Perception of English voiceless fricatives by Japanese and English native listeners under various signal-to-noise ratios *

☆Hinako Masuda and Takayuki Arai
(Graduate School of Science and Technology, Sophia University)

1 Introduction

Speech perception in noise or reverberant listening conditions is difficult for everyone, especially for non-native listeners [1-3], even if difference in the performance of the native and non-native listeners may be relatively small in quiet listening condition [1]. Garcia Lecumberri & Cooke (2006) [3] examined the effect of background noise on the perception of English intervocalic consonants by English native and non-native (Spanish) listeners. They found that non-native listeners did not reach native-like performance in quiet listening condition, and that the gap between the two listener groups grew larger with the presence of background noise.

Lambacher *et al.* (2001) [4] conducted a perceptual experiment on the identification of English voiceless fricatives /f, s, ʃ, θ, h/ in five vowel contexts /i, ɛ, a, o, u/ and three syllable positions (CV, VC, and VCV) on Japanese and English native listeners under quiet listening condition. English voiceless fricatives are difficult for Japanese native listeners to perceive, because the phonemes /f, θ/ do not exist in Japanese. The overall correct rate of Japanese native listeners was around 74%, whereas English native listeners scored around 94%. Their results also showed that Japanese native listeners did not reach native-like performance under quiet listening condition. However, in many of the studies concerning non-native speech perception, the second language proficiency of the non-native listeners has been around intermediate level.

Studies concerning non-native listeners with advanced level proficiency, Rogers *et al.* [5] compared the perceptual ability by American English native listeners and Spanish-English bilingual listeners who were exposed to English before age six. Although they found a significant difference between native and bilingual listeners in perceiving English monosyllabic words embedded in noise and reverberation, the two groups' performances were similar in quiet listening condition. Additionally, Mayo *et al.* [6] examined the perception of monosyllabic words with high and low predictability by English-Spanish bilingual listeners, and found that early exposure to a second language is advantageous in perceiving second language sounds in noise. However, they reported that even bilinguals who had been exposed to a second language since infancy did not reach native-level performance.

In sum, second language perception in noise and reverberant listening conditions are difficult for non-native listeners, but the perception in quiet listening condition may reach native-like level, especially for bilinguals. However, the tendencies seem to vary with the non-native population. The focus of the present study is to examine how second language proficiency affects perception of foreign sounds in both quiet and noisy listening conditions. The aims of the experiment are: 1) to investigate the difference in the perception of English consonants by Japanese and English native listeners in quiet and noisy listening conditions, 2) to examine the influence of second language proficiency, and 3) to compare the differences in the consonant confusions among listener groups.

* Perception of English voiceless fricatives by Japanese and English native listeners under various signal-to-noise ratios, by MASUDA, Hinako and ARAI, Takayuki (Sophia University).

2 Perceptual experiment

2.1 Participants

Twenty Japanese native listeners and six English native listeners participated in the perceptual experiment. Among the twenty Japanese native listeners, twelve participants were advanced learners of English, who had achieved higher than 850 on TOEIC® examination [7] or achieved equivalent scores on TOEFL® examination [8], and/or were placed in advanced level English classes at a university in Japan. The remaining eight Japanese participants were intermediate level learners of English, who had achieved below 650 on TOEIC® examination, and/or were placed in intermediate level English class at a university in Japan. Participants who do not have experience of living abroad received English education from age twelve at junior high schools in Japan. None of the participants reported any hearing problems. The data of participants are shown in Table 1.

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)

2.2 Stimuli

Twenty-three consonants /b, tʃ, d, f, g, h, dʒ, ʒ, k, l, m, n, p, ɹ, s, ʃ, t, θ, ð, v, w, j, z/ were embedded in /ɑ __ ɑ/ context, and were presented to the listeners. Five English voiceless fricatives /f, h, s, ʃ, θ/ were selected for analyses of correct rates (Figures 2 and 3) and confusion matrices (Tables 2-4). The speaker of the stimuli is a female Japanese-English bilingual speaker. The stimuli were recorded in a sound-proof 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. The order of the SNR was randomized. Multi-speaker babble noise was taken from NOISEX [9]. Multi-speaker babble noise was selected as noise because it resembles real-life environment that second language learners may experience difficulties in foreign language perception. The stimuli were preceded and followed by 1 second of noise.

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 USB audio amplifier (ONKYO MA-500U) and headphones (STAX SR-303 and STAX SRM-323A). The laptop computer and audio amplifier were digitally connected via USB interface.

All participants were given 23 practice trials (eighteen in noise, and five in quiet). The practice trials were not scored. After the practice trials, participants proceeded to the main experiment where 460 trials were presented (345 in multi-speaker babble noise and 115 in quiet). Stimuli were presented in the order of 1) multi-speaker babble noise (SNR = 0 dB, 5 dB, 10 dB), and 2) quiet. They were asked to listen to each stimulus and to choose the best consonant that fits to what they heard from the table of 23 consonants (see Figure 1). All experimental procedures were conducted by using Praat (Boersma *et al.*) [10].

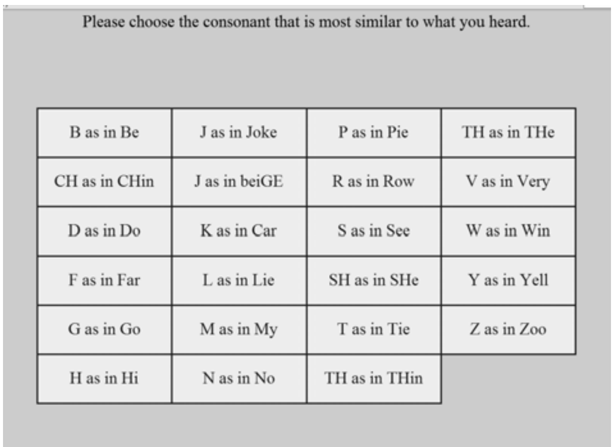

Figure 1 Experimental interface
(words extracted from Cutler *et al.*, 2004)

3 Results

3.1 General results

The average percentages of correct responses by Japanese native listeners (both intermediate and advanced learners) and English native listeners are shown in Figure 2. Although the difference between Japanese and English native listeners became most apparent at SNR = 0 dB, there was no significant difference between Japanese and English native listeners in any of the listening conditions (t-test, $p < 0.1$).

The results of intermediate learners, advanced learners, and English native listeners are shown in Figure 3. All three groups performed the best in quiet listening condition, and the performance degraded as the amount of noise increased. There was a significant difference between intermediate and advanced learners in the SNR = 10 dB condition (t-test, $p = 0.04$).

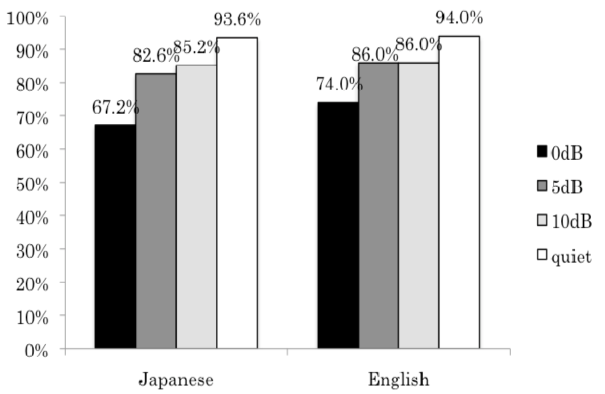

Figure 2 Average percentages of correct responses by Japanese and English native listeners in quiet and in noise

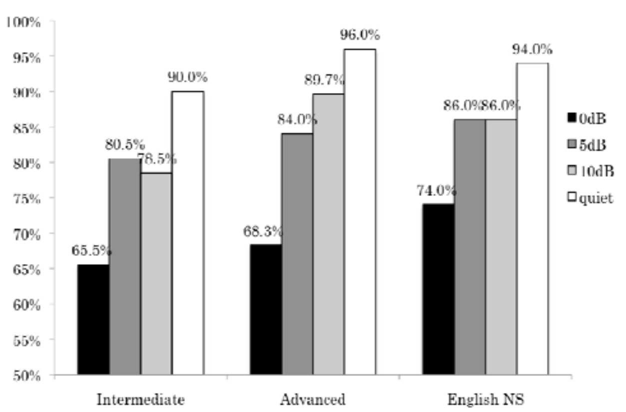

Figure 3 Average percentages of correct responses by the three listener groups in quiet and in noise

3.2 Confusion matrices in SNR = 0 dB

Overall correct rates were lowest in SNR = 0 dB for all three listener groups, thus the present paper reports only the matrices of that condition. Consonant confusion matrices of the five English voiceless fricatives in multi-speaker babble noise of SNR = 0 dB are calculated into percentages and are shown in Tables 2 to 4. Rows represent the stimuli presented to the participants, and columns represent the participants' responses.

Table 2 Consonant confusion matrix in multi-speaker babble noise of SNR = 0 dB for intermediate learners (%)

	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 Consonant confusion matrix in multi-speaker babble noise of SNR= 0 dB for advanced learners (%)

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

Table 4 Consonant confusion matrix in multi-speaker babble noise of SNR= 0 dB for English native listeners (%)

	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)

4 Discussion

The present study reported the results of the perception of English consonants in quiet and in multi-speaker babble noise (SNR = 10 dB, 5 dB, 0 dB) by twenty Japanese native listeners and six American English native listeners. The results showed that there were no significant differences in any of the listening conditions, thus implying that perception of English voiceless fricatives is difficult for English native listeners as well. The results do not correspond to those of Lambacher *et al.* [4] since the Japanese native listeners in their experiment performed significantly lower compared to English native listeners. However, some of the confusions that the Japanese native listeners in their experiment made were the same as those made by Japanese native listeners in the present study, such as the confusion of /s/ as /θ/ and /h/ as /f/.

As for the native and non-native comparison in the present experiment, the confusion patterns of Japanese and English native listeners had both similarities and dissimilarities although the overall correct rates in each listening condition was similar for both native and non-native listener groups. The confusions of /f/ as /p/, /h/ as /f/, and /θ/ as /f/ were observed in both Japanese and English native listeners, thus indicating that such confusions may be universal. Dissimilarities were observed in the confusions of /f/ as /θ/, /ʃ/ as /tʃ/, and /s/ as /θ/ which were seen only in Japanese native listeners, and Japanese confused /h/ as mostly /f/ while native listeners confused /h/ as /f/ as well as /p/. The advantage of higher second language proficiency was observed only in the percentages, and not in the confusion patterns.

5 Conclusion

The present experiment revealed that 1) English native listeners performed higher than Japanese native listeners especially under multi-speaker babble noise of SNR = 0 dB, 2) second language proficiency had a positive effect on the overall correct rates, but not on the confusion patterns, and 3) further analyses on the confusion patterns revealed that confusions of /f/ as /p/, /h/ as /f/, and /θ/ as /f/ were universal to both Japanese and English native listeners, but the confusions of /f/ as /θ/, /ʃ/ as /tʃ/, and /s/ as /θ/ were unique to Japanese native listeners. However, the results of the present experiment do not provide strong evidence due to small number of trials and participants. Confusion patterns may become more salient with increased number of trials.

Acknowledgments

The authors would like to thank Professor Shigeaki Amano for his insightful comments. This study work was partially supported by Sophia University Open Research Center from MEXT.

References

[1] Takata & Nabelek, J. Acoust. Soc. Am., 88 (2), 663-666, 1990.
[2] Cutler *et al.*, J. Acoust. Soc. Am., 116 (6), 3668-3678, 2004.
[3] Garcia Lecumberri & Cooke, J. Acoust. Soc. Am., 119 (4): 2445-2454, 2006.
[4] Lambacher *et al.*, Acoust. Sci. & Tech., 22 (5), 334-342, 2001.
[5] Rogers *et al.*, Applied Psycholinguistics, 27, 465-485, 2006.
[6] Mayo *et al.*, JSLHR, 40, 686-693, 1997.
[7] TOEIC® (Test of English for International Communication), provided by ETS (Educational Testing Service).
[8] TOEFL® (Test of English as a Foreign Language), provided by ETS (Educational Testing Service).
[9] Varga & Steeneken, Speech Communication, 12 (3), 247-251, 1993.
[10] Boersma and Weenink, Praat: doing phonetics by computer (Version 5.0.09) [Computer program].