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
SP2005-165 (2006-03)

The effect of the steady-state suppression on consonant identification by native and non-native listeners in reverberant environments

Nao HODOSHIMA† Dawn BEHNE‡ and Takayuki ARAI†

† Dept. of Electrical and Electronics Engineering, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan
‡ Dept. of Psychology, Norwegian University of Science and Technology, Trondheim, NO-7491 Norway
E-mail: † n-hodosh@sophia.ac.jp

Abstract This study investigated whether the steady-state suppression proposed by Arai et al. (Proc. Autumn Meet. Acoust. Soc. Jpn., 2001; Acoust. Sci. Tech., 2002) improved consonant identification for non-native listeners in reverberation. This study also compared the effect of steady-state suppression on consonant identification by native and non-native listeners in reverberant environments. We used steady-state suppression as a pre-processing technique which processes speech signals before they are radiated from loudspeakers in order to reduce the amount of overlap-masking. Participants were 24 native English (native listeners) and 24 Japanese speakers (non-native listeners), both with normal hearing. A diotic Modified Rhyme Test (MRT) was conducted under 2 processing conditions (with or without steady-state suppression) for 3 reverberant conditions (reverberation times of 0.4, 0.7 and 1.1 s) and a dry condition. The results showed that native listeners performed better than non-native listeners in all conditions used in this study. Although there were no significant differences between unprocessed and steady-state suppressed stimuli, and no significant interaction between the effect of the steady-state suppression and listener group under the reverberant conditions used in the current study, the effect of the steady-state suppression differed in consonant position, reverberation time and listener group. These findings imply that a pre-processing technique would be required which helps non-native listeners to identify consonants as well as native listeners do.

Keyword Speech enhancement, Non-native listeners, Reverberation, Steady-state suppression, Speech intelligibility

1. Introduction

With the increase in internationalization, more and more opportunities for listening or speaking a foreign language arise. For example, this year for the first time, a listening-comprehension test was introduced to the English exam in unified university entrance examinations in Japan. It has been reported that non-native listeners [1, 2] as well as elderly people [3, 4] and the hearing-impaired [5, 6] have more difficulty understanding speech under noisy or reverberant environments than young people with normal hearing. Therefore, less noise and reverberation are required in public spaces where non-native languages are in use, such as international airports, stations, and conference rooms. Care must also be taken in classrooms, lecture halls, etc for educational purposes.

Hazan and Simpson [7, 8] reported a speech enhancement method for non-native listeners in noise. They enhanced consonantal regions of VCV stimuli by pre-processing stimuli before adding noise, and presenting them in a background of noise with the long-term average spectrum corresponding to the speech signal. Non-native and native listeners groups obtained significantly higher intelligibility scores for the enhanced stimuli compared to natural stimuli. However, to our knowledge, no speech enhancement technique is suggested for non-native listeners in reverberation.

Steady-state suppression [9, 10] was proposed as a pre-processing approach in order to reduce the effect of overlap-masking which is the main reason reverberation reduces speech intelligibility [11, 12]. This technique suppresses steady-state portions of speech that are not necessary for syllable perception [13]. The steady-state suppression statistically improved consonant identification for young normal hearing people [14-16] and elderly people [17] under diotic listening conditions at reverberation times (RTs) of 0.7 to 1.3 s, and for young normal hearing people in a dichotic listening environment at RT of 1.3 s [18].

The purposes of the current study were 1) to investigate whether the steady-state suppression proposed by Arai et al. [9, 10] improved consonant rhyming for non-native listeners in reverberation and 2) to compare the effect of steady-state suppression on consonant rhyming by native

TABLE 1 Summary of 'native listeners' and 'non-native listeners' group characteristics.

	Native listeners	Non-native listeners
Native language	English	Japanese
Numbers	14 males, 10 females	6 males, 18 females
Age	18 to 50 years (average: 31 years)	20 to 30 years (average: 22 years)
Thresholds	less than 25 dB HL from 250 to 4 kHz	less than 25 dB HL from 125 to 8 kHz

and non-native listeners in reverberant environments. Twenty-four native English speakers and twenty-four Japanese speakers as non-native listeners participated in a listening test. A diotic Modified Rhyme Test was conducted under 2 processing conditions (with or without the steady-state suppression) for 3 reverberant conditions (RTs of 0.4, 0.7 and 1.1 s) and a dry condition.

2. Experiment

2.1. Subjects

Table 1 summarizes the two listener groups. The "native listeners" group consisted of 24 native speakers of English and the "non-native listeners" group included 24 native speakers of Japanese. Pure-tone thresholds were less than 25 dB HL from 250 to 4 kHz for native listeners and less than 25 dB HL from 125 to 8 kHz for non-native listeners. The limited frequency ranges of the thresholds of native listeners were due to the audiometers available at the two experiment sites. None of the subjects reported a history of unusual noise exposure or listening difficulties.

Non-native listeners had an average level of English proficiency in Japan based on the following criteria: 1) had English as a second language, 2) never lived abroad, 3) attained a C on the TOEIC (Test of English for International Communication) proficiency scale, attained a middle or primary rank on the English test which all first year students take at Sophia University, where the experiment was conducted. They began learning English when they were, on average, 11 years old, and had studied English for an average of 11 years

2.2. Speech materials

The version of the Modified Rhyme Test (MRT) used in Kusumoto et al. [19] was also used in the current study. The target words developed by Kruel et al. [20] were embedded within the carrier phrase, "You will mark the ____, please." All 6 lists, each composed of 50 monosyllabic words, were used. The speaker was a male native speaker of American-English (34 years old, the American standard dialect). The average intensity of the stimuli was normalized within each sentence.

TABLE 2 Reverberation times (RTs) used in the listening test.

RT (s)	0.4	0.7	1.1

2.3. Processing conditions

Two processing conditions were used in this study: original (unprocessed) speech and steady-state suppression as was used by Arai and his colleagues [9, 10]. This method calculates the *D* parameter to detect spectral transitions of a speech signal [13], and defines speech portions as steady-state when *D* is less than a specified threshold. Once a portion is considered steady-state, the amplitude of the portion is multiplied by a factor of 0.4, giving a 40% suppression rate.

2.4. Reverberant conditions

Table 2 lists the RTs used in the present study. The impulse responses corresponding to the RTs in Table 2 were obtained by multiplying exponential decays by the impulse response measured in Hamming Hall in Tokyo as described in [21]. The RT values are the average RTs derived from the Early Decay Time at the center frequencies of 0.5, 1, and 2 kHz of the 1-octave bandpass filtered impulse response.

2.5. Procedure

Each subject was tested with all six MRT lists. The lists were assigned to each of the six conditions (2 processing conditions x 3 reverberant conditions) and counterbalanced across subjects. Additionally, 50 sentences randomly selected from across the six lists were used as stimuli in the dry condition.

Two listener groups were tested in two different places. The computer-controlled experiment was conducted in a sound treated room. The stimuli were presented over headphones (AKG K271 for native listeners and STAX SR-303 for non-native listeners). The sound level was adjusted to a comfortable level for each subject beforehand, and this level was maintained throughout the experiment. In each trial, a test sentence was presented over headphones, followed by six words rhyming with the

target word presented visually on a PC screen. Subjects were forced to choose one of the rhyming words by clicking a button on the screen using a computer mouse. Once they had selected a word, the next trial was presented. The 300 reverberant stimuli were randomly presented first, followed by the 50 randomly presented stimuli in quiet condition. The reverberant condition preceded the dry condition, rather than randomizing the conditions, so that familiarity with the stimuli from the dry condition would not affect the reverberant condition. Before starting the experiment, listeners had five practice trials to become familiar with the procedure.

2.6. Results

Figure 1 shows mean percent correct by native and non-native listeners for each of the reverberant and processing conditions collapsed across word position. Figure 2 presents the results with initial and final word position shown separately. Open squares represent native listeners' mean percent correct for the unprocessed stimuli (nl_org). Filled squares represent native listeners' mean percent correct for the steady-state suppressed stimuli (nl_proc). The cross represents native listeners' mean percent correct in dry speeches (nl_d). Open circles represent non-native listeners' mean percent correct for the unprocessed stimuli (nnl_org). Filled circles represent non-native listeners' mean percent correct for the steady-state suppressed stimuli (nnl_proc). The plus symbol represents non-native listeners' mean percent correct in dry speech (nnl_d).

A 2 x 3 x 2 mixed ANOVA was carried out with listener group as a nonrepeated factor, RT, consonant position, and processing as repeated variables, and percent correct as the dependent variable. Results show that native listeners had a higher percent correct than nonnative listeners [$F(1,46) = 119.00$, $p < 0.01$]. The percent correct was also reliably higher for the 0.4 s RT than the 1.1 s RT [$F(2,92) = 68.47$, $p < 0.01$], and higher for initial consonants than for final consonants[$F(1,46) = 1881.06$, $p < 0.01$]. No reliable difference in percent correct was observed between the steady-state and unprocessed conditions. In addition to these main effects, significant interactions were observed between consonant position and listener group [$F(1,46) = 111.32$, $p < 0.01$] and between consonant position and reverberant condition [$F(2,92) = 16.20$, $p < 0.01$]. Other interactions were not significant.

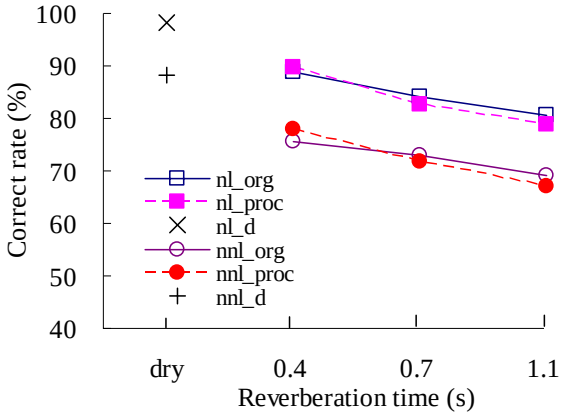

Figure 1 Native and nonnative listeners' mean percent correct presented in different reverberant and processing conditions.

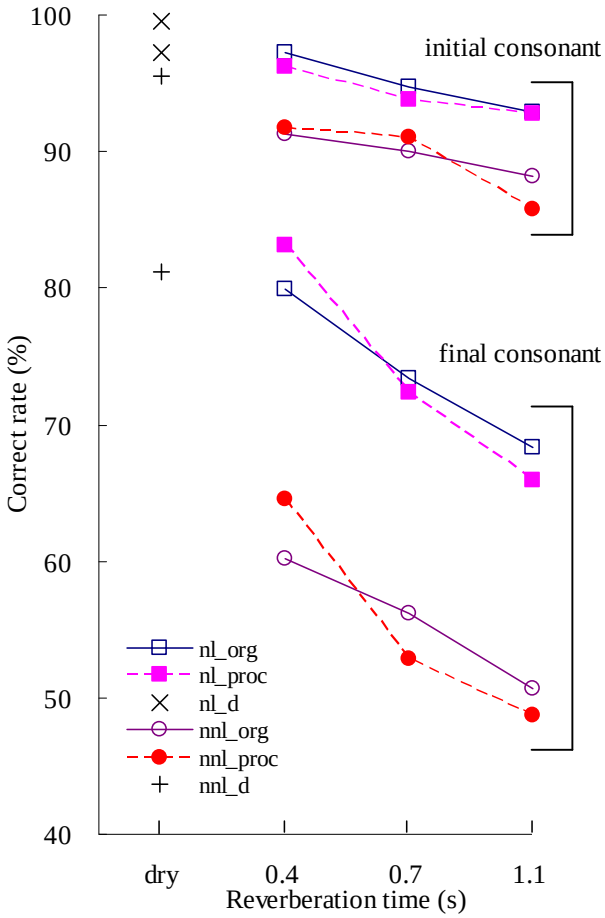

Figure 2 Native and nonnative listeners' mean percent correct for word initial and final consonants presented in different reverberant and processing conditions.

In the dry condition, a 2 x 2 mixed ANOVA was carried out with listener group as a nonrepeated factor, processing and consonant position as a repeated variable, and percent correct as the dependent variable. Results show that native listeners had a higher percent correct than nonnative listeners [$F(1,46) = 96.73$, $p < 0.01$]. The percent correct was also higher for initial consonants than for final consonants [$F(1,46) = 57.16$, $p < 0.01$]. In addition to these main effects, an interaction between listener group and consonant position was significant [$F(1,46) = 29.63$, $p < 0.01$].

3. Discussion

3.1. Native and non-native listeners

The mean percent correct was higher for native listeners than non-native listeners in the reverberant conditions. We observed the same tendency that non-native listeners had more difficulty understanding speech under reverberant environments than young people with normal hearing [1, 2], however, there was no significant interaction between listener group and reverberant condition.

As was observed in the reverberant conditions, native listeners performed better than non-native listeners. The significant main effect of listener group conflicts with [2], showing that there was no difference between native and non-native listeners in word discrimination tasks in dry speech. This inconsistency might be due to a difference between the subjects' English proficiency level. Non-native listeners in [2] lived abroad for a couple of years, while non-native listeners in this study had not necessarily been living abroad. Interestingly, in the current study, the difference between the mean percent correct for native and non-native listeners in dry speech was close to that in the reverberant conditions (the difference was 4.0 % in the dry condition, 5.1% in unprocessed stimuli and 4.8% in processed stimuli in initial consonants; 16% in the dry condition, 18.2% in unprocessed stimuli and 18.4% in processed stimuli in final consonants). This might imply that word identification scores of reverberant condition could be guessed from word identification scores in the dry condition under environments that are close to ones used in this study.

3.2. Percent correct in reverberant speech

In both word initial and final consonants the mean percent correct decreased as RT increased. This was the same tendency observed in [14-18]. In the reverberant conditions, as well as in the dry condition, the mean percent correct in initial consonants was nevertheless higher than in final consonants in reverberation, as was also observed in [22]. The preponderance of errors among final consonants likely resulted from overlap-masking produced by the reverberation of the preceding vowel [22] because the previous phoneme of the final consonant was a stressed vowel while the previous phoneme of the initial consonant was /schwa/ in the current study. In addition, the difference between the mean percent correct for initial and final consonants was larger for non-native listeners than for native listeners under the reverberant conditions (mean differences in percent correct were 20.8% for native listeners and 34.2% for non-native listeners). This indicates that non-native listeners have more difficulty in discriminating final consonants compared to native listeners with increased overlap-masking.

3.3. The effect of the steady-state suppression

Under the conditions tested in this study, processed and unprocessed speech were comparable for word initial and final consonants. Although there were no significant differences between correct rates of unprocessed and processed conditions, the effect of the steady-state suppression differed in consonant position, reverberation time and listener group. For example, the effect of the steady-state suppression was comparable for native and non-native listeners in final consonants (the mean percent correct for processed stimuli was higher than for unprocessed stimuli at a 0.4 s RT, the mean percent corrects for unprocessed stimuli was higher than for processed stimuli at 0.7 s and 1.1 s RTs). In contrast, the effect of the steady-state suppression differed in initial consonants between native and non-native listeners (the mean percent correct for processed stimuli was higher than for unprocessed stimuli at RTs of 0.4 s and 0.7 s for non-native listeners, the mean percent correct for unprocessed stimuli was higher than for processed stimuli at RTs of 0.4 s and 0.7 s for native listeners). It would be interesting to see the difference between the effect of the steady-state suppression for native and non-native listeners under other reverberant and processing conditions than those used in this study.

Previous studies [14-18] used Japanese monosyllabic words in an identification task, whereas the current study used an English consonant rhyming task. In the current

study, the same steady-state suppression parameters were used as in [14-18]. It's not clear, however, that these parameters are appropriate for English rhyming tasks. Since different materials such as tasks and languages were used in both listening tests, the appropriate parameters of the steady-state suppression might be different from previous studies [14-18]. Further investigations are needed to adjust parameters of the steady-state suppression for English rhyming tasks.

4. Conclusions

The current study 1) investigated whether the steady-state suppression proposed by Arai et al. [9, 10] improved consonant rhyming for non-native listeners of English in reverberation and 2) compared the effect of the steady-state suppression on consonant rhyming by native English listeners and non-native listeners of English in reverberant environments. The results showed that native listeners performed better than non-native listeners in all conditions used in this study. The result also showed that the mean percent correct decreased as RT increased, and was higher in initial consonants than in final consonants. In addition, the difference between the mean percent correct for initial and final consonants was larger for non-native listeners than for native listeners in reverberant conditions. Although there were no significant differences between unprocessed and steady-state suppressed stimuli, and no significant interaction between the effect of the steady-state suppression and listener group under the reverberant conditions used in the current study, the effect of the steady-state suppression differed in consonant position, reverberation time and listener group. These findings imply that a pre-processing technique would be required which helps non-native listeners to identify consonants as well as native listeners do.

Acknowledgements

This research was supported by Grants-in-Aid for Scientific Research (A-2, 16203041) and Grants-in-Aid for JSPS Fellows (176911), both from the Japan Society for the Promotion of Science. The authors would like to thank Hideki Tachibana, Kanako Ueno and Sakae Yokoyama for offering the use of the impulse response data, and Chikashi Michimata and Hirofumi Kamata for their guidance regarding the statistical analysis.

References

[1] A. K. Naœbȩlek and A. M. Donahue, "Perception of consonants in reverberation by native and non-native listeners," J. Acoust. Soc. Am., Vol.75(2), pp.632-634, 1984.

[2] Y. Takata and A. K. Núœbȩlek "English consonant recognition in noise and in reverberation by Japanese and American listeners," J. Acoust. Soc. Am., Vol.88, pp.663-666, 1990.

[3] A. K. Núœbȩlek and P. K. Robinson, "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am., Vol.71(4), pp.1242-1248, 1982.

[4] K. S. Helfer and R. A. Huntley, "Aging and consonant errors in reverberation and noise," J. Acoust. Soc. Am., Vol.90(4), pp.1786-1796, 1991.

[5] A. K. Núœbȩlek and J. M. Pickett, "Monaural and binaural speech perception through hearing aids under noise and reverberation," J. Speech Hear. Res., Vol.17, pp.724-739, 1974.

[6] T. Finitzo-Hieber and T. Tillman, "Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children," J. Speech Hear. Res., Vol.21(3), pp.440-458, 1978.

[7] V. Hazan and A. Simpson, "The effect of cue-enhancement on consonant perception by non-native listeners: Preliminary results," Proc. StiLL Workshop, pp.119-122, 1998.

[8] V. Hazan and A. Simpson, "The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects," Language and Speech, Vol.43(3), pp.273-294, 2000.

[9] T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto and T. Kitamura, "Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments," Proc. Autumn Meet. Acoust. Soc. Jpn. Vol.1, pp.449-450, 2001 (in Japanese).

[10] 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. Tech., Vol.23, pp.229-232, 2002.

[11] R. H. Bolt and A. D. MacDonald, "Theory of speech masking by reverberation," J. Acoust. Soc. Am., Vol.21, pp.577-580, 1949.

[12] A. K. Núœbȩlek and L. Robinette, "Influence of precedence effect on word identification by normally hearing and hearing-impaired subjects," J. Acoust.

Soc. Am., Vol.63, pp.187-194, 1978.
[13] S. Furui, “On the role of spectral transition for speech perception,” J. Acoust. Soc. Am., Vol.80, pp. 1016-1025, 1986.
[14] N. Hodoshima, T. Inoue, T. Arai and A. Kusumoto, “Suppressing steady-state portions of speech for improving intelligibility in various reverberant environments,” Acoust. Sci. Tech., Vol.25, 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] N. Hodoshima and T. Arai, “Investigating an optimum suppression rate of steady-state portions of speech that improves intelligibility the most as a pre-processing approach in reverberant environments,” J. Acoust. Soc. Am., Vol.118, pp.1930, 2005.
[17] Y. Miyauchi, N. Hodoshima, K. Yasu, N. Hayashi, T. Arai and M. Shindo, “A preprocessing technique for improving speech intelligibility in reverberant environments: The effect of steady-state suppression on elderly people,” Proc. Eurospeech, pp.2769-2772, 2005.
[18] N. Hodoshima, T. Goto, N. Ohata, T. Inoue and T. Arai, “The effect of pre-processing approach for improving speech intelligibility in a hall: Comparison between diotic and dichotic listening conditions,” Acoust. Sci. Tech., Vol.26(2), pp.212-214, 2005.
[19] A. Kusumoto, T. Arai, K. Kinoshita and N. Hodoshima, “Modulation enhancement of speech by preprocessing for improving intelligibility in reverberant environment,” Speech Com., Vol.45(2), pp.101-113, 2005.
[20] E. J. Kreul, J. C. Nixon and K. D. Kryter, “A proposed clinical test of speech discrimination,” J. Speech Hear. Res., Vol.11, pp.536-552, 1968.
[21] N. Hodoshima, T. Arai and A. Kusumoto, “Enhancing temporal dynamics of speech to improve intelligibility in reverberant environments,” Proc. Forum Acusticum Sevilla, 2002.
[22] V. O. Knudsen, “The hearing of speech in auditoriums,” J. Acoust. Soc. Am. Vol.1(1), pp.56-82, 1929.