TASK DIFFERENCES IN THE IDENTIFICATION OF ENGLISH CONSONANTS IN NOISE BY JAPANESE LISTENERS

Hinako Masuda

Seikei University, Tokyo, Japan
h-masuda@st.seikei.ac.jp

ABSTRACT

Research has demonstrated the challenges that listeners face when they are presented with non-native sounds in noise. The challenges have been observed in both low- and high-level perception. However, the relationship between the two types of processing remains unknown. The present paper reports the results of two experiments. Experiment 1 explored the identification of English consonants in intervocalic context, and Experiments 2 explored the identification of words, both in quiet and noisy conditions by native Japanese listeners. Consonants that Japanese listeners often have difficulty to perceive accurately (/l/ and /r/) are extracted and discussed. Results showed that 1) task differences did not have a significant effect on overall consonant identification, 2) task differences affected the identification of /l/ and /r/, and 3) predictability played an important role in word identification.

Keywords: speech perception, word, consonant, non-native, noise.

1. INTRODUCTION

Perceiving speech in non-idealistic (e.g. noisy, reverberant or both) environments is difficult, yet we are constantly exposed to them. Although even native listeners suffer accurate perception in severe listening environments, the challenge is obviously larger for non-native listeners. Due to globalization, learning and using foreign languages outside of traditional classrooms have become common. Therefore, investigation of how listeners perceive non-native speech in real-life environments is important, and the demand for development of language learning materials that consider such conditions are increasing.

Numerous studies have been carried out regarding non-native speech perception, targeting from low-level (e.g. segment level perception such as consonants and vowels) to high-level processing (e.g. word and sentence level perception), both in quiet and adverse listening environments. For example, a vast number of experiments on the identification and discrimination of non-native consonants and vowels have been conducted (e.g. [2, 5, 6, 10]). Such research generally suggest that non-native listeners' perception is often influenced by their first language, and that when non-native sounds are presented in adverse listening environments, they are less likely to tolerate the detrimental effect of background disturbances such as noise compared to the native listeners.

Similarly, research has also looked into non-native word perception (e.g. [4, 7, 9, 11, 12]), which also generally suggest, like the research on segments, that non-native listeners are vulnerable to noise. In the case of high-level perception, however, the factors that affect perception are more complex compared to perception of segments. That is, factors such as lexical familiarity, predictability and proficiency in the target language must also be considered, as listeners are more likely to perceive familiar words and predictable words more accurately than words that are unfamiliar and unpredictable [7, 11], and more proficient non-natives are able to perceive words more accurately than less proficient non-natives (e.g. [16]).

Traditionally, experiments reported in most papers have focused on one type of experiment (e.g. either consonant identification task or word identification task). The present paper aims to bridge this gap. Specifically, the present study aims to investigate whether the ability to perceive segments (consonants) is carried over to the perception of words. To examine this, two experiments are compared. Experiment 1 is a consonant identification task where non-native (Japanese) listeners are asked to listen to consonants in an intervocalic context. Experiment 2 is a word identification task where the same non-native listeners are asked to listen to predictable and unpredictable words in sentences. Both experiments were conducted in quiet and noisy listening conditions to investigate the effect of background noise. To look deeper into how listeners' native language affects perception, the results of consonants /l/ and /r/ are extracted and discussed, as they are the two consonants that Japanese listeners often suffer when listening to speech in English (e.g. [1, 13, 15].

2. EXPERIMENTS

The procedures for the data collection was approved by the Seikei University Ethics Committee.

2.1. Experiment 1: Consonant identification task

2.1.1. Participants

Thirty Japanese listeners of English (L1: Japanese, L2: English) participated in Experiment 1. All participants were university students at the time of experiment (age range 18-22) and had no prior experience of studying phonetics. Their TOEIC® scores ranged from 420-700 (average 556, median 545). None of the participants reported any hearing problems.

2.1.2. Stimuli

The stimuli consisted of 17 English consonants (/b d ʤ f h ʒ l ɹ s t ʃ θ ð v w j z/). The stimuli were produced by two speakers: male American English speaker and female Japanese-American English bilingual speaker. The recordings by the American speaker were made using a microphone (Blue Microphones Yeti) and the Japanese-American English bilingual speaker's recordings were made using a digital sound recorder (TASCAM Linear PCM recorder DR-100MKII) and a microphone (Audio Technica AT4022), both at a sampling frequency of 48 kHz, later downsampled to 16 kHz.

The consonants were embedded in an intervocalic context in the carrier sentence "You are about to hear /aCa/ on the tape again" where C represents the target consonant. The stimuli were presented to the listeners in quiet and noisy (multispeaker babble noise at Signal-to-Noise Ratio (SNR) of 0dB) listening conditions. The stimuli were preceded and followed by one second of noise. Each participant listened to a total of 340 stimuli (17 consonants x 2 speakers x 2 listening conditions x 5 repetitions).

2.1.3. Procedure

The experiment took place in a soundproof room. A laptop computer and the software Praat [3] was used to carry out the experiment. An identification task was designed, in which participants were asked to listen to the stimuli with a pair of headphones (SONY MDR CD900ST) and to choose the consonant that they heard from a list of 23 consonants (/p b t d k g t ʃ ʤ m n f v θ ð s z ʃ ʒ h ɹ j w l/ presented in the form of, for example, "B as in Be"). The stimuli were presented only once. Participants were presented to practice trials prior to the main experiment (12 in noise and quiet conditions, respectively). All participants were paid upon completion of the experiment.

2.2. Experiment 2: Word identification task

2.2.1. Participants

Twenty listeners from the subject pool in Experiment 1 participated in Experiment 2.

2.2.2. Stimuli

The stimuli used in Experiment 2 were produced by a male native American English speaker. The stimuli were recorded in a soundproof room using a microphone (Audio Technica AT4022) and a digital recorder (TASCAM Linear PCM recorder DR-100MKII) at a sampling frequency of 48 kHz, later downsampled to 16 kHz.

The stimuli consisted of 140 sentences (70 words x 2 predictability types) which were formed based on the Speech in Noise Test (SPIN). One hundred sentences were presented in noisy condition and 40 sentences were presented in quiet condition. Multispeaker babble noise at a signal-to-noise ratio (SNR) of 0 dB was used to create the noisy listening condition. The stimuli were preceded and followed by one second of noise. The target word was always the last word of the sentence and was always monosyllabic. Half of the sentences in each listening environment (50 in noisy and 20 in quiet condition) were contextually predictable (e.g. "We were swimming in the *pool*."), and the remaining half were contextually unpredictable (e.g. "She was talking about the *pool*."). The target words in noisy and quiet listening conditions were different.

2.1.3. Procedure

The experiment took place in a soundproof room. A laptop computer and the software Praat [3] was used to present the stimuli. Participants were asked to listen to the stimuli with a pair of headphones (SONY MDR CD900ST) and to write down the final word of the sentence on the answer sheet provided by the experimenter. The sentences were also visually presented to the listeners at the top of the experimental interface. The stimuli were presented only once. Prior to the main experiment, the participants were presented with six practice trials (three in noise and three in quiet conditions). All participants were paid upon completion of the experiment.

3. RESULTS

3.1. Experiment 1 (Consonant identification)

Figures 1 shows the merged accuracy rates of all 17 consonants (analyzed items: 30 participants x 17 consonants x 2 conditions x 2 speakers x 5 repetitions = 10200 items). As expected, listeners' accuracy rates were higher when stimuli were presented in quiet condition. Statistical analysis using a two-way analysis of variance with speaker (monolingual or bilingual) and condition (quiet or noisy) as independent variables was carried out. The results did not show a significant interaction between condition and speaker ($F(1,8772)=2.77$, $p=.09$). Significant main effects were observed in condition ($F(1,8772)=60.73$, $p<.001$), but not in speaker ($F(1,8772)=0.51$, $p=.47$).

Figure 1: Accuracy rates of all 17 consonants in quiet/noise spoken by bilingual/monolingual speakers (%).

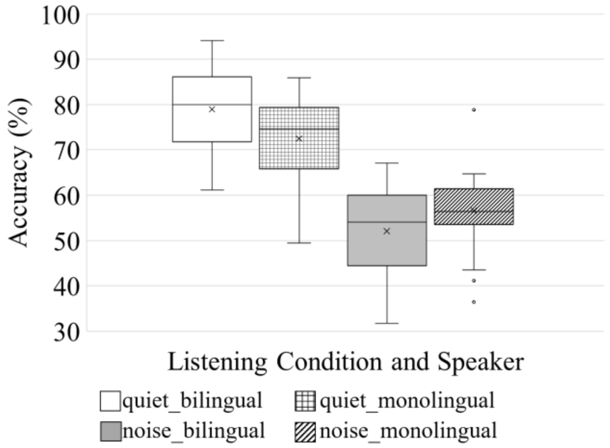
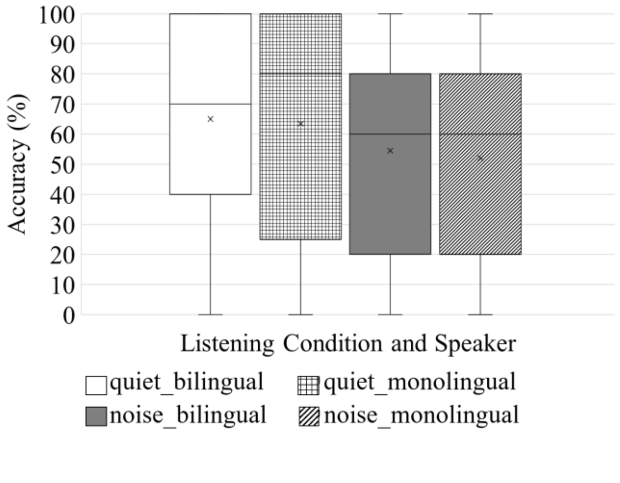


Figure 2 shows the extracted results of /l/ and /r/ (analyzed items: 30 participants x 2 consonants x 2 conditions x 2 speakers x 5 repetitions: 1200 items), the two consonants that are often documented as perceptually challenging to native listeners of Japanese. Again, listeners' accuracy rates were higher when stimuli were presented in quiet condition. A significant main effect was observed in condition ($F(1,1072)=14.93$, $p<.001$), but not in speaker ($F(1,1072)=0.06$, $p=.80$).

Figure 2: Accuracy rates of /r/ and /l/ in quiet/noise spoken by bilingual/monolingual speakers (%).



3.2. Experiment 2 (Word identification)

Figure 3 shows the accuracy rates of all words in quiet (40 sentences)/noisy (100 sentences) conditions in predictable/unpredictable context (analyzed items: 20 participants x 70 words x 2 predictability = 2800 items). Accurate responses were only those with correct spellings (i.e. misspelled words were categorized as wrong answers). Statistical analysis using a two-way analysis of variance with predictability (predictable or unpredictable) and condition (quiet or noisy) as independent variables was carried out. The results showed a significant interaction between predictability and condition ($p<.001$). Significant main effects of predictability ($p<.001$) and condition ($p<.001$) were observed. Post-hoc using Tukey-Kramer test showed significant differences ($p<.01$) in the accuracy rates between quiet_high and quiet_low, quiet_high and noise_high, quiet_high and noise_low, quiet_low and noise_high, quiet_low and noise_low, and noise_high and noise_low, meaning accuracy rates decreased significantly as listening conditions and predictability became less ideal.

Figure 3: Accuracy rates of all words in quiet/noise and predictable/unpredictable conditions (%) (high: high predictability, low: low predictability).

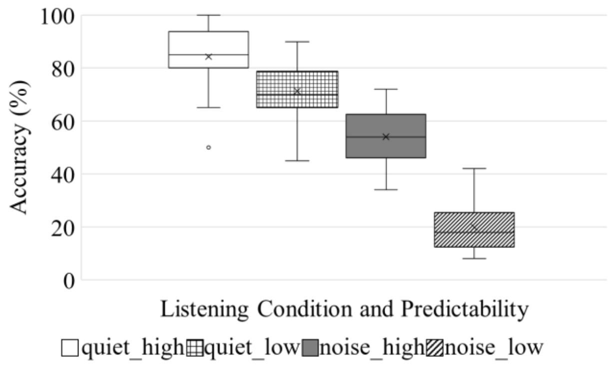
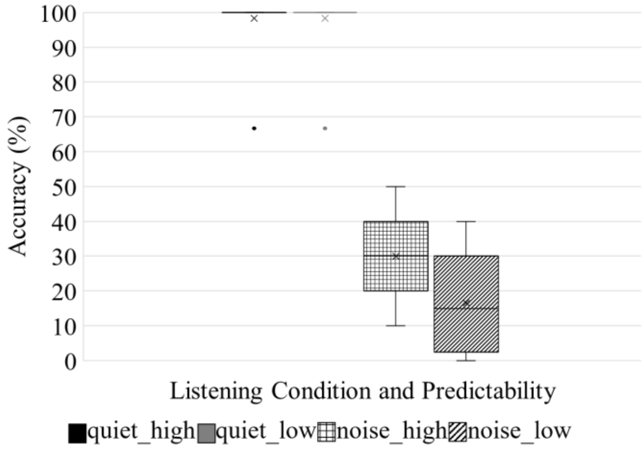


Figure 4 shows the accuracy rates of words that have initial consonants /l/ and /r/. Although the number of words is limited to only three, the near-perfect scores in quiet condition show that participants had little difficulty in perceiving both predictable and unpredictable words if there is no noise in the background (analyzed items: 20 participants x 3 words x 2 predictability = 120 items). However, as expected, accuracy rates decreased dramatically in noisy listening condition (analyzed items: 20 participants x 10 words x 2 predictability = 400 items). Two-way analysis of variance showed a significant interaction between predictability and condition ($p<.01$), and a trend in significance between predictability and condition ($p=.07$). Significant main effects of condition (p<.001) and predictability (p<.001) were observed.

Figure 4: Accuracy rates of words with initial consonants /l/ and /r/ in quiet/noise and predictable/unpredictable conditions (%).



4. DISCUSSION

The present paper reported the results of two experiments: Experiment 1 on the identification of consonants in intervocalic context and Experiment 2 on the identification of English words. Because the present paper does not report the results of native listeners, as reference, previous research with similar experimental designs shows that native listeners achieved approximately 80% in a consonant identification task in a noisy condition at SNR=0dB [6] as well as in unpredictable word identification in a noisy condition at SNR=0dB [12]. Comparison with these studies show that the impact of presence of noise and the absence of predictability is greater for non-native listeners.

Comparison of the overall results of the two experiments in the present study (Figures 1 and 3) show that while task difference (non-word or word) does not seem to affect the overall accuracy of consonant perception (with the exception of unpredictable words in noisy condition in Experiment 2), it affected the perception of /l/ and /r/ (Figures 2 and 4).

Accuracy rates were similar in noisy condition in Experiment 1 (Figure 1, approximately 55-60%) and noisy condition + high predictability words in Experiment 2 (Figure 3, approximately 55%). This result suggests that listeners were able to use predictability in perceiving words accurately (thus also perceiving consonants accurately) in noisy condition. On the contrary, accuracy rate of unpredictable words in noisy condition stands around 20%. Compared to previous studies on experiments on unpredictable words in noisy conditions, this score was unexpectedly low, considering 1) target words in Experiment 2 were lexically familiar to Japanese listeners, and 2) previous research shows native listeners achieved around 80% accuracy rate in the identification of unpredictable words in noisy (SNR=0dB) condition (e.g. [12, 14]). The results obtained from Experiment 2 demonstrates how difficult it is for non-native listeners to perceive unpredictable words in noise.

Although the accuracy rates of /l/ and /r/ in intervocalic context did not differ significantly compared to the results of all the consonants (approximately 70-80% in quiet condition, and approximately 55-60% in noisy condition), the results of /l/ and /r/ show a wide distribution. This may be caused by the relatively wide range of English proficiency level of the participants, assuming from TOEIC scores (lower to upper intermediate level).

Moreover, the results of /l/ and /r/ extracted from Experiments 1 and 2 (Figures 2 and 4) showed clear effect of task differences. Despite the small number of word items analysed in Experiment 2, accuracy rates of words with initial consonants /l/ and /r/ were close to 100% in quiet condition, while accuracy in noise_high condition (predictable words in noise) decreased to 30% (accuracy rates of all words: approximately 55% in noise_high). This result indicates that /l/ and /r/ words are even more adversely affected in noise compared to words that begin with other consonants. Meanwhile, noise_low condition (unpredictable words in noise) did not differ significantly (/l/ and /r/ words: 15%, all words: 20%), which suggests that SNR=0dB causes floor effect for lower to upper intermediate level learners of English, not just for /l/ and /r/ but all words.

5. CONCLUSION

The present paper reported the results of two experiments: identification of English consonants (Experiment 1) and words (Experiment 2) in quiet and noisy conditions by native Japanese listeners. The results indicated that listening condition and predictability affects accuracy, and that task difference (consonant or word identification) affects particularly the perception of /l/ and /r/. Further analyses are necessary regarding differences in individual consonants as well as effect of proficiency.

7. REFERENCES

[1] Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R. & Yamada, T. 2004. Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32, 233–250.

[2] Best, C. T., McRoberts, G. W., Nomanthemba, M. 1998. Examination of Perceptual Reorganization for

Nonnative Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants. *J. Exp. Psychol. Hum. Percept. Perform.* 14, 345–360.
[3] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.43. /http://www.praat.org/
[4] Bradlow, A.R., Alexander, J.A. 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J. Acoust. Soc. Am.* 121(4), 2339-2349.
[5] Cummins, J. 1981. Age on Arrival and Immigrant Second Language Learning in Canada: A Reassessment. *Appl. Linguist.* 11, 132–149.
[6] Cutler, A., Garcia Lecumberri, M. L., Cooke, M. 2008. Consonant identification in noise by native and non-native listeners: effects of local context. *J. Acoust. Soc. Am.* 124, 1264–1268.
[7] Florentine, M. 1985. Non-native Listeners' Perception of American-English in Noise. *Proc. Inter-Noise* 85 1021–1024.
[8] Halle, P. A., Segui, J., Frauenfelder, U., Meunier, C. 1998. Processing of Illegal Consonant Clusters: A Case of Perceptual Assimilation? *J. Exp. Psychol. Hum. Percept. Perform.* 24, 592–608.
[9] Kalikow, D. N., Stevens, K., Elliott, L. L. 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J. Acoust. Soc. Am*. 61, 1337–1351.
[10] Lecumberri, M. L. G., Cooke, M. 2006. Effect of masker type on native and non-native consonant perception in noise. *J. Acoust. Soc. Am.* 119, 2445–2454.
[11] Mayo, L. H., Florentine, M., Buus, S. 1997. Age of second-language acquisition and perception of speech in noise. *J. Speech Lang. Hear. Res.* 40, 686–693.
[12] Meador, D., Flege, J. E., Mackay, I. R. A. 2000. Factors affecting the recognition of words in a second language. *Biling. Lang. Cogn.* 3, 55–67.
[13] Miyawaki, K., Jensins, J.J., Strange, W., Liberman, A.M., Verbrugge, R., Fujimura, O. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics* 18, 331–340.
[14] Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., Abrams, H. B. Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. 2006. *Applied Psychophysics* 27, 465–485.
[15] Takagi, N. & Mann, V. 1995. The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics* 16, 380–406.
[16] Takata, Y., & Nabelek, A. K. 1990. English consonant recognition in noise and in reverberation by Japanese and American listeners. *J. Acoust. Soc. Am.* 88(2), 663–666.