Effect of consonant manner on L2 speech perception in multispeaker babble noise

Hinako MASUDA[1]

[1]Seikei University, Japan

ABSTRACT

Past research has shown that perceiving speech in non-laboratory environments is challenging for both native and non-native listeners, even if they are able to perform well in laboratory (quiet) environment. The present study reports the impact of manner of articulation on English consonant identification in multispeaker babble noise by Japanese learners of English. Twenty-five American English listeners and 17 Japanese listeners were presented with 23 American English consonants embedded in the context "You are about to hear a__a". Stimuli were presented with and without multipeakerbabble noise. Participants were asked to listen to the stimuli, and to choose the consonant they heard from 23 consonant choices. Three signal-to-noise ratios were adapted for the noisy environments: 10dB, 5dB, and 0dB. Analysis was carried out as Language (Japanese or English), Manner (stops, fricatives, affricates), and Condition (SNR = 10dB, 5dB, and 0dB) as independent variables and Accuracy rates as the dependent variable. Statistical analyses demonstrated significant interactions between Language and Condition, as well as Condition and Manner.

Keywords: Speech Perception, L2, Consonants, Japanese, English
I-INCE Classification of Subjects Number(s): 63.3

1. INTRODUCTION

Speech perception in non-laboratory listening environments can be difficult for all listeners, both native and non-native, and difficulty varies among different types of sounds. For example, a study by Miller & Nicely (1954) (1) on the identification of English consonants in background noise by native listeners of English demonstrated that voicing and nasality features of consonants were maintained even when they were exposed to noise as severe as signal-to-noise ratio (SNR) of -12 dB, while place of articulation was prone to be negatively affected by noise even at SNR = 6 dB.

A number of research have shown that accurate perception of non-native speech sounds is difficult regardless of the listeners' proficiency in the target language (2-5). The difficulty undoubtedly becomes even greater when such sounds are presented in non-laboratory environments, i.e. in background noise and reverberation. Previous studies on non-native speech perception in adverse listening environments have demonstrated that 1) even non-native listeners that were able to perform near-native in identifying non-native sounds in quiet listening condition had difficulty as listening environments became degraded, for both non-words (6) and real words (7), 2) vowels are generally more accurately perceived than consonants (6), and 3) the kind of background noise used as a masker has an effect on the performance of both native and non-native listeners (8).

The present study investigates the effect of consonant manner on accurate identification of English consonants by native (English) and non-native (Japanese) listeners in both quiet and noisy listening environments. A number of studies have claimed that fricatives, especially voiceless fricatives, are difficult for Japanese listeners to perceive accurately (9, 10) as well as English approximants such as /r/ and /l/ (11-13, among many others). One of the reasons behind such difficulties lies in the differences between the phonological systems of English and Japanese. Concerning consonants, it is well-known that English approximants /r/ and /l/ do not exist in Japanese and is often merged into one Japanese approximant /ɾ/. Fricatives, especially voiceless fricatives, are also a set of consonants that Japanese listeners struggle to identify accurately. In

1 h-masuda@st.seikei.ac.jp

Japanese, there are phonetically five voiceless fricatives [ɸ s ɕ ç h] which are not phonologically distinct (14). Taking such differences into consideration along with the general sonority of consonants (stops < affricates < fricatives < nasals < approximants) and results demonstrated by Miller & Nicely (1954) (1) that the nasal feature is tolerant of noise, this study hypothesizes that Japanese listeners' difficulty of accurately identifying non-native sounds in noise increases in the order of nasals > affricates > stops > approximants > fricatives.

2. IDENTIFICATION EXPERIMENT

2.1 Participants

Twenty-five American English listeners and 17 Japanese listeners participated in the identification experiment. All participants were university students in their 20s and 30s at the time of the experiment, and none of the listeners reported any hearing difficulties. The Japanese listeners were recruited in Japan, and the American English listeners were recruited in both Japan and the United States.

2.2 Stimuli

The stimuli used in the identification experiment consisted of 23 American English consonants [p b t d k g tʃ ʤ m n f v θ ð s z ʃ ʒ h ɹ j w l] embedded in an intervocalic context /aCa/. The nasal consonant /ŋ/ was excluded from the stimuli set because it rarely appears in the intervocalic position in English. The stimuli were presented in the carrier phrase "You are about to hear XXX". The stimuli were produced by a Japanese-English bilingual speaker, and were recorded in a sound-attenuated chamber using a digital sound recorder (Marantz PMD660) and a microphone (SONY ECM-23F5) at a sampling frequency of 48 kHz. The sounds were later downsampled to 16 kHz.

2.3 Procedure

A laptop was used to present the stimuli to the listeners and to record their responses. The identification experiment consisted of three sessions: 1) practice session, 2) stimuli presented in noisy listening conditions, and 3) stimuli presented in quiet listening condition. The stimuli were presented in SNRs (Signal-to-Noise Ratios) of 10dB, 5dB and 0dB for the noisy listening conditions. Multi-speaker babble noise (15) was selected as background noise which is more challenging than stationary noise and best depicts native and non-native difference in noise (Lecumberri & Cooke, 2006).

The stimuli were presented to the listeners through headphones (STAX SR-303 or STAX SRM-323A; English listeners recruited in the United States used Sennheiser HD280 Pro) and a USB amplifier (ONKYO MA-500U). English listeners recruited in the United States were presented to the stimuli directly from a Mac computer. Each stimulus was presented to the listener only once. After listening to the stimulus, listeners were asked to identify the consonant between the vowels, and to select one of the 23 consonants from the experimental interface on the computer screen that was most close to what they heard (e.g. "B as in Be" for /aba/, see Figure 1). The trials automatically proceeded after listeners selected one of the consonants. Inter-stimulus interval was 0.3 seconds, and response time was not measured. The practice session consisted of 23 trials (6 trials * 3 SNR conditions + 5 trials in quiet condition), and did not provide feedback. Participants then proceeded to the noisy listening conditions, which were randomized across three SNR conditions. Lastly, participants proceeded to the quiet listening condition. The main part of the experiment, 2) and 3), consisted of 460 trials in total (23 consonants * 4 listening conditions * 5 repetitions) which took around 40 minutes to complete. All experimental procedures were carried out using the computer program Praat (16).

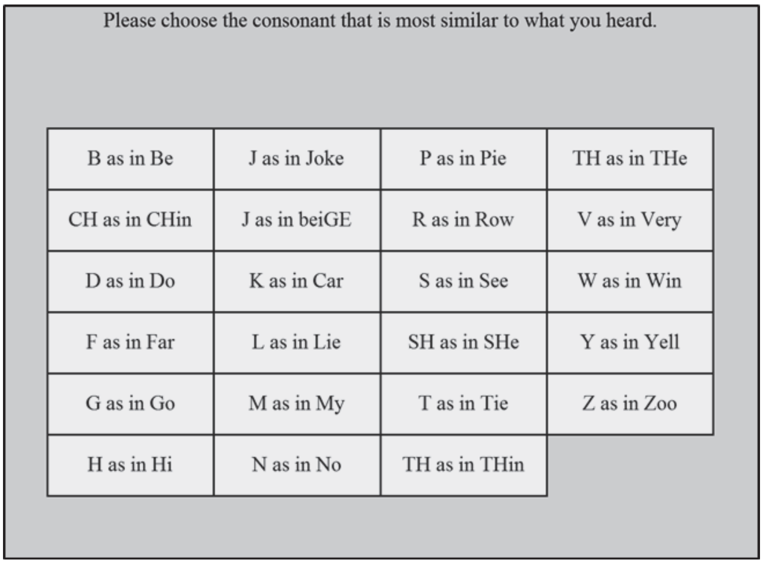

Figure 1 Experimental interface

3. RESULTS & DISCUSSION

Figures 2 and 3 show accurate identification rates of each consonant manner in the four listening conditions by English (Figure 2) and Japanese (Figure 3) listeners. As expected, both speaker groups performed best in the quiet listening condition, and performance decreased as listening conditions became more challenging. As demonstrated in Miller & Nicely (1954)'s study, nasals best tolerated the degradation of listening conditions amongst others in both listener groups, with a small drop in accuracy rate even in the most challenging condition.

Statistical analysis was carried out as Language Group, Manner, and Condition as independent variables and Accuracy rates as the dependent variable. Analysis of Variance observed significant Main Effect of Language (F (1, 40) = 8.94, p = 0.004, η^2= 0.09), Condition (F (3, 40) = 74.51, p = 0.00, η^2= 0.19), and Manner (F (2, 40) = 36.35, p = 0.00, η^2= 0.16), and significant interactions between Language and Condition (3, 40) = 4.26, p = 0.006, η^2= 0.01) and Condition and Manner (F (6, 40) = 2.82, p = 0.01, η^2= 0.01). Unexpectedly, English listeners performed worse than Japanese listeners; therefore, this paper discusses the two languages separately. To look at differences in perceptual difficulty among consonant manners within the language group, *post-hoc* analysis using pairwise comparisons with Holm adjustment for *p*-values was carried out (summarized in Table 1). For the Japanese listeners, there were significant differences in accuracy rates between all but affricates/approximants. For the English listeners, significant differences were observed in all but affricates/fricatives and stops/approximants.

Considering the performance in the quiet condition as baseline, English listeners were able to tolerate noise and performed steadily in SNR = 10dB and 5dB for only nasals, whereas Japanese listeners performed steadily in nasals and stops. The hypothesis that difficulty increases in the order of nasals > affricates > stops > approximants > fricatives was confirmed for Japanese listeners with an exception of reverse order of affricates/stops. Although English listeners' performance of affricates were higher than fricatives in N0 (quiet environment) to N2 (SNR = 5 dB), it resembled the performance of fricatives in N3 (SNR = 0dB). Japanese listeners' performance of affricates, on the other hand, remained high even in the most difficult condition.

Japanese listeners' data was further analyzed considering their experience of living abroad. There was a significant interaction between having the experience of living abroad and accurate perception of fricatives (t (7795) = 4.028, $p < 0.00$) and approximants (t (7795) = 3.28, p = 0.001), with higher performance by those with experience of living abroad. Considering that fricatives and approximants are the two groups of consonants that Japanese listeners struggle to accurately identify as non-native listeners, this result suggests that exposure to English facilitates accurate perception of challenging consonants.

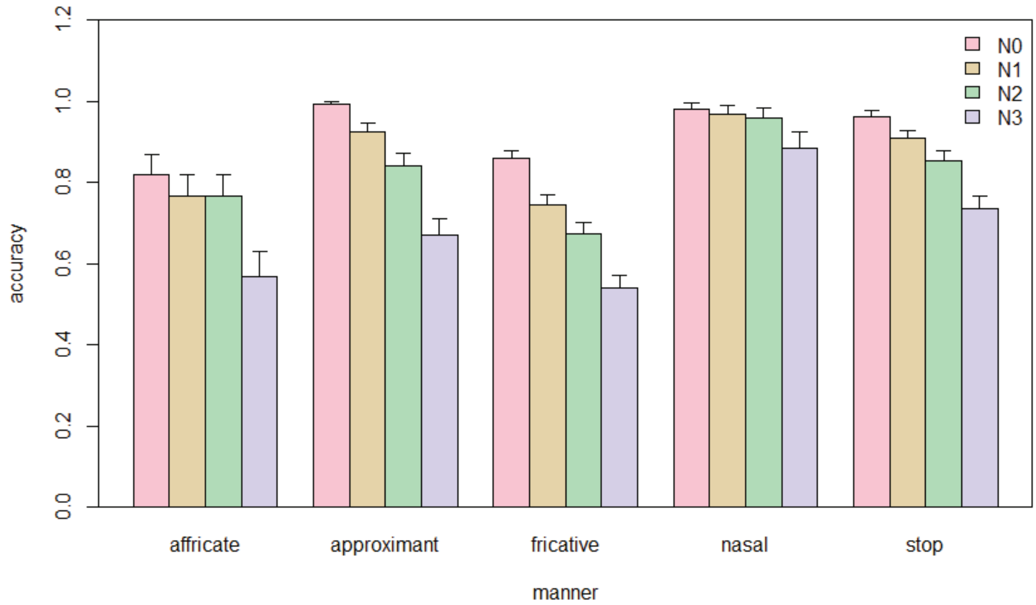

Figure 2 Accurate identification rates of each consonant manner by English listeners
(N0: Quiet condition, N1: SNR = 10dB, N2: SNR = 5dB, N3: SNR = 0dB)

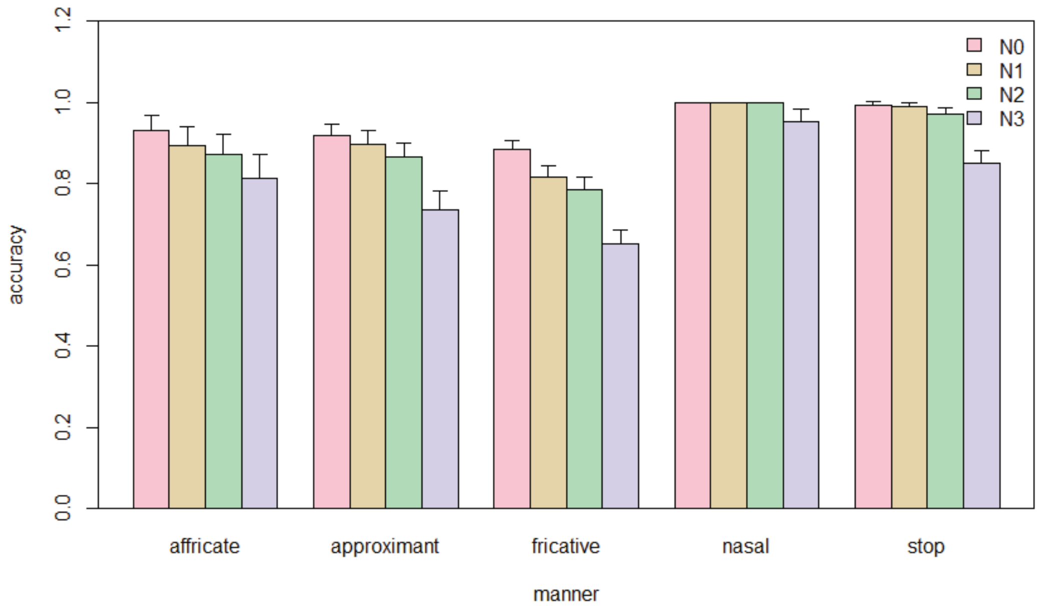

Figure 3 Accurate identification rates of each consonant manner by Japanese listeners
(N0: Quiet condition, N1: SNR = 10dB, N2: SNR = 5dB, N3: SNR = 0dB)

Table 1 Summary of post-hoc analysis using the Holm adjustment for *p*-values
(E: English listeners, J: Japanese listeners)

	Affricate	Approximant	Fricative	Nasal
Approximant	E: 9.9e-16 J: 0.144	-	-	-
Fricative	E: 0.11 J: 3.9e-10	E: < 2e-16 J: 8.0e-10	-	-
Nasal	E: < 2e-16 J: 2.4e-09	E: 7.9e-09 J: < 2e-16	E: < 2e-16 J: < 2e-16	-
Stop	E: < 2e-16 J: 1.1e-06	E: 0.47 J: 3.5e-16	E: < 2e-16 J: < 2e-16	E: 2.0e-08 J: 0.025

4. CONCLUSIONS

This paper investigated the effect of consonant manner on the identification of English consonants in quiet and noisy listening environments by English and Japanese listeners. Results demonstrated that fricatives were most difficult to accurately identify for both listener groups, and that nasals tolerate noise the most. Within Japanese listeners, some advantages were observed for those with experience of living abroad – they identified fricatives and approximants significantly higher than those without experience of living abroad. The two types of consonants have been known as challenging for Japanese learners of English. The results suggest that experience of living abroad is beneficial in achieving higher identification rates in challenging consonants.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant number 2658011 (Grant-in-Aid for Challenging Exploratory Research).

REFERENCES

1. Miller GA., Nicely PE. An analysis of perceptual confusions among some English consonants. J Acoust Soc Am. 1954; 27(2): 338-352.
2. Florentine M. Non-native listeners' perception of American-English in noise. Proc. Inter-noise 85. 1985; 1021-1024.
3. Mayo LH, Florentine M, Buus S. Age of second-language acquisition and perception of speech in noise. J Speech Lang Hear Res. 1997; 40, 686–693.
4. Rogers CL, Lister JJ, Febo DM, Besing JM, Abrahams HB. Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. Appl Psycholing. 2006; 27: 465–485.
5. Masuda H. Misperception patterns of American English consonants by Japanese listeners in reverberant and noisy environments. Speech Commun. 2016; 79: 74-87.
6. Cutler A, Weber A, Smits R, Cooper N. Patterns of English phoneme confusions by native and non-native listeners. J Acoust Soc Am. 2004; 116(6): 3668-3678 .
7. Nabelek A, Donahue AM. Perception of consonants in reverberation by native and non-native listeners. J. Acoust. Soc. Am. 1984; 75(2): 632-634.
8. Garcia Lecumberri ML, Cooke M. Effect of masker type on native and non-native consonant perception in noise. J Acoust Soc Am. 2006; 119(4): 2445-2454.
9. Lambacher S, Martens W, Nelson B, Berman J. Identification of English voiceless fricatives by Japanese listeners: The influence of vowel context on sensitivity and response bias. Acoust Sci Technol. 2001; 22(5): 334-343.
10. Masuda H, Arai T. Identification of English voiceless fricatives in multispeaker babble noise by native Japanese and English listeners: Influence of English proficiency. Acoust Sci Technol. 2013; 34(5):

356-360.
11. Miyawaki K, Strange W, Vergrugge R, Liberman A, Jenkins J, Fujimura O. An effect of linguistic experiment. The discrimination of [r] and [l] by native speakers of Japanese and English. Percept Psychophys. 1975; 18(5): 331-340.
12. Logan J, Lively S, Pisoni D. Training Japanese listeners to identify English /r/ and /l/: A first report. J Acoust Soc Am. 1991; 89(2): 874-886.
13. Aoyama K, Flege JE, Guion SG, Akahane-Yamada R, Yamada T. Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. J Phon. 2004; 32: 233-250.
14. Vance TJ. The Sounds of Japanese. Cambridge University Press, Cambridge; 2008.
15. Varga A, Steeneken HJM. Assessment for automatic recognition II: NOISEX-92: a database and an experiment to study the effect of additive noise on speech recognition systems. Speech Commun. 1993; 12(3): 247-251.
16. Boersma P, Weenink D. Praat: doing Phonetics by computer [Computer program] Version 5.3.66, retrieved from. http:// www.praat.org/; 2014.