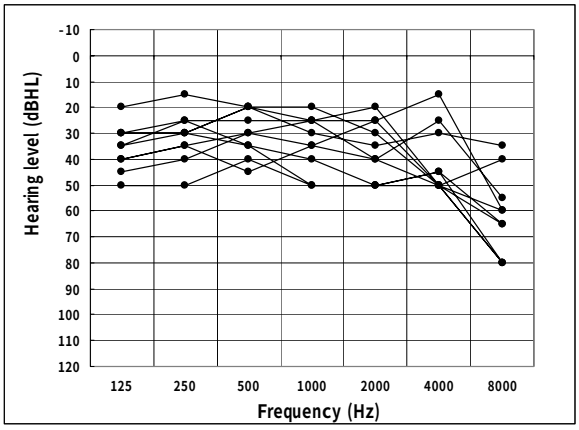

FIG. 3. Audiograms of group C (other types of hearing loss)

3.3. Procedure

Instructions for each experiment were displayed on a computer screen in a soundproof chamber. Before presentation of the stimulus through headphones (STAX SR-303), the sound pressure level was adjusted to a suitable level for each subject. To familiarize the subject to the experimental procedure, several practice trials were permitted prior to starting each experiment. We presented each stimulus only once in each trial, and the subject wrote the word on an answer sheet provided after each presentation. After each trial finishes, next trial screen was displayed by clicking a guide button on the screen, and the next stimulus was presented. Stimuli in each experiment were randomly presented.

4. Results

4.1. Intelligibility of a monosyllable in speech

FIGURE 4 shows the intelligibility of a monosyllable for the three groups. For total subjects, intelligibility of original speech was 88.3% (σ =0.18) and intelligibility of processed speech was 89.5 % (σ =0.15), which did not represent a significant difference due to processing.

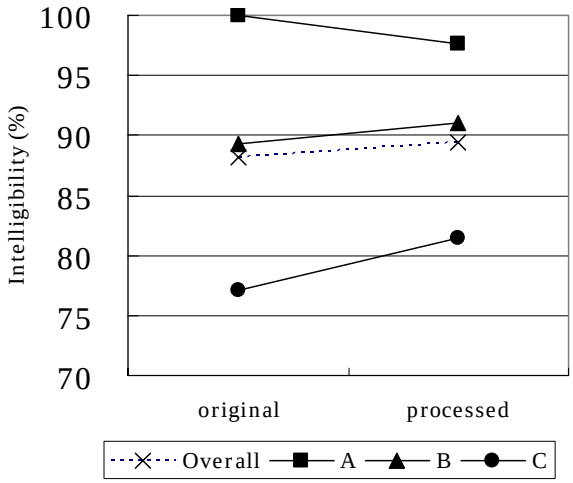

FIG. 4. Results of intelligibility of original and processed monosyllables in speech according to each subject group.

4.2. Intelligibility of a word in speech

FIGURE 5 shows the relation between hearing levels and intelligibility of a word in speech. In addition, for total subjects, intelligibility of original speech was 96.7% (σ =0.03) and intelligibility of processed speech was 96.9% (σ =0.02). There is no significant difference due to processing.

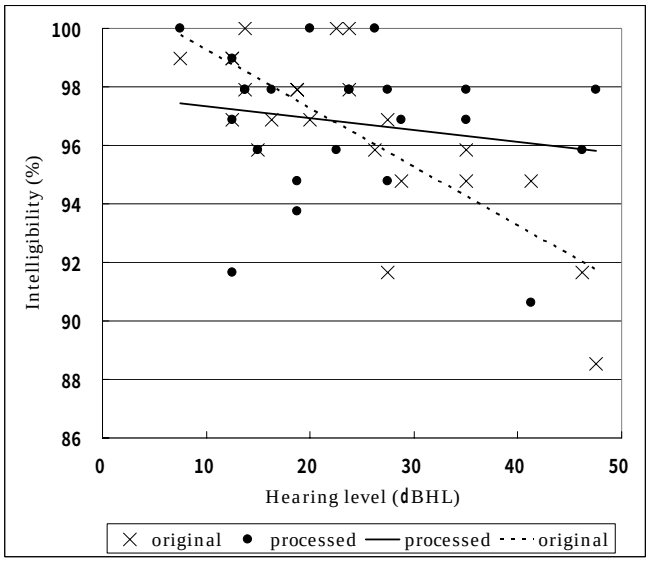

FIG. 5. Results of the intelligibility of a word.
Solid and dashed lines are the linear regressions for the results of the original and processed words, respectively.

TABLE III. Results of the intelligibility of a monosyllable in white noise.

	S/N [dB]	-10	-5	0	5	10	15	20
A	Original [%]	3.7	33.3	68.5	79.6	92.6	96.3	94.4
	Processed [%]	1.9	25.9	59.3	75.9	77.8	88.9	96.3
B	Original [%]	8.3	37.5	45.8	66.7	87.5	91.7	95.8
	Processed [%]	8.3	25.0	41.7	70.8	79.2	91.7	87.5
C	Original [%]	8.3	33.3	61.7	70.0	80.0	78.3	83.3
	Processed [%]	3.3	38.3	65.0	63.3	73.3	86.7	86.7
Overall	Original [%]	6.8	34.8	61.4	72.7	85.6	87.1	89.4
	Processed [%]	3.8	31.1	58.3	68.9	75.8	87.9	90.2

4.3. Intelligibility of a monosyllable in a white noise

TABLE III shows the intelligibility of a monosyllable in a white noise for total subjects and for each group. Various effects of processing for each S/N were seen for total subjects and for each group, however, tendentious and significant effects due to processing were not observed.

5. Discussion

We begin by discussing the results of the intelligibility of a monosyllable in speech (FIG. 4). For group A, the processing had no effect on intelligibility since the intelligibility of the original speech was the ceiling. However, for groups B and C, while processing did produce a positive effect, this effect was not significant because we had few subjects. In addition, in group A, a minor positive effect for processing in S/N=20dB was seen (TABLE III). it might be indicated the possibility of the effect for group A in the unclean condition that would not lead a ceiling intelligibility of original speech. In a fact, some previous studies showed the effect of steady-state suppression in reverberant condition [13, 14, 15]. The relation between intelligibility and forward masking by the vowel “a”, which is the end of the carrier sentence preceding this target, is under investigation now.

Next we discuss the results of the intelligibility of a word in speech (FIG. 5). A significant effect of processing was again not seen for the total subjects, for any familiarity group or for position of the monosyllable in the word. This may be due to the ceiling effect since the total subjects had 96% or more intelligibility of original words. It is suggested by the experiment that the intelligibility of a word of the elderly, especially of those with hearing loss, may be superior to the intelligibility of a monosyllable. However, the two linear regressions in FIG. 5 show there was a tendency for intelligibility among the subjects with the degradation of hearing level to be improved with processed speech compared to original speech because the intelligibility of processed speech was improved better than the intelligibility of original speech with the degradation of hearing level. A strong negative correlation (R=-0.772) was seen for the linear regression of the original speech but correlation for the linear regression of the processed speech was seldom seen (R=-0.188). The relation between temporal masking and intelligibility is currently under investigation.

With regard to the intelligibility of a monosyllable in a white noise (TABLE III), though processed speech was not only suppressed its vowel but also well suppressed a white noise which was stationary, the effect of processing was no tendentious and was not shown significant effect for intelligibility of a monosyllable. Then, we discuss in terms of /ba/ in S/N=0dB, where intelligibility declined, and /pa/ in S/N=0dB, where intelligibility improved. Many subjects recorded the voiced stop consonant /ba/ as a voiceless stop consonant /pa/. The addition of a white noise resulted in the erroneous suppression of the voiced stop consonant portion. On the other hand, since /pa/ was not suppressed in the consonant portion, intelligibility was improved. A white noise has its components in several frequency bands and the present technique could not discriminate the expect steady-state bands of a consonant in the environment. Thus, considering application of the present technique to a hearing aid, unstable results will likely occur if a white noise is contained in a monosyllable. Reliable operation requires improvement of the calculation of D and its threshold, and we aim to make such improvement in the near future.

6. Conclusions

In this study, we investigated whether steady-state suppression of vowels would afford any benefits for improved intelligibility of a monosyllable in speech, of a word and of a monosyllable in white noise. The results showed that the consonant enhancement technique using steady-state suppression might be effective for improving the intelligibility of a monosyllable in speech in individuals with hearing impairment. As intelligibility of a word may be relative to hearing level, steady-state suppression of a vowel may be effective in proportion to the degradation of hearing level. Under the conditions of a white noise, it is suggested by this experiment that a device such as a hearing aid is required for more accurate calculation of *D* and its threshold value. Furthermore, we continue our investigation of the relation between temporal masking and intelligibility in order to find the effect of a masking from neighbor vowel in a monosyllable in speech and in a word.

7. Acknowledgements

This research was supported in part by Grants-in-Aid for Scientific Research (A-2, 16203041) from the Japan Society for the Promotion of Science. We wish to express our many thanks to the participants from the CHIYODA CITY Silver Human Resources Center.

Reference

[1] S. E. Gehr and M. S. Sommers, “Age differences in backward masking,” J. Acoust. Soc. Am., 106(5), pp. 2793-2799, 1999.

[2] R. D Kent and C. Read, The Acoustic Analysis of Speech, T. Arai and T. Sugawara, ed., KAIBUNDO, Tokyo, 1996.

[3] S. Gordon-Salant, “Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing,” J. Acoust. Soc. Am., 80(6), pp. 1599-1607, 1986.

[4] E. Kennedy, H. Levitt, A. C. Neuman, and M. Weiss, “Consonant-vowel intensity ratios for maximizing consonant recognition by hearing–impaired listeners,” J. Acoust. Soc. Am., 103(2), pp. 1098-1114, 1998.

[5] Y. Yoshizumi, T. Mekata, Y. Yamada, R. Suzuki, Y. Tanaka, A. Kawano and S. Funasaka, “Speech enhancement algorithms for hearing impaired subjects: An evaluation for hearing impaired subjects,” Trans. Tech. Comm. Psychol. Pshysiol. Acoust., The Acoustic Society of Japan., H-93-18, 1993.

[6] K. Yonemoto, N. Kurauchi, “Auditory function and effective speech processing in sensorineural hearing impaired,” Trans. Tech. Comm. Psychol. Pshysiol. Acoust., The Acoustic Society of Japan., H-90-11, 1990.

[7] J. Shidara, K. Kodera and M. Suzuki, “Improvement of consonant confusion by digital compression,” Audiology Japan 39, pp. 284-290, 1996.

[8] A. Hayashi, S. Imaizumi, T. Harada, H. Seki and H. Hosoi, “Relationships between ear’s temporal window & VOT perception: Experimental considerations,” Trans. Tech. Comm. Psychol. Physiol. Acoust., The Acoustic Society of Japan., H-90-10, 1990.

[9] K. Kobayashi, Y. Hatta, K. Yasu, N. Hodoshima, T. Arai and M. Shindo, “A study of monosyllable enhancement for elderly listeners by steady-state suppression”, Technical Report of IEICE, SP2004-155, pp. 7-12, 2005.

[10] T. Arai, K. Yasu and N. Hodoshima, “Effective speech processing for various impaired listeners,” Proceedings the 18th International Congress on Acoustics, 2, pp. 1389-1392, 2004.

[11] S. Furui, “On the role of spectral transition for speech perception,” J. Acoust. Soc. Am., 80(4), pp. 1016-1025, 1986.

[12] H. Hermansky and N. Morgan, “RASTA processing of speech,” IEEE Trans. Speech and Audio Process., 2, pp. 578-589, 1999.

[13] T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto and T. Kitamura, “Effects on suppressing steady-state portions of speech on intelligibility in reverberant environments," Acoust. Sci. and Tech., 23(4), pp. 229-232, 2002.

[14] N. Hodoshima, T. Inoue, T. Arai, A. Kusumoto and K. Kinoshita, "Suppressing steady-state portions of speech for improving intelligibility in various reverberant environments," Acoust. Sci. and Tech., 25(1), pp. 58-60, 2004.

[15] N. Hodoshima, T. Arai, T. Inoue, K. Kinoshita and A. Kusumoto, "Improving speech intelligibility by steady-state suppression as pre-processing in small to medium sized halls," Proc. Eurospeech, pp. 1365-1368, 2003.

[16] W. Strange, J. Jenkins and T. L. Johnson, “Dynamic specification of coarticulated vowels,” J. Acoust. Soc. Am., 74(3), pp. 695-705, 1983.

[17] A. Jongman, R. Wayland and S Wong, “Acoustic characteristics of English fricatives,” J. Acoust. Soc. Am., 108(3), pp. 1252-1263, 2000.

[18] K. Forrest, G. Weismer, P Milenkovic and R. N. Dougall, “Statistical analysis of word-initial voiceless obstruents: Preliminary data,” J. Acoust. Soc. Am., 84(1), pp. 115-123, 1988.

[19] S. Sakamoto, Y Suzuki, S Amano, K. Ozawa, T. Kondo, and T. Sone, "New lists for word intelligibility test based on word familiarity and phonetic balance," Journal of the Acoustic Society of Japan., 54(12), pp. 842-849, 1998.