Perception of voiced English consonants in quiet and multi-speaker babble noise by Japanese and English native listeners *

◎Hinako Masuda and Takayuki Arai
(Faculty of Science and Technology, Sophia University)

1 Introduction

Theoretical models in second language learning, such as Perceptual Assimilation Model (Best, 1995) [1] and Speech Learning Model (Flege, 1995) [2], claim that the difficulty of foreign speech learning lies in how similar or dissimilar a given sound is from the sounds in the listener's native language. These models imply that foreign sound perception is determined not by the existence of a particular phoneme in the listener's native language inventory, but whether the target sound is similar or dissimilar from the sounds in the listeners' native inventory, although discussion continues on the definitions of the terms 'similarity' and 'dissimilarity'.

One may be able to perceive speech sounds in a quiet listening condition; however, speech perception is negatively affected in the presence of background noise, even for native listeners. Needless to say, the difficulty becomes larger for non-native listeners [3-5]. The difference between native listeners and non-native listeners with beginner to intermediate level proficiency is often observed in both quiet and noisy listening conditions; however the performance of non-native listeners with high proficiency is more complex.

Although the measurement of proficiency varies among studies, previous research [6-10] have demonstrated that non-natives with high fluency may experience difficulty in perceiving foreign speech sounds in background noise and reverberation, even if they may be able to perform close to native listeners in quiet. The previously conducted research by the authors (which data will also be used in the present report) [9, 10] analyzed the perception of English consonants /f, h, s, ʃ, θ, ɹ, l/ in quiet and background noise (SNR = 10, 5, 0 dB) by English native listeners and Japanese native listeners with intermediate and advanced level English proficiency. The analyses on the English voiceless fricatives showed that while all three listener groups struggled to perceive the sounds in background noise, the confusions of /f/ as /θ/, /ʃ/ as /tʃ/ and /s/ as /θ/ were observed in both intermediate and advanced learners but not in native listeners. This indicates that native language influence was observed in the non-native listeners regardless of their English proficiency. Analyses on the English /ɹ/ and /l/, on the other hand, revealed that advanced learners performed native-like for /ɹ/, but showed significant differences with the native listeners in both quiet and noise for /l/. Such results support the study by Aoyama *et al*. [11] in which they claim that the acquisition of /ɹ/ takes less effort compared to /l/ because of its dissimilarity with the Japanese /ɾ/.

In sum, second language perception in noisy and reverberant listening conditions is difficult for non-native listeners, and the performance of advanced level learners is complex, i.e. accuracy of perception in terms of correct rates and confusion patters varies among phonemes. Previous studies by the authors focused on the analyses of /f, h, s, ʃ, θ, ɹ, l/ [9] and /ɹ/ and /l/ [10]; this paper will focus on the perception of /b, v, ð, z, dʒ, ʒ/ in quiet and multi-speaker babble noise (SNR = 10, 5, 0 dB) by English native listeners and Japanese native listeners with intermediate and advanced level English proficiency, using the data collected in previous studies [9, 10]. The present study will examine how second language proficiency affects perception of voiced English consonants in both quiet and noisy listening conditions with varying signal-to-noise ratios, and we will discuss the differences in the performances of the listeners by creating confusion matrices.

* Perception of voiced English consonants in quiet and multi-speaker babble noise by Japanese and English native listeners, by MASUDA, Hinako and ARAI, Takayuki (Sophia University).

2 Perceptual experiment

2.1 Participants

Thirty listeners participated in the present experiment, and they are divided into three groups according to their language learning background (see Table 1). Three participants were excluded from analyses due to insufficient participant information. Twenty-one Japanese native listeners participated as the non-native listener group (15 male, 9 female), and six English native listeners (2 male, 4 female) participated as the control group. None of the participants reported any hearing problems.

Table 1 Data of participants [9, 10]

	Intermediate learners	Advanced learners	English natives
Number of participants	N = 8	N = 13	N = 6
Age (Range)	Mean 23.0 (20-31)	Mean 26.4 (20-35)	Mean 20.8 (20-21)

Japanese native listeners with higher proficiency were grouped as advanced level learners of English, who had achieved higher than 850 on TOEIC® examination (Test of English for International Communication, provided by Educational Testing Service) or achieved equivalent scores on TOEFL® examination (Test of English as a Foreign Language, also provided by Educational Testing Service), and/or were placed in advanced level English classes at a university in Japan. The remaining participants are grouped as 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 at least six years of English education from age twelve at junior high schools and high schools in Japan.

Six English native listeners participated in the experiment as control listeners. All six participants were recruited in Japan. Three participants had Japanese, Chinese, or Korean background, and moved to the United States in the first few years of life and were exposed to American English thereafter. All three participants considered American English as their native language, and Japanese, Chinese, or Korean as a foreign language. The level of Japanese was in the range of beginner to intermediate in all participants. All English native listeners had less than one year of residence in Japan, except for one participant with the Japanese background who had lived in Japan for two years, from birth to age two.

2.2 Stimuli

Twenty-three English 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 presented to the listeners. Among the 23 consonants, data of six voiced consonants /b, v, ð, z, dʒ, ʒ/ are analyzed and reported in the present study. 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.

Participants first listened to the stimuli with background noise (multi-speaker babble noise at SNR = 0 dB, 5 dB, 10 dB), and then proceeded to the perception of the stimuli in quiet. The stimuli embedded in noise were presented in random order. The stimuli in noise were preceded and followed by 1 second of noise. The added multi-speaker babble noise was taken from NOISEX [12]. Multi-speaker babble noise was selected as background noise since it most likely resembles real-life environment that second language learners may experience difficulties in foreign language perception, such as in a school cafeteria.

2.3 Procedure

A laptop computer was used to present the stimuli and to record the listeners' responses. All experimental procedures were conducted by using Praat [13]. 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

(six each in SNR = 0 dB, 5 dB, 10 dB, and five in quiet). The practice trials were not scored nor given feedback. 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). 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, as shown in Figure 1.

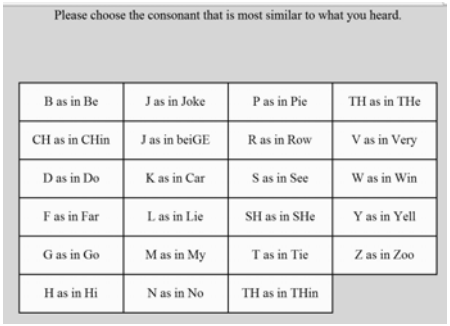

Figure 1 Experimental interface
(Reference: Cutler *et al.*, 2004)

3 Results

The results of /b, v, ð, z, dʒ, ʒ/ were analyzed, and the average correct rates are shown in Figure 2. Two-way analysis of variance found significant effects of both listener groups (p = 0.01) and listening conditions (p < 0.01) in the average correct rates. Correct rates decreased as listening conditions degraded for both native and non-native listeners. Intermediate learners' performance degraded dramatically with the presence of noise; however, advanced learners maintained a native-like performance throughout all of the conditions.

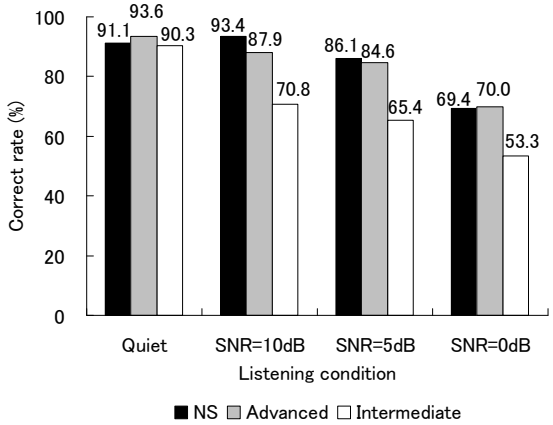

Figure 2 Correct rates of the perception of /b, v, ð, z, dʒ, ʒ/ by the three listener groups

Further analyses on each consonant revealed that the advanced learners' tendency in performing native-like or intermediate learners-like varies among consonants and listening condition. The perception of /b/ by advanced level learners, for example, resembled the results of intermediate level learners in the condition of SNR = 0 dB (Figure 3). The perception of /z/ by advanced learners, however, did not resemble intermediate learners in any of the conditions; rather, they were able to maintain native-like performance in all of the conditions (Figure 4).

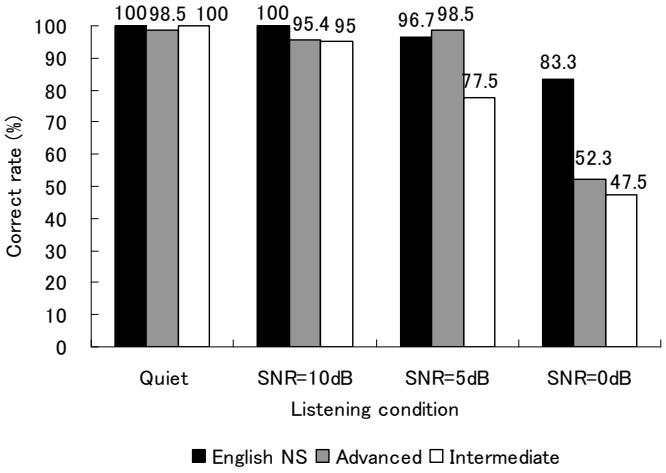

Figure 3 Correct rates of the perception of /b/ by the three listener groups

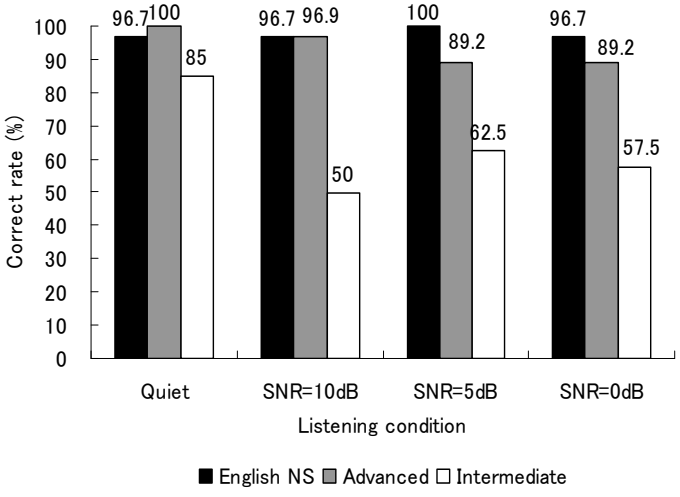

Figure 4 Correct rates of the perception of /z/ by the three listener groups

4 Discussion

Consonant confusion matrices of the six consonants by the three listener groups in SNR = 0 dB are shown in Tables 2 – 4. The correct rates of /b/ in SNR = 0 dB condition were similar in advanced and intermediate level learners, as mentioned previously, at approximately 50%. It is interesting to note that the similarity in the two listener groups was not only observed in the correct rates but also in the consonant confusion pattern. Advanced level learners confused /b/ as /ð/, a tendency observed also in intermediate level learners but not in English native listeners. The correct rates and confusion of /b/ suggest that advanced level learners are also influenced by their first language, just as intermediate level learners are.

Table 2 Confusion matrix
(English natives in SNR = 0 dB) (%)

	b	v	ð	z	dʒ	ʒ	others
b	83.3	13.3					3.3 (d)
v	60.0	30.0					3.3 (f), 6.7 (p)
ð	6.7	6.7	50.0				3.3 (d), 10.0 (t), 23.3 (θ)
z			3.3	96.7			
dʒ					83.3	10.0	6.7 (tʃ)
ʒ					26.7	73.3	

Table 3 Confusion matrix
(advanced learners in SNR = 0 dB) (%)

	b	v	ð	z	dʒ	ʒ	others
b	52.3	23.1	12.3				7.7 (d), 1.5 (g), 1.5 (ɹ), 1.5 (w)
v	49.2	47.7					1.5 (p), 1.5 (w)
ð	1.5	3.1	80.0	13.8			1.5 (θ)
z			10.8	89.2			
dʒ					72.3	21.5	6.2 (tʃ)
ʒ					21.5	78.5	

Table 4 Confusion matrix
(intermediate learners in SNR = 0 dB) (%)

	b	v	ð	z	dʒ	ʒ	others
b	47.5	15.0	20.0				2.5 (d), 2.5 (g), 2.5 (l), 7.5 (θ), 2.5 (w)
v	52.5	32.5	7.5				2.5 (f), 5.0 (θ)
ð	2.5	5.0	75.0	5.0			12.5 (θ)
z			37.5	57.5			5.0 (θ)
dʒ					55.0	40.0	5.0 (tʃ)
ʒ					47.5	52.5	

5 Conclusion

The present study reported the results of the perception of /b, v, ð, z, dʒ, ʒ/ in quiet and multi-speaker babble noise at SNR = 0, 5, 10 dB by Japanese native listeners with intermediate and advanced level English proficiency, and English native listeners. Average overall correct rates showed that advanced level learners' performance resembled that of the English native listeners, although further analyses showed that such tendency varied among consonants. Furthermore, consonant confusion matrices by the three listener groups in SNR = 0 dB condition showed that the confusion of /b/ as /ð/ was unique to Japanese native listeners, regardless of their English proficiency.

Acknowledgments

This study work was partially supported by Sophia University Open Research Center from MEXT.

References

[1] Best, *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, 171-204, 1995.
[2] Flege, *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, 233-277, 1995.
[3] Takata & Nabelek, *J. Acoust. Soc. Am.*, 88 (2), 663-666, 1990.
[4] Cutler *et al.*, *J. Acoust. Soc. Am.*, 116 (6), 3668-3678, 2004.
[5] Garcia Lecumberri & Cooke, *J. Acoust. Soc. Am.*, 119 (4): 2445-2454, 2006.
[6] Rogers et al., *Applied Psycholinguistics*, 27, 465-485, 2006.
[7] Mayo et al., *JSLHR*, 40, 686-693, 1997.
[8] Florentine, *Proc. Inter-noise 85*, 1021-1024, 1985.
[9] Masuda & Arai, *Proc. ASJ spring meeting*, 471-474, 2011.
[10] Masuda & Arai, *Proc. ASJ spring meeting*, 477-480, 2012.
[11] Aoyama *et al.*, *Journal of Phonetics*, 233-250, 2004.
[12] Varga & Steeneken, *Speech Communication*, 12 (3), 247-251, 1993.
[13] Boersma &Weenink, Praat: doing Phonetics by computer (Version 5.0.09) [Computer program].