



Perception of voiceless fricatives by Japanese listeners of advanced and intermediate level English proficiency

Hinako Masuda, Takayuki Arai

Graduate School of Science and Technology, Sophia University, Tokyo, Japan

h-masuda@sophia.ac.jp, arai@sophia.ac.jp

Abstract

Numerous research has investigated how first language influences the perception of foreign sounds. The present study focuses on the perception of voiceless English fricatives by Japanese listeners with advanced and intermediate level English proficiency, and compares their results with that of English native listeners. Listeners identified consonants embedded in /a __ a/ in quiet, multi-speaker babble and white noise (SNR=0 dB). Results revealed that intermediate level learners scored the lowest among all listener groups, and /th-/s/ confusions were unique to Japanese listeners. Confusions of /th-/f/ were observed among all listener groups, which suggest that those phoneme confusions may be universal.

Index Terms: Japanese, English consonants, English proficiency, non-native, speech perception

1. Introduction

Learning a foreign language is more or less troublesome for everyone. The difficulty of speech perception for non-native listeners is caused by several reasons such as the starting age of foreign language learning, the quality and quantity of the input in the target language, the number of years being exposed to the target language, first language influence, and many more. Among those, numerous studies in the past have investigated the influence of native phoneme knowledge on the perception of foreign language sounds [1-3].

Garcia Lecumberri & Cooke [1] reported that non-native listeners (Spanish) were not able to identify a certain set of English consonants in quiet, and that the difference between native and non-native listener groups grew larger when stimuli were embedded in babble, speech-shaped, and competing English and Spanish speech noise. Their results also revealed that babble noise was the most difficult listening condition for both native and non-native listeners. Influence of L1 was observed in consonant confusion, in which they claim that it may be due to the differences in the phonetic cues used for phoneme perception, and that the cues of non-native listeners may not have been fully developed.

Cutler *et al.* [2] compared the identification abilities of English vowels and consonants in quiet and babble noise by native and non-native (Dutch) listeners. The results revealed that non-native listeners scored lower than native listeners in both listening conditions, although, contrary to Garcia Lecumberri & Cooke's findings, the disadvantages of non-native listeners were not affected by signal-to-noise ratios. However, both positive and negative L1 influence was observed for the non-native listeners. Non-native listeners' performance was similar to native listeners when the target phonemes were presented at similar syllable positions as in their native language (Dutch), whereas the perception of phonemes that do not exist in Dutch were significantly worse compared to native listeners.

In the case of Japanese native listeners perceiving English sounds, Lambacher *et al.* [3] conducted a five-alternative, forced-choice test of five English voiceless fricatives /f, s, ʃ, θ, h/ in five vowel contexts /i, ε, 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 high at around 74%, whereas English native listeners scored around 94%. Their results showed that non-native listeners' performance did not reach native-like level even under quiet listening condition. However, the Japanese participants had little or no exposure to training in English pronunciation or listening in school, although all listeners had six years of a typical grammar-based English education at school. This suggests that the level of English proficiency may have been around beginner to intermediate level.

Heretofore described was mainly non-native speech perception regarding L1 monolinguals or L2 intermediate learners. Concerning non-native listeners with advanced level proficiency, Rogers *et al.* [4] compared American English native listeners with Spanish-English bilinguals 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. Furthermore, Mayo *et al.* [5] investigated the perception of monosyllabic words with high and low predictability by English-Spanish bilinguals, and claimed 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.

The two studies on bilingual listeners [4,5] suggest that although they may have native-like performance under a quiet listening condition, the gap between non-native and native listeners are likely to grow larger with the presence of noise and reverberation. However, the term "bilingual" is ambiguous, and, as Grosjean [6] states, factors associated with bilinguals are often diverse, insufficient or controversial. For that, and language processing of bilinguals being so complex, studies on bilinguals are still scarce, and many aspects remain unknown.

The aims of the present experiment is to investigate 1) the difference in the perceptual ability of voiceless English fricatives by Japanese native listeners with advanced and intermediate level English proficiency, 2) the effect of background noise on the perception of voiceless fricatives by Japanese native listeners with advanced and intermediate level English proficiency, and 3) whether advanced learners perform similar to English native listeners. The Japanese participants in the present study are Japanese native speakers

with advanced and intermediate-level English proficiency. The present study will set TOEIC® scores [7] as the main criteria for assessing participants' English level, and will label the higher assessed, "advanced level learners", and will not use the term "bilinguals" in order to avoid misinterpretation.

2. Experiment

2.1. Participants

The 19 listeners in the present experiment are divided into three groups according to their language background. 17 Japanese native listeners participated as the non-native group, and two English native listeners participated as the native control group. Japanese native listeners are divided into two groups regarding their English level. The data of participants are shown in Table 1.

2.1.1. Intermediate level learners of English

Ten intermediate level English learners (6 female, 4 male) participated in the experiment. All are native listeners of Japanese. Intermediate-level learners do not have the experience of living abroad for more than one month, with an exception of those who had lived abroad for more than one month but had more exposure to Japanese than English, and did not reach 850 in the TOEIC® examination. Participants who do not have experience of living abroad received typical English education at a Japanese junior high school from age 12. None of the participants reported any hearing problems.

2.1.2. Advanced level learners of English

Seven advanced level English learners (5 female, 2 male) participated in the experiment. All are native listeners of Japanese. The criteria for being labeled as advanced learners in English proficiency are whether they 1) achieved higher than 850 in the TOEIC® examination or achieved equivalent scores in the TOEFL® [8] examination, and/or 2) were placed in advanced English classes at university. None of the participants reported any hearing problems.

2.1.3. English native listeners (control group)

Two English native listeners (2 male) participated in the experiment as control listeners. All are native listeners of American English. None of the participants reported any hearing problems.

Table 1. Data of participants.

	Intermediate learners	Advanced learners	English native listeners
Participants	N=10	N=7	N=2
Age	19 - 30 (Mean=23.1)	19 - 27 (Mean=23.1)	27, 41
Length abroad (year;month)	3 wks - 6;0 (Mean=2;10) N=7	0;9 - 9;10 (Mean=3;9) N=6	

2.2. Material

Twenty-three consonants /b, f, d, f, g, h, dʒ, ʒ, k, l, m, n, p, ɹ, s, ʃ, t, θ, ð, v, w, j, z/ were presented to the listeners. Among

the 23 consonants, voiceless fricatives /f, s, ʃ, θ, h/ were selected for analyses, which are some of the phonemes Japanese native listeners struggle to perceive.

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 consonants were embedded in /a __ a/ context, and were presented in the order of 1) quiet, 2) multi-speaker babble noise (SNR=0 dB), and 3) white noise (SNR=0 dB). The two noises were taken from NOISEX [9]. The stimuli were preceded and followed by 1 second of noise, and were presented to the listeners at SNR=0 dB. All experimental procedures were conducted by using Praat [10].

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 (6 in quiet, 9 in multi-speaker babble noise, and 8 in white noise). The practice trials were not scored. After the practice trials, participants proceeded to the main experiment where 138 trials were presented (23 consonants x 2 repetitions x 3 listening conditions). They were asked to listen to each stimulus and to choose the best consonant that fits to what they heard from the 23 consonants, as shown in Figure 1 (words extracted from Cutler *et al.* [2]).

Please choose the consonant that is most similar to what you heard.			
B as in Be	J as in Joke	P as in Pie	TH as in The
CH as in CHin	J as in beiGE	R as in Row	V as in Very
D as in Do	K as in Car	S as in See	W as in Win
F as in Far	L as in Lie	SH as in She	Y as in Yell
G as in Go	M as in My	T as in Tie	Z as in Zoo
H as in Hi	N as in No	TH as in THin	

Figure 1: Twenty-three consonant choices for Experiment 1.

2.4. Results

2.4.1. General results of Japanese native listeners

The average correct rates in the three listening conditions by Japanese native listeners with advanced and intermediate English levels are shown in Figure 2. Advanced learners scored higher than intermediate level learners in all listening conditions, although no significant difference between the two listener groups was found in any of the conditions. However, significant differences were found among the average correct

rates between quiet and babble noise, quiet and white noise, and babble and white noise (Chi-square test, $p < 0.01$).

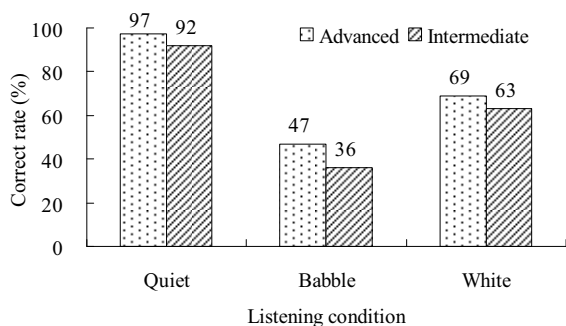


Figure 2: Average correct rates by advanced and intermediate level learners (%).

The approximate percentages of correct responses under quiet listening condition were 97% for advanced learners, and 92% for intermediate level learners. Chi-square test found a significant trend between the two groups of listeners' correct rates ($p = 0.1$). Consonant confusion matrices in quiet by advanced and intermediate level learners are shown in Tables 2 and 3, respectively. Percentages may not add up to 100% due to rounding. Rows represent the stimuli presented to the listeners, and columns represent the listeners' responses.

Table 2. Consonant confusion matrix of advanced learners in quiet (%).

	f	θ	h	s	ʃ
f	100				
θ		86		14	
h			100		
s				100	
ʃ					100

Table 3. Consonant confusion matrix of intermediate level learners in quiet (%).

	f	θ	h	s	ʃ
f	100				
θ	25	70		5	
h			100		
s		10		90	
ʃ					100

The approximate percentages of correct responses under multi-speaker babble noise at SNR=0 dB were 47% for advanced learners, and 36% for intermediate level learners. Chi-square test found a significant trend between the two groups of listeners' correct rates ($p = 0.11$). Consonant confusion matrices in multi-speaker babble noise at SNR=0 dB by advanced and intermediate level learners are shown in Tables 4 and 5, respectively. The phoneme /p/ is added to the response column since some of the target consonants were frequently confused as /p/.

Table 4. Consonant confusion matrix of advanced learners in multi-speaker babble noise at SNR=0 dB (%).

	f	θ	h	s	ʃ	p	others
f	21					71	b=7
θ	50	43				7	
h	43		0			57	
s				100			
ʃ					71		ʒ=7, ʧ=21

Table 5. Consonant confusion matrix of intermediate level learners in multi-speaker babble noise at SNR=0 dB (%).

	f	θ	h	s	ʃ	p	others
f	30					60	b=5, v=5
θ	55	20	5	5		10	h=5
h	40		0			60	
s		15		85			
ʃ					45		ʧ=45, dʒ=5, ʒ=5

The percentages of correct responses under white noise at SNR=0 dB were 69% for advanced learners, and 63% for intermediate level learners. Chi-square test found no significant difference between the two groups of listeners' correct rates ($p = 0.37$). Consonant confusion matrices in white noise at SNR=0 dB by advanced and intermediate level learners are shown in Tables 6 and 7, respectively.

Table 6. Consonant confusion matrix of advanced learners in white noise at SNR=0 dB (%).

	f	θ	h	s	ʃ	p	others
f	79	14					δ=7
θ	21	57		14			δ=7
h			86				g=14
s		43		36	7		k=14
ʃ					86		ʧ=14

Table 7. Consonant confusion matrix of intermediate level learners in white noise at SNR=0 dB (%).

	f	θ	h	s	ʃ	p	others
f	60	15		15			k=10
θ	15	45		30	5		t=5
h	15		85				
s		50		25	10		g=5, s=5, k=5
ʃ					100		

2.4.2. Results of English native listeners

The two English native listeners' numbers of errors are shown in Figure 3. The tendency of consonant confusions was similar to those of Japanese native listeners. English native listeners confused /θ/ as /ð/, f, t/, and /f/ as /θ/, s, p/. Some common errors among Japanese and English native listeners were confusion of /θ/ as /f/, /f/ as /θ/, and /f/ as /p/.

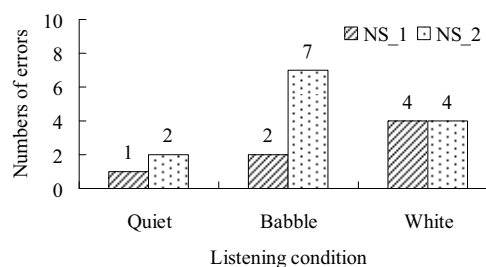


Figure 3: English native listeners' numbers of errors.

3. General Discussion

Overall, the results of the intermediate level learners were always the lowest among the three listener groups in each listening condition. The results of the experiment showed that 1) a significant trend in correct rates was found between advanced and intermediate level learners in quiet and babble conditions, but not in white noise, 2) both advanced and intermediate level learners' performance deteriorated with the presence of noise, and 3) some of the errors that Japanese native listeners made were among those made by English native listeners.

According to Lambacher *et al.* [3], Japanese native listeners were likely to confuse /h/ as /f/, /θ/ as /s/, and /s/ as /θ/ even under quiet listening condition, whereas English native listeners made few confusions, except for the confusion of /f/ as /θ/ (error rate of approximately 11%). The Japanese native listeners in the present experiment, both advanced and intermediate level listeners, performed higher than the Japanese listeners in Lambacher *et al.*, under quiet condition; however, our results under noise resembled their confusion tendencies. The Japanese native listeners in the present experiment were likely to confuse /h/ as /f/ under multi-speaker babble noise, and /θ/ as /s/, and /s/ as /θ/ under white noise, which correspond well with the confusion results of Lambacher *et al.* On the other hand, English native listeners in the present study, though only two participants, did not confuse /θ/ as /s/, and /s/ as /θ/, even under noise. Such results illustrate the influence of L1 (Japanese) on the perception of foreign sounds; Japanese listeners were likely to confuse /θ/ and /s/, for example, because /θ/ does not exist in the Japanese phonemic inventory, and were probably assimilated to /s/. The interesting finding is that not only /θ/ was perceived as /s/, but /s/ was perceived as /θ/ as well. Perceiving phonemes that exist in one's native inventory as a foreign sound is a result of hypercorrection, and may be one of the evidence that proves the confusion of the two phonemes. Other phoneme confusions, such as /θ/ as /f/, /f/ as /θ/, and /f/ as /p/ were observed in both English and Japanese native listeners, which suggest that these confusions may be universal, and are not specified to Japanese native listeners.

Although English native listeners in Lambacher *et al.* [3] made few errors in quiet, English native listeners in Cutler *et al.* [2] and Miller & Nicely [11] made errors under noisy listening conditions that are similar to the results of Japanese native listeners in the present experiment. For example, English native listeners in Cutler *et al.* confused /f/ with /p/ at rate of 22.9%, /s/ with /θ/ at a rate of 10%, /ʃ/ with /tʃ/ at rate of 19.6%, /h/ with /p/ at rate of 14.6%, and /θ/ with /f/ at rate of 13.3% under babble noise of SNR=0 dB. Such confusions were some of those made by advanced and intermediate level learners of the present study, which indicate that even English native listeners' performance is deteriorated in the presence of noise at SNR=0 dB. Furthermore, Miller & Nicely's experiments on the perception of English consonants under noise at SNR=0 dB, +12 dB (in frequency response of 200-6500 Hz) also showed that English native listeners confused /s/ as /θ/, and /θ/ as /f/, under noise at SNR=0 dB by approximately 15% and 34%, respectively. Like the Japanese participants in the present study, /θ/ was perceived as /f/ by approximately 28% even under SNR=+12 dB, which is relatively quiet listening condition. The results of the two previous studies suggest that even English native listeners are prone to confuse phonemes under noisy conditions. The overall correct rates of English native listeners and Japanese advanced learners of English are expected to rise under a less demanding listening condition; that is, SNR=0 dB may have been too severe of a condition for even English native listeners. Further, much closer investigation regarding listening conditions with various SNRs is needed to clarify the difference among Japanese native listeners with advanced and intermediate level English proficiency, and how they differ from English native listeners.

4. Conclusion

The present study investigated the perception of English consonants under quiet and noisy listening conditions at SNR=0 dB by Japanese native listeners with advanced and intermediate level English proficiency, and the results were compared to that of English native control group. Results showed that both advanced and intermediate level learners performed similarly, although intermediate level learners always performed the lowest among the three listener groups. The results of the Japanese native listeners were compared to English native listeners in both the present study and those in previous studies [2, 11], and showed that some of the phoneme confusions of the Japanese native listeners were similar to that of English native listeners.

Although the results attained by the present experiment did not show major differences between Japanese and English native listeners, the gap between the two may widen if the SNR is higher. Taking the results of the present experiment into consideration, perceptual experiment under noisy conditions with a wider range of SNR is needed to clarify the difference among advanced learners, intermediate level learners, and English native listeners.

5. Acknowledgments

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

6. References

- [1] Garcia Lecumberri, M.L. and Cooke, M., "Effect of masker type on native and non-native consonant perception in noise", *J. Acoust. Soc. Am.*, 119 (4): 2445-2454, 2006.
- [2] Cutler, A., Weber, A., Smits, R. and Cooper, N., "Patterns of English phoneme confusions by native and non-native listeners", *J. Acoust. Soc. Am.*, 116 (6): 3668-3678, 2004.
- [3] Lambacher, S., Martens, W., Nelson, B. and Berman, J., "Identification of English voiceless fricatives by Japanese listeners: The influence of vowel context on sensitivity and response bias", *Acoust. Sci. & Tech.*, 22 (5): 334-342, 2001.
- [4] Rogers, C.L., Lister, J.J., Febo, D.M., Besing, J.M., and Abrams, H.B., "Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing", *Applied Psycholinguistics*, 27: 465-485, 2006.
- [5] Mayo, L.H., Florentine, M. and Buus, S., "Age of second-language acquisition and perception of speech in noise", *JSLHR*, 40: 686-693, 1997.
- [6] Grosjean, F., "Studying bilinguals: Methodological and conceptual issues", *Bilingualism: Language and Cognition*, 1: 131-149, 1998.
- [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, A. and Steeneken, H.J.M., "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 Communication*, 12 (3): 247-251, 1993.
- [10] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer [Computer program]". Retrieved from <http://www.praat.org/>
- [11] Miller, G.A. and Nicely, P.E., "An analysis of perceptual confusions among some English consonants", *J. Acoust. Soc. Am.*, 27 (2): 338-352, 1955.