

ACOUSTICAL LETTER

Identification of English voiceless fricatives in multispeaker babble noise by native Japanese and English listeners: Influence of English proficiency[1]

Hinako Masuda* and Takayuki Arai†

Department of Science and Technology, Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan

(*Received 7 December 2012, Accepted for publication 19 April 2013*)

Keywords: L2 perception, English, Voiceless fricatives, Background noise, Proficiency
PACS number: 43.71.+m [doi:10.1250/ast.34.356]

1. Introduction

Listening to speech in noisy environments is more difficult than in quiet environments, especially for non-native listeners, even if their performance in quiet environments does not fall far below that of native listeners [1–3]. Previous works on the perception of English vowels and consonants in various contexts such as /VC/, /CV/ [2], and /VCV/ [3] by native and non-native listeners have shown that non-native listeners' perception in quiet environments did not reach that of native listeners, and that the difference between the two listener groups increased when target sounds were presented in noisy listening environments. This tendency of non-native listeners applies not only to non-natives with intermediate-level proficiency but also to advanced-level learners [4–7].

Florentine (1985) [4] demonstrated that even highly fluent non-native listeners do not always perform as well as native listeners. Highly fluent non-native listeners and English native listeners took the SPIN test (Speech Perception in Noise [8]), in which participants were asked to listen to English monosyllabic nouns with high and low predictability in a quiet environment and in babble noise. When non-native listeners that achieved higher than 95% accuracy rates in the quiet environment, suggesting that they were highly proficient in English, proceeded to take the test in babble noise, they were unable to reach nativelike levels of performance. This result clearly shows the negative impact background noise has on the perception of foreign sounds regardless of the level of proficiency of the non-native listener.

Mayo *et al.* (1997) [5] carried out an experiment on the perception of monosyllabic English words in quiet and noisy listening environments by English native listeners and English-Spanish bilinguals who had acquired English as infants, toddlers, and after puberty. Their results showed that although early exposure to a second language improved foreign speech perception (i.e., the infant and toddler groups performed higher than the postpuberty group), even the performance of the infant group did not reach native listeners' scores when sounds were presented in noise.

Another study on bilinguals by Rogers *et al.* (2006) [6] examined the perceptual ability of English monosyllabic words in noisy and reverberant listening environments by English native listeners and English-Spanish bilinguals. The results showed that the bilinguals' performance in adverse listening environments fell short of that of the native listeners with significant differences, even though the bilinguals attained perfect scores in a quiet environment. The bilingual participants were first exposed to Spanish from birth, and then to English before the age of six. This implies that even early bilinguals do not perform as well as native listeners in adverse environments and are likely to be influenced the language they were first exposed to.

The perception of English consonants by Japanese listeners has been frequently examined. A well-known case is their difficulty to perceive English /ɹ/ and /l/ [7,9–11]. Adachi *et al.* (2006) [9] and Ueda *et al.* (2007) [10] compared the ability to identify /ɹ/ and /l/ by Japanese and English native listeners in a quiet environment and in background noise. They found that English native listeners were able to identify the sounds with perfect scores in the quiet environment and that their performance degraded to approximately 70% in background noise. The Japanese native listeners' performance, on the other hand, fell far below that of the English native listeners: they were only able to perceive the two sounds with approximately 65% [9] and 70% [10] performance even in a quiet environment, which decreased to approximately 55% in background noise of SNR (signal-to-noise ratio) $= -15$ dB [9] and SNR $= -21$ dB [10]. However, Akahane-Yamada *et al.* (1996) [11] claimed that the difficulty in perceiving /ɹ/ and /l/ by Japanese listeners can be improved by perceptual training, and that perceptual training also has a positive effect on the production of the two sounds.

The authors examined how English proficiency and the amount of noise affects the identification of English /ɹ/ and /l/ in a quiet environment and in background noise by Japanese and English native listeners [7], which is part of the data set used from the recordings of stimuli in the present study. Non-native Japanese participants were divided into intermediate- and advanced-level groups, and were presented with the target sounds in a quiet environment and in multispeaker babble noise at SNR $= 10$ dB, 5 dB, and 0 dB. Their results showed that while advanced-level learners had nativelike performance for the perception of /ɹ/, the results

[1]This paper includes our previous work, Masuda and Arai "Perception of English voiceless fricatives by Japanese and English native listeners under various signal-to-noise ratios," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 471–474 (2011). Additional analyses are introduced.

*e-mail: h-masuda@sophia.ac.jp
†e-mail: arai@sophia.ac.jp

for /l/ became more like those of the intermediate-level learners as the amount of background noise increased. This result suggests that it is important to take the non-native population's proficiency into consideration as their performance varies, as well as using a variety of SNRs, to examine the differences in how performance degrades among participants with varying proficiency.

Another difficulty that Japanese native listeners often face is the perception of English fricatives. This difficulty occurs because the number of fricatives in Japanese is fewer than that in English. English has a total of nine fricatives /f v θ ð s z ʃ ʒ h/ [12] while Japanese has three /s z h/ [13].* Lambacher *et al.* (2001) [17] performed a five-alternative, forced-choice (5AFC) perceptual experiment on Japanese and English native listeners to identify English voiceless fricatives with five vowel contexts in a quiet environment. Results revealed that Japanese listeners achieved high accuracy rates with an overall average rate of 74%, with /ʃ/ being the highest at approximately 88% and /θ/ the lowest at approximately 55%. The English native listeners scored an average rate of approximately 94%, with /f/ being the lowest at approximately 86% and /s ʃ/ being the highest at approximately 97%. By comparing the results of the two listener groups, we can speculate that the result of the Japanese listeners is influenced by the phonological system of their native language, i.e., Japanese listeners had difficulty in identifying /θ/ because the phoneme does not exist in their native language.

An experiment previously conducted by the authors compared the perception of English voiceless fricatives by Japanese native listeners with intermediate- and advanced-level English proficiency in a quiet environment and in two types of background noise (multispeaker babble noise and white noise, both at SNR = 0 dB), and found a significant trend between the two listener groups' accuracy rates [18]. The participants in their study attained lowest scores for /θ/ in the quiet environment (86% for advanced learners and 70% for intermediate learners), in agreement with the results of Lambacher *et al.* However, they concluded that using only one SNR condition is insufficient for accurate measurement of the influence of background noise in the two learner groups, and suggested that various SNRs should be taken into account in listening environments in order to observe the effect of proficiency on the perception of foreign sounds in background noise. Thus, here we report the results of the perception of English voiceless fricatives in a quiet environment and in multispeaker babble noise at SNR = 10 dB, 5 dB, and 0 dB.

In summary, perceiving foreign speech sounds in adverse listening environments such as in background noise is difficult for non-native listeners, regardless of their proficiency in the target language. Although many studies exist on the perception of second-language sounds by non-native listeners, there are still unknown issues regarding the perception of second-language consonants in noise by non-native listeners with varying second-language proficiency. The perceptual performance of advanced-level learners is especially difficult to investigate because there are no set criteria for defining what advanced level refers to [19]. We are particularly interested in how performance differs between advanced-level learners and native listeners in background noise. In this study, we aim to observe the perception of English voiceless fricatives by English and Japanese native listeners, taking the Japanese listeners' English proficiency into account. Moreover, the perceptual environment includes four conditions: 1) no noise, multispeaker babble noise at 2) SNR = 10, 3) SNR = 5 dB, and 4) SNR = 0 dB, in order to observe not only the effect of the proficiency of the listeners but also the amount of background noise.

We address two research questions: 1) What is the effect of the amount of multispeaker babble noise in perceiving voiceless English fricatives?, and 2) What is the impact of English proficiency on speech perception? The final goal of our research is for the results obtained from the present perceptual experiment to contribute to the fields of second-language acquisition and second-language pedagogy, including CALL (computer-assisted language learning) systems, particularly in developing materials for English proficiency-based perceptual training containing background noise for non-native Japanese listeners. We hope to capture characteristics specifically common to Japanese listeners by looking into not only overall correct rates but also the confusion between consonants encountered by learners. The use of background noise with various SNRs will also enable us to understand the mechanism of speech perception by learners with different proficiencies.

Table 1 Data of participants.

	Intermediate learners	Advanced learners	English native listeners
Number of participants	$N = 8$	$N = 12$	$N = 6$
Mean age (range)	23.0 (20–31)	26.7 (20–35)	20.8 (20–21)

2. Perceptual experiment

2.1. Participants

Twenty-six listeners participated in the experiment: 20 Japanese and six English native listeners (see Table 1). Among the 20 Japanese native listeners, 12 participants were grouped as advanced learners of English, who had achieved a score higher than 850 in TOEIC® (Test of English for International Communication provided by ETS) or equivalent scores in TOEFL® (Test of English as a Foreign Language provided by ETS) and/or were placed in advanced-level English classes at a university in Japan. The remaining eight Japanese participants were grouped as intermediate-level learners of English, who had achieved a score lower than 650 in TOEIC® and/or were placed in an intermediate-level English class at a university in Japan. Participants did not have experience of living abroad and had studied English at a junior high school in Japan from the age of twelve. None of the participants reported any hearing problems.

*There are phonetically a total of seven fricatives in Japanese [ɸ s z ɕ ʑ ç h] [14], and the occasional occurrence of [β ð ɣ] in rapid speech as allophones of /b d g/ [15,16].

2.2. Stimuli

Twenty-three consonants /b ʧ d f g h ʤ ʒ k l m n p ɹ s ʃ t θ ð v w j z/ were embedded in a /ɑ __ ɑ/ context. Five English voiceless fricatives /f h s ʃ θ/ were selected for analysis. The speaker of the stimuli was a female Japanese-English bilingual speaker. The stimuli were recorded in a soundproof room using a digital sound recorder (Marantz PMD 660) and a microphone (Sony ECM-23F5) at a sampling frequency of 48 kHz. The stimuli were later downsampled to 16 kHz. Stimuli were presented in the order of 1) multispeaker babble noise (SNR = 0 dB, 5 dB, 10 dB), and 2) no noise. The stimuli embedded in noise were presented to the listeners in random order. Multispeaker babble noise was taken from NOISEX [20]. Multispeaker babble noise was selected as the background noise for the present experiment because it resembles a real-life environment in which second-language learners may experience difficulties in foreign language perception. The stimuli were preceded and followed by one second of noise. All experimental procedures were carried out using the computer program Praat [21].

2.3. Procedure

A laptop computer was used to present the stimuli and to record the listeners' responses. Participants were presented with the stimuli through a USB audio amplifier (Onkyo MA-500U) and headphones (Stax SR-303 or Stax SRM-323A). The laptop computer and audio amplifier were digitally connected via a USB interface.

All participants were given 23 practice trials (18 in noise and 5 without noise). The practice trials were not scored nor were any feedback given. After the practice trials, the participants proceeded to the main experiment, in which 460 trials were presented (345 in multispeaker babble noise and 115 in a quiet environment). They were asked to listen to each stimulus and to choose the consonant that most closely fitted to what they heard from a table of 23 consonants (see Fig. 1).

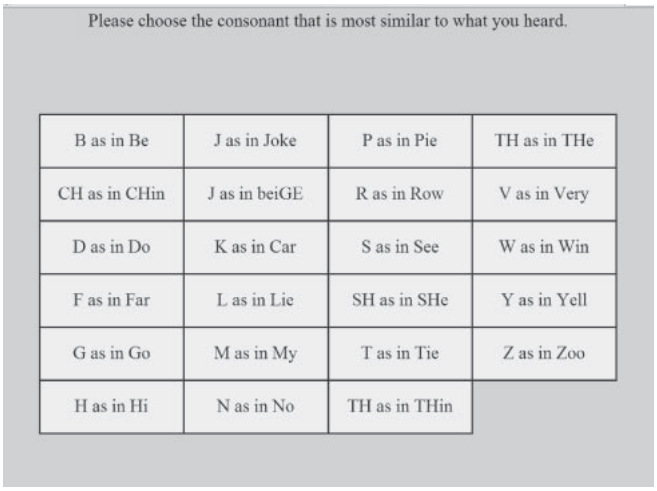

Fig. 1 Experimental interface [2].

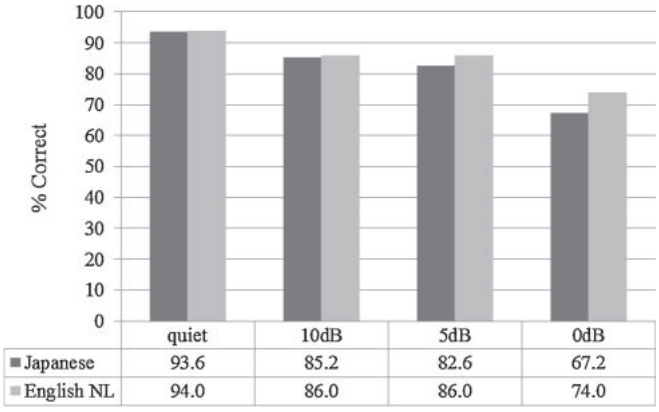

Fig. 2 Average correct rates of voiceless fricatives for Japanese and English native listeners in quiet environment and in noise.

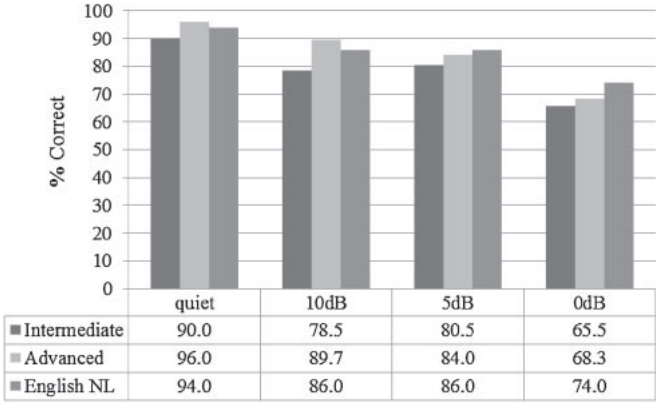

Fig. 3 Average correct rates of voiceless fricatives for intermediate learners, advanced learners, and English native listeners in quiet environment and in noise.

3. Results

3.1. Average correct rates

The combined average correct rates for both Japanese intermediate- and advanced-level learners and those for English native listeners are shown in Fig. 2. Overall, the English native listeners achieved higher correct rates than the Japanese listeners under all conditions. The performance degraded as background noise increased for both groups. The two-factor factorial ANOVA showed a significant main effect of listening conditions [$F(3, 96) = 19.9$, $p < 0.001$] but not listener groups [$F(1, 96) = 1.06$, $p = 0.30$]. *Post hoc* comparisons using the Tukey-Kramer test revealed significant differences in the listening conditions of Quiet and SNR = 5 dB ($p < 0.05$), Quiet and SNR = 0 dB ($p < 0.01$), SNR = 10 dB and SNR = 0 dB ($p < 0.01$), and SNR = 5 dB and SNR = 0 dB ($p < 0.01$). The interaction of the two factors was not significant [$F(3, 96) = 0.28$, $p = 0.83$].

Figure 3 shows a detailed graph of the average correct rates of intermediate learners, advanced learners, and English native listeners. Listeners' performance degraded as background noise increased. The two-factor factorial ANOVA showed a significant main effect of both listening conditions [$F(3, 92) = 20.38$, $p < 0.001$] and listener groups [$F(2, 92) = 2.96$, $p = 0.05$]. The interaction of the two factors was not significant [$F(6, 92) = 0.39$, $p = 0.87$]. *Post hoc* comparisons using the Tukey-Kramer test showed significant trends in

Table 2 Intermediate learners' confusion matrix for SNR = 0 dB (%).

Stimuli \ Response	f	h	s	ʃ	θ	p	Others
f	65.0			2.5	10.0	12.5	10.0 (b)
h	35.0	55.0			2.5	5.0	2.5 (b)
s			90.0	5.0	5.0		
ʃ				85.0	2.5		5.0 (t), 7.5 (tʃ)
θ	57.5				35.0		5.0 (t), 2.5 (ð)

Table 3 Advanced learners' confusion matrix for SNR = 0 dB (%).

Stimuli \ Response	f	h	s	ʃ	θ	p	Others
f	70.0	1.7			11.7	15.0	1.7 (l)
h	46.7	41.7				11.7	
s			90.0		5.0		5.0 (ð)
ʃ				91.7			6.7 (tʃ), 1.7 (ʒ)
θ	38.3		1.7		56.7	1.7	1.7 (ð)

Table 4 English native listeners' confusion matrix for SNR = 0 dB (%).

Stimuli \ Response	f	h	s	ʃ	θ	p	Others
f	70.0					30.0	
h	16.7	66.7				16.7	
s			100				
ʃ			3.3	96.7			
θ	56.7				36.7	3.3	3.3 (ð)

the listener groups Intermediate and Advanced ($p = 0.063$), Intermediate and NL ($p = 0.075$), and Advanced and NL ($p = 0.069$), and significant differences in the listening conditions of Quiet and SNR = 5 dB ($p < 0.05$), Quiet and SNR = 0 dB ($p < 0.01$), SNR = 10 dB and SNR = 0 dB ($p < 0.01$), and SNR = 5 dB and SNR = 0 dB ($p < 0.01$).

3.2. Confusion matrices

Consonants were confused most often under the SNR = 0 dB condition, as indicated by the lowest correct rates in all listener groups. Confusion matrices in the case of SNR = 0 dB are shown in Tables 2–4 to illustrate the similarities and differences in consonant confusion among the three listener groups. Rows represent the stimuli presented to the participants, and columns represent the participants' responses.

4. Discussion and conclusion

The Japanese non-native and English native groups' average correct rates showed no statistically significant differences. However, significant differences were observed among listener groups when the non-native listeners were further categorized into subgroups of intermediate- and advanced-level learners. To look further into the differences among listener groups, we performed additional analysis of the consonant confusion patterns.

The confusion matrices showed both similarities and differences among the three groups. All three listener groups had difficulty perceiving /θ/ and /h/ at SNR = 0 dB, which showed that the difficulty in identifying /θ/ is not limited to Japanese listeners. For all listener groups, the rate of correct identification of sibilants /s/ and /ʃ/ was high, and difficulty in identification was observed for the nonsibilants (/f h θ/). The confusion patterns showed that for all listener groups, nonsibilants were rarely confused as sibilants.

Table 5 Range of intensity (in dB) of the target consonants with 20 babble noise segments at SNR = 0 dB. The "current" levels (in dB) show the intensity of the babble noise segment used in the experiment.

	f	h	s	ʃ	θ
range (dB)	−1.7–1.6	−1.2–1.4	−1.2–1.7	−0.7–1.3	−1.1–1.4
current (dB)	1.4	0.8	1.3	1.3	1.4

The most common confusion for /θ/ was with /f/ for all listener groups. This result is not in agreement with a previous study [17], in which Japanese native listeners confused /θ/ with both /s/ and /f/ even in a quiet listening environment; the confusion of /θ/ with /s/ was rarely observed for the participants in the present study. In the case of /h/, its confusion with /f/ was observed in all three listener groups, but its confusion with /p/ increased with decreasing English proficiency. Although the confusion of /f/ with /p/ was seen in all three listener groups, the confusion of /f/ with /θ/ was only observed for intermediate and advanced learners.

In the current experiment, we used an identical babble noise segment for all experimental stimuli. However, there was no guarantee that the babble noise segment added to the target consonant had the mean intensity because babble noise fluctuates over time. Therefore, we measured the intensity of the target consonant with the babble noise when the total stimulus SNR was 0 dB. Twenty different babble noise segments were used to measure the distribution of the intensity for the five consonants. Table 5 shows the minimum and maximum levels relative to the mean in decibels ("range" in this table). The ranges were approximately two to three decibels for all consonants. This table also shows the intensity of the babble noise segment (with the target consonants) used in the experiment ("current" in Table 5); the "current" levels were 0.8 to 1.4 dB higher than the mean levels owing to the temporal fluctuation of the noise. This means that the babble noise conditions used in the current experiment were slightly more severe than the average condition but the difference was not large.

The detailed analysis in the present study showed both similarities and differences in the identification of English voiceless fricatives by the three listener groups, and confirmed the importance of observing confusion patterns and considering second-language proficiency. Defining second-language proficiency, however, is difficult and depends on context. Although the present study adopted TOEIC/TOEFL scores to measure the learners' English proficiency, there is no universally agreed definition of what intermediate- or ad-

vanced-level learners are [19]. Further analysis is needed regarding learners' language background such as the age of acquisition and experience of living in English-speaking countries.

References

[1] Y. Takata and A. Nabelek, "English consonant recognition in noise and in reverberation by Japanese and American listeners," *J. Acoust. Soc. Am.*, **88**, 663–666 (1990).

[2] A. Cutler, A. Weber, R. Smits and N. Cooper, "Patterns of English phoneme confusions by native and non-native listeners," *J. Acoust. Soc. Am.*, **116**, 3668–3678 (2004).

[3] M. L. Garcia Lecumberri and M. Cooke, "Effect of masker type on native and non-native consonant perception in noise," *J. Acoust. Soc. Am.*, **119**, 2445–2454 (2006).

[4] M. Florentine, "Non-native listeners' perception of American-English in noise," *Proc. Inter-noise 85*, pp. 1021–1024 (1985).

[5] L. H. Mayo, M. Florentine and S. Buus, "Age of second-language acquisition and perception of speech in noise," *J. Speech Lang. Hear. Res.*, **40**, 686–693 (1997).

[6] C. L. Rogers, J. J. Lister, D. M. Febo, J. M. Besing and H. B. Abrams, "Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing," *Appl. Psycholinguist.*, **27**, 465–485 (2006).

[7] H. Masuda and T. Arai, "Perception of /r/ and /l/ in quiet and multi-speaker babble noise by Japanese and English native listeners," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 477–480 (2012).

[8] 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).

[9] T. Adachi, R. Akahane-Yamada and K. Ueda, "Intelligibility of English phonemes in noise for native and non-native listeners," *Acoust. Sci. & Tech.*, **27**, 285–289 (2006).

[10] K. Ueda, R. Akahane-Yamada, R. Komaki and T. Adachi, "Identification of English /r/ and /l/ in noise: The effects of baseline performance," *Acoust. Sci. & Tech.*, **28**, 251–259 (2007).

[11] R. Akahane-Yamada, Y. Tohkura, A. R. Bradlow and D. B. Pisoni, "Does training in speech perception modify speech production?" *Proc. 4th Int. Conf. Spoken Language Processing*, pp. 606–609 (1996).

[12] P. Carr, *English Phonetics and Phonology: An Introduction* (Blackwell Publishers, Oxford, 1999).

[13] H. Okada, "Japanese," *The Handbook of the International Phonetic Association: A Guide to Use the International Phonetic Alphabet* (Cambridge University Press, Cambridge, 1999), pp. 117–119.

[14] T. J. Vance, *An Introduction to Japanese Phonology* (State University of New York Press, New York, 1987).

[15] T. Arai, N. Warner and S. Greenberg, "Analysis of spontaneous Japanese in a multi-language telephone-speech corpus," *Acoust. Sci. & Tech.*, **28**, 46–48 (2007).

[16] S. Kawakami, *Nihongo Onsei Gaisetsu* (Ofusha, Tokyo, 1977) (in Japanese).

[17] S. Lambacher, W. Martens, B. Nelson and J. Berman, "Identification of English voiceless fricatives by Japanese listeners: The influence of vowel context on sensitivity and response bias," *Acoust. Sci. & Tech.*, **22**, 334–342 (2001).

[18] H. Masuda and T. Arai, "Perception of voiceless fricatives by Japanese listeners of advanced and intermediate level English proficiency," *Proc. INTERSPEECH*, pp. 1866–1869 (2010).

[19] F. Grosjean and P. Li, *The Psycholinguistics of Bilingualism* (Wiley-Blackwell, Oxford, 2013).

[20] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, **12**, 247–251 (1993).

[21] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer [Computer program]," Retrieved from http://www.praat.org/