Preliminary study of the perception of English consonants by Japanese learners of English*

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

1 Introduction

Listening to speech in a noisy environment is troublesome for everyone, especially for non-native listeners [1-4]. The difficulty of speech perception for non-native listeners is due to several reasons such as interference from their native language, the length and quality of exposure to the target language, age of acquisition, differences in the usage of phonetic cues, and many more.

Cutler *et al.* [3] performed English vowel and consonant identification experiments in quiet and babble noise on native and non-native (Dutch) listeners. Their study found that while non-native listeners' performance was significantly lower than that of the native listeners in both quiet and babble noise, the disadvantage of the non-native listeners were not affected by the signal-to-noise ratios. However, they found both positive and negative L1 interference for the non-native listeners; 1) non-native listeners performed similarly to native listeners when phonemes were presented at similar syllable positions as in their native language (Dutch), and 2) non-native listeners performed significantly worse than native listeners for phonemes that do not exist in their native language.

Lecumberri and Cooke [4] reported that non-native listeners (Spanish) were less able to identify English consonants in quiet compared to native listeners, and the difference between the two listener groups became larger when consonants were presented in babble, speech-shaped, and competing English and Spanish speech noise. They also found that babble noise was most difficult for both native and non-native listeners. Influence of L1 on the non-native listeners were seen among the patterns of consonant confusion. They claim that this may be due to the differences in the phonetic cues used for perceiving phonemes that are not in their native language, and that non-native listeners' use of cues are not fully developed.

In the case of Japanese native listeners' perception of English consonants, Ueda *et al.* [5] showed that Japanese native listeners' identification of English liquid sounds /r/ and /l/ in noise did not reach native-like performance even after 15 days of training in quiet. Japanese native listeners are well-known for confusing English /r/ and /l/ liquid phonemes due to phonemic differences in Japanese and English.

A study by Mayo *et al.* [6] on the perception of monosyllabic words with high and low predictability by English-Spanish bilinguals showed that early exposure to a second language is advantageous in perceiving sounds in noise. They also claim that duration of exposure is not as influential as the age of exposure to a second language. However, they also state that even bilinguals who had been exposed to English since infancy did not perform as high as native listeners. Rogers *et al.* [7] compared American English native listeners with Spanish-English bilinguals who were exposed to English before age 6. They found a significant difference between native and bilingual non-native listeners in perceiving sounds under noise and reverberation, but not in quiet. The two studies on bilinguals indicate that non-native listeners, especially bilinguals, perform similar to native listeners in quiet condition, but error rates increase when sounds are presented in noise.

A number of studies on non-native speech perception in noise has been conducted, however, research on bilinguals' speech perception is still scarce, and many aspects of bilingualism are yet

* Preliminary study of the perception of English consonants by Japanese learners of English, by MASUDA, Hinako and ARAI, Takayuki (Sophia University).

unknown. The purpose of the present paper is to find out the difference in the perceptual ability of English consonants by Japanese native listeners with advanced and intermediate-level English learners in order to investigate how English learning background affects the perception of L2 consonants.

2 Experiment

2.1 Participants

Nineteen Japanese native listeners participated in the experiment. The data of the participants are shown in Table 1. Eleven were advanced learners of English, who 1) had achieved higher than 850 in TOEIC examination or achieved equivalent scores in TOEFL examination, and/or 2) were placed in advanced English classes at university.

The remaining 8 were intermediate-level learners of English who had not lived abroad for more than one month (with an exception of those who had lived abroad but had more exposure to Japanese than English, and had not achieved 850 in TOEIC examination). Participants who do not have experience of living abroad received English education from age 12 at junior high schools in Japan. None of the participants reported any hearing problems.

Table 1 Data of participants

	Advanced learners	Intermediate learners
Participants	N=11	N=8
Age	19－29 (Mean=22)	19－24 (Mean=20)
Length abroad (year;month)	2 wks－9;10 (Mean=3;10) N=9	3 wks－6;0 (Mean=2;3) N=6

2.2 Material

Twenty-three English consonants /b, ʧ, d, f, ɡ, h, ʤ, ʒ, k, l, m, n, p, ɹ, s, ʃ, t, θ, ð, v, w, j, z/ were presented to participants, and 8 consonants /b, f, h, l, s, ʃ, ɹ, v/ were selected for analysis, which are among some of the phonemes Japanese native listeners struggle to perceive due to phonemic differences in Japanese and English. All consonants were embedded in /ɑ___ɑ/ context. The vowel /ɑ/ was used in order to reduce coarticulatory differences among the stimuli [4].

The speaker of the stimuli is a female, Japanese-English bilingual speaker. The recording of the materials took place 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 three listening conditions were 1) quiet, 2) multi-speaker babble noise, and 3) white noise, presented in that order. Multi-speaker babble noise and white noise were taken from NOISEX [8]. Noise was added to the sound by using MATLAB [9]. The stimuli embedded in noise were presented at an SNR of 0 dB. 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 participants’ response. All procedure of the experiment was conducted by using Praat [10]. 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.* [3]).

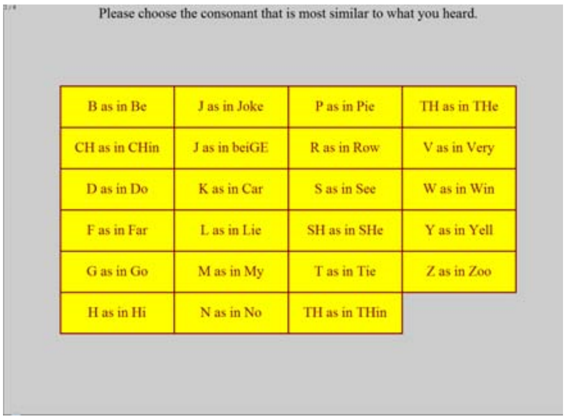

Figure 1 Twenty-three consonant choices for Experiment 1

2.4 Results

2.4.1. Quiet

Consonant confusion matrices in quiet for advanced and intermediate-level learners are calculated into percentages, and are shown in Tables 2 and 3, respectively. Percentages may not add up to 100 due to rounding. Rows represent the stimuli presented to participants, and columns represent participants' responses. /th/ includes both /θ/ and /ð/.

The approximate percentages of correct responses were 98% for advanced learners and 86% for intermediate-level learners. Advanced learners made a total of 4 errors (approximately 2%) out of 176 stimuli (11 participants x 8 consonants x 2 repetitions), while intermediate-level learners made a total of 18 errors (approximately 14%) out of 128 stimuli (8 participants x 8 consonants x 2 repetitions). Chi-square test found a significant difference between the two groups of listeners' number of errors in quiet condition (p<0.01).

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

	b	f	h	l	s	ʃ	ɹ	v	others
b	95								d (4)
f		100							
h			100						
l				100					
s					100				
ʃ						100			
ɹ							100		
v	13							86	

Table 3 Consonant confusion matrix in quiet for intermediate-level learners (%)

	b	f	h	l	s	ʃ	ɹ	v	others
b	100								
f		100							
h			100						
l				56			37		th (6)
s					87	6			th (6)
ʃ						100			
ɹ				31			68		
v	25							75	

2.4.2. Multi-speaker babble noise at 0 dB

Consonant confusion matrices in babble noise for advanced and intermediate-level learners are shown in Tables 4 and 5, respectively. The approximate percentages of correct responses were 49% for advanced learners and 39% for intermediate-level learners. Advanced learners made a total of 89 errors (approximately 51%) out of 176 stimuli, and intermediate-level learners made a total of 78 errors (approximately 61%) out of 128 stimuli. Chi-square test found a significant trend between the two groups of listeners' number of errors in multi-speaker babble noise (p=0.07).

Table 4 Consonant confusion matrix in babble noise for advanced learners (%)

	b	f	h	l	s	ʃ	ɹ	v	others
b	72							27	th (27), g (4), z (4), w (4)
f		40							p (59)
h		36	4						p (59)
l	4			22			9	4	m (36), th (4), n (4), w (13)
s					95				th (4)
ʃ						72			ʧ (18), Z (9)
ɹ				4			95		
v	50							31	th (9), p (9)

Table 5 Consonant confusion matrix in babble noise for intermediate-level learners (%)

	b	f	h	l	s	ʃ	ɹ	v	others
b	56								th (31), d (12)
f	12	31							p (43), th (12)
h		43							p (56)
l				6			12		n (25), m (37), w (18)
s					81	6			th (12)
ʃ						50			ʧ (25), Z (12), ʤ (12)
ɹ				37			62		
v	62							37	

2.4.3. White noise at 0 dB

Consonant confusion matrices in white noise for advanced and intermediate-level learners are shown in Tables 6 and 7, respectively. The percentages of correct responses were approximately 70% for advanced learners and 62.5% for intermediate-level learners. Advanced learners made a total of 53 errors (approximately 30%) out of 176 stimuli, and intermediate-level learners made a total of 48 errors (37.5%) out of 128 stimuli. Chi-square test found no significant difference between the two groups of listeners' number of errors in white noise (p=0.17).

Table 6 Consonant confusion matrix in white noise for advanced learners (%)

	b	f	h	l	s	ʃ	ɹ	v	others
b	81							18	
f		63			4				th (31)
h			95						g (4)
l				59			13		w (27)
s					22	13			k (13), th (36), t (9), g (4)
ʃ						95			ʧ (4)
ɹ				31			68		
v	27							72	

Table 7 Consonant confusion matrix in white noise for intermediate-level learners (%)

	b	f	h	l	s	ʃ	ɹ	v	others
b	81							18	
f		50							th (37), k (12)
h		12	87						
l				50			31		w (18)
s		6			12	6			ʧ (12), th (56), g (6)
ʃ						100			
ɹ				43			50		th (6)
v	31							68	

3 General Discussion

The results of the experiment shows that 1) both listener groups performed the best in quiet condition, 2) multi-speaker babble noise was more difficult than white noise for both listener groups, and 3) similar tendencies in the types of errors could be observed in the two listener groups.

Advanced learners achieved significantly higher scores than intermediate-level learners in quiet condition; however, no significant differences were seen when target sounds were presented in noise. Such results suggest that even advanced learners have difficulties in perceiving L2 sounds in noise. Similar L1 interference could be observed in the two listener groups in phoneme confusion in noise: /b/ was confused with /th/, /f/ as /p/, /h/ as /p/, /l/ as /m/, /ʃ/ as / ʧ/, and /v/ as /b/ in babble noise, and /f/ as /th/, /s/ as / ʃ/, /ɹ/ as /l/, and /v/ as /b/ in white noise.

4 Conclusion

The present study investigated the identification of English consonants by Japanese native speakers with advanced and intermediate-level learners of English. The results of consonant identification revealed that even advanced learners who have had experience of living abroad or achieved high scores on TOEIC examination perform similar to intermediate-level learners, at least in the perception of consonants in noise. However, further research is necessary to investigate whether returnees who have had experience of receiving education in English and advanced learners who have never lived abroad differ, and to compare them with the performance of the native listeners.

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.

References

[1] Nabelek & Donahue, JASA, 75(2), 632-634, 1984.
[2] Takata & Nebelek, JASA, 88(2), 663-666, 1990.
[3] Cutler *et al.*, JASA, 116(6), 3668-3678, 2004.
[4] Garcia Lecumberri & Cooke, JASA, 119(4), 2445-2454, 2006.
[5] Ueda *et al.*, Acoust. Sci. & Tech., 28(4), 251-259, 2007.
[6] Mayo *et al.*, JSLHR, 40, 686-693, 1997.
[7] Rogers *et al.*, Applied Psycholinguistics, 27, 465-485, 2006.
[8] Varga & Steeneken, Speech Commun., 12(3), 247-251, 1993.
[9] MATLAB, http://www.mathworks.com/
[10] Boersma & Weenink, Praat: doing phonetics by computer (Version 5.0.09) [Computer program].