TECHNICAL REPORT

Intelligibility of English phonemes in noise for native and non-native listeners

Takahiro Adachi[1,2,*], Reiko Akahane-Yamada[1,†] and Kazuo Ueda[3,‡]

[1]*ATR Human Information Science Laboratories,
2–2–2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619–0288 Japan*
[2]*Department of Cyber Society and Culture Course, The Graduate University for Advanced Studies, Hayama-cho, Miura-gun, Kanagawa, 240–0193 Japan*
[3]*Kyushu University, 4–9–1, Shiobaru, Minami-ku, Fukuoka, 815–8540 Japan*

(*Received 27 September 2005, Accepted for publication 21 January 2006*)

Abstract: The effect of noise and presentation level on the perception of English consonants by native listeners and non-native listeners were examined. English words contrasting in /r/–/l/, /b/–/v/ and /s/–/th/ sounds, which are known to be difficult to distinguish for native speakers of Japanese, were presented to both native speakers of American English (AE listeners) and those of Japanese (J) in white noise and in pink noise at sytematically changed signal-to-noise ratios (SNR). Words were also presented at various presentation levels. The effects of noise and presentation level differed by phonetic contrast and language group. The /b/–/v/ and /s/–/th/ contrasts were more affected by noise at high SNRs, while /r/–/l/ was tolerant for noise when the SNR was higher than −3 dB. The presentation level affected AE listeners' identification of /b/–/v/, but not of other contrasts. J listeners' perception was affected less than that of AE listeners, possibly because the flooring effect for J listeners' identification performance was low, even for original stimuli.

Keywords: Speech, Speech intelligibility, Perception, Noise

PACS number: 43.71.Gv [doi:10.1250/ast.27.285]

1. INTRODUCTION

Many studies have reported that speech perception is impaired in noise (e.g, [1]). Furthermore, non-native listeners are less tolerant than native listeners (e.g, [2,3]). Ueda *et al.* [4] measured the identification accuracy of American English words contrasting in /r/ and /l/ in noise, whereby stimuli were presented to native speakers of American English and those of Japanese at systematically varied signal-to-noise ratios (SNR). Results showed that the identification accuracy of native speakers of Japanese was affected by noise at higher SNRs, i.e., with softer noise. Miller and Nicely [5] have shown that the extent of noise tolerance varies depending on the consonants. There is a possibility that the effect of noise on the Japanese speakers' perception of English speech varies according to the consonant. Thus, in this study, we have investigated the intelligibility of the English /b/–/v/ and /s/–/th/ contrasts as well as /r/–/l/ contrast under various SNR conditions for native speakers of American English and those of Japanese. The effect of the sound-pressure level was also examined, because there is a possibility that the presentation level affects the identification performance in a different way depending on the consonants or the native languages of the listener.

2. METHOD

2.1. Stimuli

Fifty pairs of English words minimally contrasting in /r/ and /l/, 30 pairs with contrast in /b/ and /v/, and 30 pairs with contrast in /s/ and /th/ were used as word materials. There were 220 words (110 pairs) in total. It has been reported that Japanese speakers' accuracy of perceiving the English /r/ and /l/ differs depending on the position of these phonemes in the word [6]. Thus, /r/–/l/ word pairs with contrast in various positions were used in this experiment; ten pairs had contrast in the initial singleton, 13 pairs in the initial consonant cluster, nine pairs in intervocalic positions, six pairs in the final consonant cluster, and 12 pairs in the final singleton. For

[*] e-mail: tadachi@atr.jp
[†] e-mail: yamada@atr.jp
[‡] e-mail: ueda@design.kyushu-u.ac.jp

/b/–/v/, 20 pairs had contrast in the initial position, three pairs in intervocalic positions, one pair in the final consonant cluster, and six pairs in the final singleton. For /s/–/th/, 16 pairs had contrast in the initial position, one pair in an intervocalic position, 12 pairs in the final singleton, and one pair in the intervocalic cluster.

Two native speakers of American English, one male and one female, each produced these 220 words in an anechoic chamber. These were recorded onto a DAT tape and sampled onto a computer disk at a 44.1 kHz sampling frequency and 16-bit resolution. Later, the utterances were saved onto word-by-word files.

Sound pressure levels (SPL) were measured with an IEC coupler (Bruel & Kjar, type 2231) with the word files being output through over headphones (STAX SRM-1/ML-2). The amplitude of the sound signal for each word file was changed in order to equalize the SPL. The amplitude was adjusted until the SPL measured with the Fast and A curve characteristics showed 59 dB on average for words with contrast in /r/–/l/ and /s/–/th/, and 65 dB on average for /b/–/v/, in order to maintain high intelligibility for native speakers. These SPL-adjusted files were used as original signals.

White noise and pink noise for each original signal were generated with the noise generator (B&K, Type 1049). The noise signals were 400 ms longer in duration than the original signals, with 5 ms onset and 5 ms offset tapers with linear envelopes. Then, the noise was added to the original signals at various SNRs such that the noise started 200 ms earlier and lasted 200 ms longer than the original signals. The SNR for each phonetic contrast is shown in Table 1.

The amplitudes of the original 220 signals were further varied in order to generate stimulus continua by varying the SPL from 39 dB(A) to 69 dB(A) in 5 dB steps, resulting seven in stimuli on each continuum.

2.2. Participants

Participants comprised 11 native speakers of Japanese (J listeners) and 12 native speakers of American English (AE listeners) living in Japan. The J listeners were undergraduate students, and none of them had any experience of living abroad for more than three months. The J listeners ranged in age from 19 to 26, and the AE listeners from 23 to 43. A hearing screening test performed at 15 dB HL for frequencies from 250 to 8,000 Hz showed all participants to have normal bilateral hearing acuity.

Table 1 SNR of original signal and added noise.

Contrast	SNR [dB]
/r/–/l/	9, 3, 0, −3, −9, −15
/b/–/v/	12, 9, 6, 3, 0, −3, −6, −9*
/s/–/th/	9, 3, 0, −3, −9, −15

*−9 dB condition was tested only for white noise.

2.3. Procedure

Two alternative identification tasks were used in the experiment. On each trial, two of a minimal pair contrasting in /r/–/l/, /b/–/v/ or /s/–/th/ were displayed visually in English ortho-graphical form on two separate buttons on a computer display. Then, one of the words was played binaurally over headphones (STAX, SRM-1/ML-2) at a fixed listening level of 59 dB(/r/–/l/, /s/–/th/) or 65 dB(/b/–/v/) HL, when the original signal was played. A higher listening level was used for /b/–/v/ stimuli, because a preliminary experiment showed that even AE listeners did not show high accuracy if they were played at 59 dB. Participants identified the word they had heard and selected the corresponding button by clicking the mouse. No feedback was provided for participants' responses.

Each participant sat in front of a CRT monitor and a keyboard in a sound-proof chamber. The experiments were self-paced, and presentations of stimuli and data collection were controlled by a PC.

The experiment was conducted over three days. On the first day, the effect of white noise was assessed. First, /r/–/l/ words uttered by the male speaker were tested in one block under different SNR conditions (from 9 dB to −15 dB, as shown in Table 1). Words contrasting in /r/ and /l/ with all SNR conditions were presented in a random order. Second, /b/–/v/ words by same speaker were tested in the same manner as for /r/–/l/ words. Third, /s/–/th/ words by speaker 1 were tested in the same manner. Then, /r/–/l/, /b/–/v/, and /s/–/th/ words uttered by the female speaker were tested in the same manner as for the stimuli by the male speaker. On the second day, the effect of pink noise was assessed, and on the third day, the effect of the presentation level was assessed in the same manner as the first day's experiment.

3. RESULTS

3.1. Effect of Noise

Within each language group, the response for each stimulus was pooled across participants, and the correct response rates for each contrast (/r/–/l/, /b/–/v/, and /s/–/th/), each SNR condition (original and SNR conditions shown in Table 1), and each noise (white noise and pink noise) were calculated. Figure 1 shows the J listeners' correct response rates for each contrast as a function of SNR. Expectedly, the correct response rates for the original stimuli were very low, 0.66 for /r/–/l/, 0.70 for /b/–/v/, and 0.69 for /s/–/th/, when data from the white-noise series and pink-noise series were averaged. Interestingly, however, the correct response rate decreased as SNR decreased.

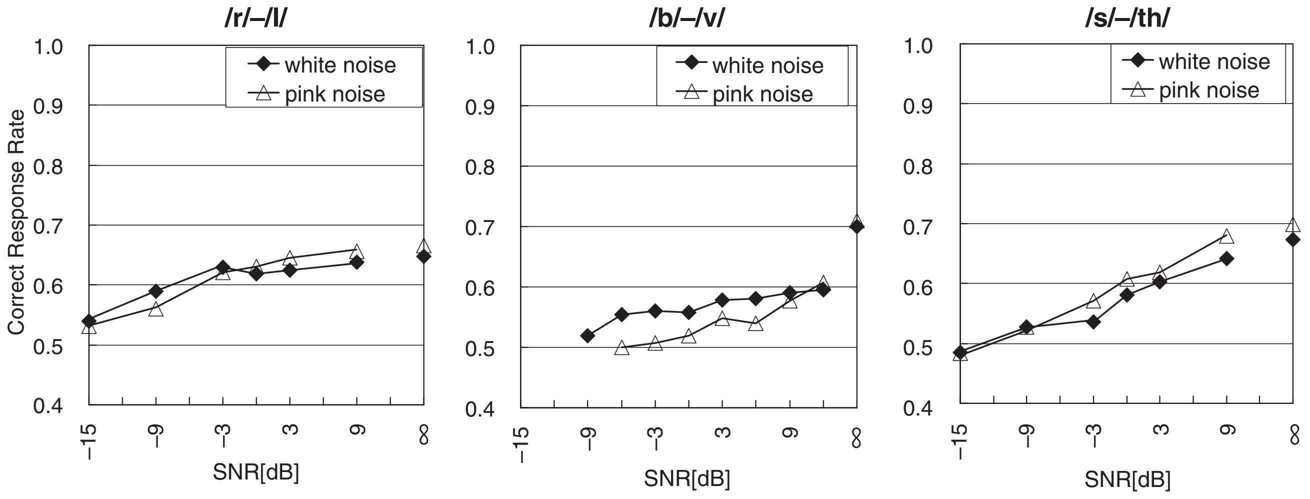

Fig. 1 J listeners' correct response rates for each contrast as a function of SNR. The SNR for the original signal is presented with ∞.

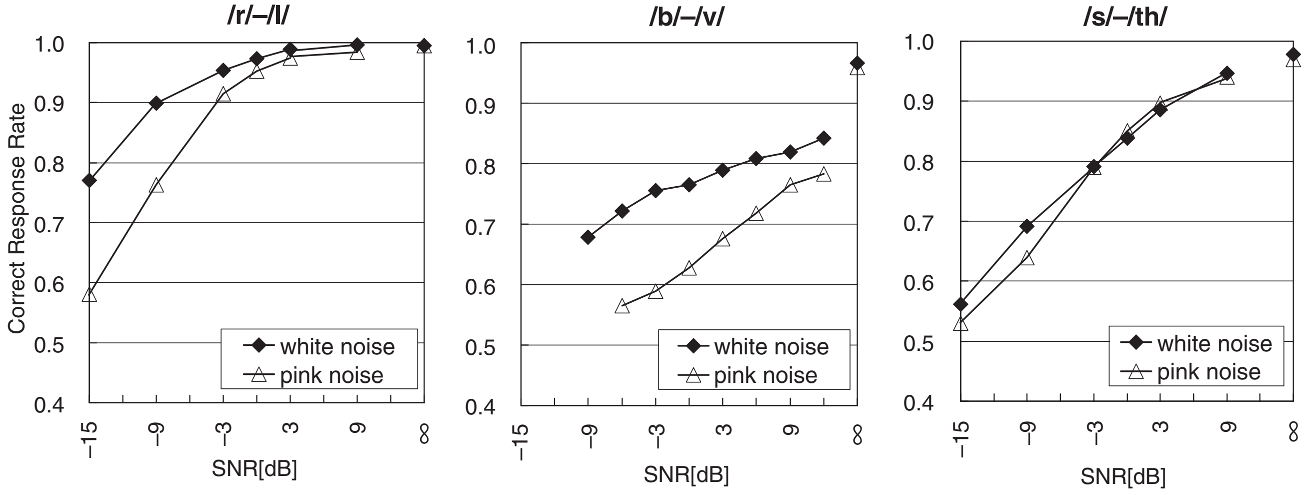

Fig. 2 AE listeners' correct response rates for each contrast as a function of SNR. The SNR for the original signal is presented with ∞.

AE listeners' response patterns are shown in Fig. 2. The responses for /r/–/l/ remained high until SNR of −3 dB, and decreased rapidly when SNR became smaller than −3 dB. In contrast, the effect of noise appeared at a higher SNR for /b/–/v/ or /s/–/th/ contrasts. For /r/–/l/ and /b/–/v/, the effect of pink noise was greater than that of white noise, although there was no difference in effects of white and pink noise for /s/–/th/.

The correct response rates for each phonetic contrast, noise, SNR condition, and language group were calculated. The arc-sine-transformed values were submitted to three separate ANOVAs by phonetic contrast. In each ANOVA, a three-factor design was used, in which noise (white and pink) and SNR conditions (Table 1) were within-subject variables, while language group (J and AE) was a between-subjects variable. For /b/–/v/ contrast, there were no data for pink noise under the −9 dB SNR condition. Thus, data for the −9 dB condition were not included in the /b/–/v/ analysis, in order to make the design factorial.

A significant interaction between language group and SNR was observed for /r/–/l/: $F(5, 105) = 88.870$, $p < 0.01$; /b/–/v/: $F(6, 126) = 8.345$, $p < 0.01$; and /s/–/th/: $F(5, 105) = 58.043$, $p < 0.01$. Other interactions were not significant. For all three contrasts, the main effects of language group (/r/–/l/: $F(1, 21) = 398.586$, $p < 0.01$; /b/–/v/: $F(1, 21) = 59.199$, $p < 0.01$; /s/–/th/: $F(1, 21) = 205.328$, $p < 0.01$) and SNR (/r/–/l/: $F(5, 105) = 217.236$, $p < 0.01$; /b/–/v/: $F(6, 126) = 41.763$, $p < 0.01$; /s/–/th/: $F(5, 105) = 250.844$, $p < 0.01$) were significant.

3.2. Effect of Presentation Level

Figures 3 and 4 show the correct response rates for the /r/–/l/, /b/–/v/, and /s/–/th/ stimuli for J listeners and AE listeners as a function of the sound presentation level.

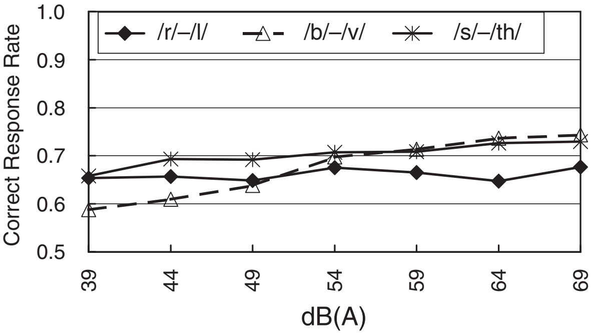

Fig. 3 J listeners' correct response rates for each contrast as a function of dB(A).

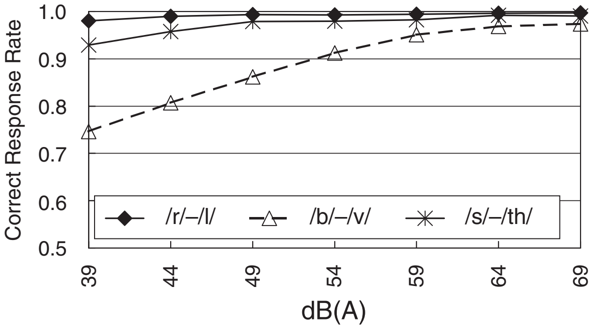

Fig. 4 AE listeners' correct response rates for each contrast as a function of dB(A).

Arc-sine-transformed correct response rates were submitted to a three-factor ANOVA in which phonetic contrast (/r/–/l/, /b/–/v/, and /s/–/th/) and presentation level (39, 44, 49, 54, 59, 64 and 69) were within-subject variables, while language group (J and AE) was a between-subjects variable.

A significant interaction between language group and presentation level was observed ($F(6, 126) = 5.142$, $p < 0.01$). Other interactions were not significant. The main effects of language group ($F(1, 21) = 123.285$, $p < 0.01$) and presentation level ($F(6, 126) = 34.439$, $p < 0.01$) were significant.

4. DISCUSSION

Results indicated that the perception of English words was affected by noise for both AE listeners of and J listeners. The rate of correct identification decreased when SNR decreased. The effect of noise differed by phonetic contrast and language group. For AE listeners, pink noise had a greater effect than white noise when SNRs were identical for /r/–/l/ and /b/–/v/, whereas the effects were similar for /s/–/th/. For J listeners, the effect of noise was weaker than for AE listeners, the function was gradual, and there were small differences in the effect between pink noise and white noise. These were possibly due to the flooring effect, because the J listeners' identification accuracy was very low, correct response rate of less than 0.7, even for original stimuli.

Interestingly, AE listeners' results showed notable differences with respect to phonetic contrast. The /r/–/l/ contrast was not affected by SNR of −3 dB, but when SNR was lower than that, perception was severely degraded. In contrast, /b/–/v/ and /s/–/th/ were affected even under higher SNR conditions, and the accuracy fell almost linearly as SNR decreased. Dominant acoustic characteristics of /b/, /v/, /s/, and /th/ are nonperiodic noise with rather broad spectra; a burst for /b/ and fricatives for /v/, /s/ and /th/. These consonants may be buried easily in broad-spectrum white noise or pink noise. In contrast, the acoustic difference between /r/ and /l/ is a frequency of the third formant (F_3), which is a periodic signal showing its of spectral peak in a rather low frequency band, between 1 kHz and 3 kHz. These consonants may be more tolerant to noise than consonants with broad spectra.

The effect of presentation level also differed auording to phonetic contrast and language group. AE listeners' perception was greatly affected by presentation level for /b/–/v/ contrast, while other contrasts were affected to only a small extent. J listeners' perception was also affected by the presentation level, but the effect was not large, possibly due to the flooring effect.

5. CONCLUSION

The present data showed the effect of noise and presentation level when listening to native phonetic contrasts and difficult non-native phonetic contrasts. It was demonstrated that the effect differed depending on phonetic contrast, even for native listeners. For the phonetic contrasts measured in this study, this difference can be explained in terms of the the acoustic characteristics of the signal and noise. It was also shown that non-native listeners were less affected by noise and presentation level. However, this result may be due to the flooring effect, because we examined phonetic contrasts, that are difficult for non-native listeners to differentiate. Further examination is necessary to clarify the general effect of noise and presentation level for non-native listeners.

ACKNOWLEDGEMENT

The authors are grateful to Professor T. Yamada for his helpful comments and M. Nishida for her assistance in conducting the experiments.

This research was conducted as part of "Research on Human Communication" with funding from the National Institute of Information and Communications Technology.

REFERENCES

[1] D. N. Kalikow, K. N. Stevens and L. L. Elliott, "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc.*

Am., **61**, 1337–1351 (1977).

[2] M. Florentine, “Speech perception in noise by fluent non-native listeners,” *Trans. Tech. Comm. Physiol. Acoust.*, H-85-16 (1985).

[3] S. J. van Wijngaarden, H. J. Steeneken and T. Houtgast, “Quantifying the intelligibility of speech in noise for non-native listeners,” *J. Acoust. Soc. Am.*, **112**, 3004–3013 (2002).

[4] K. Ueda, R. Komaki and R. Akahane-Yamada, “The effect of noise on /r/–/l/ phonemes perception,” *Proc. 65th Annu. Meet. Jpn. Psychol. Assoc.*, p. 120 (2001).

[5] G. A. Miller and P. E. Nicely, “An analysis of perceptual confusions among some English consonants,” *J. Acoust. Soc. Am.*, **27**, 338–352 (1955).

[6] R. Komaki, R. Akahane-Yamada and S. Katagiri, “Effect of native language on the production of second-language speech segments,” *Acoust. Sci. & Tech.*, **23**, 163–165 (2002).