Transcommunication Vol.11-1 Spring 2024
Graduate School of International Culture and Communication Studies

Article

L2 perception of Japanese /Cju/ and /Cjo/ sequences by Mandarin listeners in quiet and multi-speaker babble noise

Yili Liu

Abstract

This study investigated the perception of Japanese /Cju/ and /Cjo/ sequences by native Mandarin listeners under different listening conditions. It further examined the effects of (1) Japanese proficiency, (2) phonetic context, (3) syllable position, and (4) multi-speaker babble noise on non-native speech perception. An identification test was conducted in which native Japanese and Mandarin speaking participants of varying Japanese proficiency levels (beginner, intermediate, and advanced levels) identified /u/ and /o/ in / CjV/ syllables. The test was conducted in both quiet and multi-speaker babble noise environments. The results revealed that, despite variations in Japanese proficiency and preceding consonant acoustic features, most L1 Mandarin listeners could distinguish between /Cju/ and /Cjo/ units. The confusion of the back vowels /u/ and /o/ in /CjV/s, reported in previous studies, was rarely found. Syllable position was confirmed to affect learners' perceptual performance. Furthermore, the significant differences between native and non-native listeners in the noisy condition demonstrate that noise compromised non-native identification more than native identification. In particular, the highly proficient performance exhibited by intermediate and advanced learners in noisy environments suggests that they likely employed similar signal-processing strategies as native listeners. This study contributes empirical evidence for vowel identification in both quiet and noisy environments.

Key words: Non-native, Speech perception, Japanese /Cju/ and /Cjo/ sequences, Noise

1. Introduction

In the field of research on second language acquisition, it has been suggested that the ability to comprehend and speak in a second language (L2) is influenced by similarities and differences in learners' first language (L1) and L2 (Elvin & Escudero, 2019). This influence is known as cross-linguistic influence. Mismatches between the native and target phonemic inventories interfere with the development of L2 speech acquisition. Consequently, some L2 segments are found to be more challenging to acquire than those that already exist in learners' L1 category. For instance, native speakers of Spanish, learning English as an L2, often encounter difficulties with the contrast between English tense and lax vowels /ɪ/ and /i/, as these distinctions are not present in Spanish (Escudero & Boersma, 2004). Similarly, it has been also shown that Japanese learners of English struggle to distinguish English /r/ and /l/, as both sounds are perceived as the same phoneme /r/, i.e. alveolar tap [ɾ](Aoyama et al., 2004).

Current theories and studies on L2 speech perception are generally centered on how L1 influences L2 performance to predict L2 difficulties. Two of the predominant theoretical models aimed at predicting the degree of difficulty in learning non-native sounds based on how these sounds are mapped onto learners' L1 phonetic categories, are the Speech Learning Model (SLM) and the Perceptual Assimilation Model of L2 speech learning (PAM-L2) (Flege,1995; Best & Tyler, 2007). Specifically, SLM posits that L2 sounds similar to an existing L1 category should be more difficult to acquire than those not closely associated with any existing L1 category. On the other hand, PAM-L2 is an extended version of Perceptual Assimilation Model (PAM), originally devised to account for naïve listeners' perception of non-native sounds. In PAM-L2, six assimilation types have been proposed to predict difficulties in L2 speech perception based on the articulatory similarities and dissimilarities between L1 and L2. A larger number of studies support the hypotheses of SLM and PAM-L2 (e.g. Aoyama et al., 2004; Tyler, 2019). Therefore, this study aims to explore whether the difficulty that Chinese learners of Japanese face with regard to vowel discrimination within /CjV/ sequences could be explained by SLM or PAM-L2.

Recently, suprasegmental features, such as the contrasts of short and long vowels have received extensive attention in the field of Japanese speech learning. However, few studies have specifically focused on the acquisition of Japanese vowels by Chinese speaking learners. Native speakers of Mandarin Chinese encounter difficulties in distinguishing between /u/ and /o/ in /CjV/ sequences as observed in words like /kjoositu/ ('classroom'), especially at beginner levels (Lin, 1981; Sugiyama, 1985; Kitamura, 1992). The challenge in distinguishing between /Cju/ and /Cjo/ may be due to the absence of /o/ in Mandarin Chinese. Given that the Japanese vowel /o/ is not present in the Mandarin

Chinese vowel inventory, Chinese speakers often adopt the diphthong /au/ as a similar sound (Liu, 1984; Sugiyama, 1985; Yu, 1985). While much research has focused on the production of /Cju/ and /Cjo/ by Chinese speakers, more evidence is needed regarding the perception of /u/ and /o/ in /CjV/ sequences. Additionally, questions remain about the learning stage(s) at which learners can identify /Cju/ and /Cjo/, as the same error pattern is rarely reported among higher-level learners.

Another factor that affects speech perception is noise. It can affect people's general concentration, for example, environmental sounds such as falling rain may enhance concentration and productivity, while speech sounds like other people's conversations may distract and cause discomfort. With regard to speech perception, noise can also prevent listeners from hearing speech perfectly. While native listeners can take advantage of acoustic-phonetic and contextual cues to compensate for information loss, understanding speech in noise poses a challenge for non-native listeners due to imperfections in their target language knowledge (Lecumberri et al., 2010). Even after extensive exposure to the target language, non-native listeners may not achieve native-like speech recognition (Florentine, 1985). This non-native disadvantage is particularly observed when the contrasts do not occur phonemically in the learners' L1, leading to misunderstandings and confusion (Flege, 2003; Grimaldi et al., 2014). Different syllable positions and acoustic features of neighboring segments also appear to increase the processing load for non-native listeners (Beckman, 1998; Cilibrasi, 2016). Therefore, it is important to explore how non-native listeners handle L2 speech sounds under noisy conditions.

The present study aimed to examine the effects of linguistic proficiency, phonetic context, syllable position, and noise on the identification of Japanese /u/ and /o/ in /CjV/ sequences by L1 Mandarin listeners. In particular, the study addressed whether the L1 Mandarin listeners could distinguish /Cju/ and /Cjo/ pairs and identified which factors pose greater challenges.

2. Literature review

This section reviews relevant research to provide theoretical support and identify the gaps in previous studies. The literature review focuses on comparing the Japanese and Mandarin Chinese vowel systems. Additionally, the research on Japanese /CjV/ sequences and L2 perception in noise is also discussed.

2.1 Vowel system of Chinese and Japanese

Standard Japanese features five vowels /i, e, a, o, u/, all of which exhibit phonetic contrast in duration (short vs. long) and are monophthongal (Maddieson & Disner, 1984; Vance,

2008). Based on articulatory features, Japanese [i] is a high front vowel, almost identical in quality to the cardinal [i]. [e] is a mid-front vowel, with the tongue position falling between IPA [e] and IPA [ɛ]. Japanese [a] falls between the cardinal vowel [a] and [ɑ]. [o] is a weakly rounded vowel with a mid-back tongue position. /u/ is described as a high-back unrounded vowel [ɯ], and it has been reported that the tongue position of [ɯ] becomes more fronted after a front semivowel [j] (Vance, 2008).

On the other hand, Mandarin Chinese exhibits a more complex vowel inventory compared to Japanese. While there is general consensus regarding the nature of high vowels /i/, /y/, and /u/, as well as the low vowel /a/, the description of mid vowels is still under investigation. The variation of the mid vowel [e], [ə], [o], [ɤ], [ɛ] and [ɔ] occurs in different and complementary contexts; thus, they are treated as allophones of a single phoneme (Wiese, 1997). The most widely accepted analysis of the Mandarin Chinese vowel system includes five vowels /i, y, a, ə, u/ (Duanmu, 2003; Lin, 2007). There are two front high vowels [i] and [y]. The tongue position of Mandarin [i] is similar to that of the cardinal [i], while [y] is produced with a tongue position similar to [i] but with rounded lips. The low central [a] is fairly similar to the Japanese [a]. In contrast to the Japanese back unrounded [ɯ], Mandarin [u] is rounded and positioned further back. [ə] is an unrounded mid vowel that is tenser and falls further back than the English schwa. Furthermore, Mandarin Chinese features four major diphthongs in: /ai, au, ei, ou/.

Since Japanese /a/ and /i/ share similar articulation with Mandarin Chinese, these two vowels appear to be relatively easy for Mandarin speakers to acquire (Sugiyama, 1985; Yu, 1985; Wang et al.,1987; Kitamura, 1992). However, L1 Mandarin learners of Japanese often encounter difficulties when producing the less similar Japanese vowels /e/, /u/ and /o/. For example, /e/ may be assimilated to /ei/, /ai/ or /ie/(Chen, 1962; Mizutani & Otsubo,1971; Kitamura, 1992). Similarly, the assimilation of /u/ and /o/ to Chinese vowel categories has been commonly observed in previous studies. There is still argument regarding the difficulty of acquisition of the Japanese /o/. Liu (1984) and Wang et al. (1987) describe /o/ is easy to acquire for Mandarin speakers, while Kitamura (1992) holds the opposite opinion, stating that since /o/ only occurs as a complementary allophone in Mandarin Chinese, its acquisition is challenging for them. Therefore, further investigation should be conducted to provide more evidence for L2 Japanese vowel acquisition.

2.2 Previous studies on Japanese /CjV/ sequences

Japanese has a group of palatalized consonants /Cj/. Due to the articulatory features of the semi-vowel /j/, the tongue position of the preceding consonant in /Cj/ assimilates, approaching the hard palate. Palatalized consonants are associated with vowels /a/, /u/, and /o/ to form /CjV/ sequences referred to as *yo-on* in Japanese (Labrune, 2012).

The difficulties associated with /CjV/ sequences have been widely discussed in both L1 and L2 acquisition. L1 Japanese learners at an early stage of language acquisition tend to reduce /j/ or insert a vowel to cope with the production difficulties of /CjV/ sequences (Paradis, et al., 1985; Endo, 1990; Tsurutani, 2004). However, L2 learners of Japanese also suffer from producing and perceiving /CjV/ sequences. For example, vowel epenthesis can be observed in L1 English and Russian learners' utterances (Tsurutani, 2004; Watanabe, 2011). Korean speaking learners of Japanese tend to simplify /CjV/ sequences, producing, for example, /nu/ instead of /nyu/ (Sukegawa, 1993; Kondo, 2011).

In contrast to English or Korean speakers, a typical error pattern for L1 Mandarin learners of Japanese, especially among beginner learners, is segment substitution, such as /tosjokan/ ('library') → /tosjukan/ (Sukegawa, 1993; Liu, 1983). Lin (1981) attributed the confusion between /u/ and /o/ in /CjV/ sequences to unfamiliarity with the target language and cross-linguistic influences. Sugiyama (1985) suggested that 'although Japanese do not contrast /u/ and /ɯ/, beginner Mandarin learners of Japanese tend to apply Chinese rounded [u] to pronounce Japanese unrounded [ɯ], leading to this vowel confusion.' However, the above studies attempted to explain the issue through comparisons of the phonological features of Japanese and Mandarin, relying only on questionnaires or researchers' observations rather than experimental data. Therefore, further exploration of perception and empirical evidence is required to provide a full picture of the perception of /CjV/ sequences.

Furthermore, Kitamura (1992) indicated that the confusion of the Japanese /Cju/ and /Cjo/ sequences was caused by the corresponding sound /Cjou/ in Mandarin. Since /o/ is absent in Mandarin Chinese, Mandarin speaking learners of Japanese adopt the 'similar' diphthong /ou/ in their L1 to produce both Japanese /Cju/ and /Cjo/. Zhu (2011) compared the articulatory features of Japanese /ju/ and /jo/ with Chinese /jou/ using MRI to examine why Japanese /ju/ produced by L1 Mandarin speakers is misheard as /jo/. The results revealed that due to coarticulation, the Japanese back high vowel [ɯ] tends to be more frontal and central in /ju/, but Mandarin /u/ does not change its tongue position in any case. These findings indicate that differences in acoustic features between Japanese and Mandarin lead to mishearing from the perspective of articulatory phonetics. Nevertheless, it remains unclear whether the phonetic features of preceding consonants affect non-native speech perception.

2.3 L2 speech perception in noise

Conversations often take place in noisy environments rather than quiet settings. Typically, native speakers are able to deal with these imperfect conditions by taking advantage of their native linguistic knowledge. However, non-native speech perception in

adverse conditions poses a considerable challenge, particularly for L2 listeners. Takata and Nábělek (1990) reported that L2 listeners were more adversely affected by listening conditions and reverberations than native speakers. This disparity arises from native speakers' ability to utilize linguistic cues, such as contextual and semantic information, to compensate for information loss.

The performance of non-native listeners in noise varies with different maskers. Lecumberri and Cooke (2006) conducted an identification task to explore the effect of three maskers, i.e. stationary noise, multi-speaker babble noise, and competing speech, on consonant perception. The results indicated that the competing speech had the least effect, but multi-speaker babble noise significantly affected the perception of both native and non-native listeners. Furthermore, it was revealed that L2 learners were more adversely hindered by energetic masking than native speakers.

Additionally, the relationship between learners' L2 proficiency and their perception in noise has been discussed in previous studies. Kilman et al. (2014) examined the influence of L2 proficiency on speech perception under different maskers, suggesting that L2 proficiency decisively affects speech recognition in adverse conditions. Higher proficiency L2 learners had an advantage in L2 speech perception compared to less proficient listeners. However, Masuda and Arai (2013) reported contradictory results, stating that L2 proficiency did not significantly affect speech perception under noisy conditions. These divergent findings indicate that the effect of learners' proficiency under noisy conditions requires further investigation.

Moreover, it is crucial to explore the effect of word position on the perception of /Cju/ and /Cjo/ units. Word initial positions are described as marked and perceptually salient, while word-final positions are less active and often undergo deletion (Beckman, 1998; Smith, 2002). Cilibrasi (2016) examined how word position affects perception and suggested that consonant clusters in nonword initial positions are detected more accurately than those in other positions. Therefore, under degraded auditory conditions, it is presumed that the identification of the word-final position vowel is more challenging than the word-initial position.

2.4 Research Questions

Despite a substantial body of literature on L2 vowel speech perception (Flege et al., 1997; Flege, 1993; Bohn & Flege, 1990; Tsukada, 2011) and comparative studies between native and non-native listeners in speech recognition under adverse conditions, respectively (Nábělek & Donahue, 1984; Takata & Nábělek, 1990; Cutler et al., 2004; Bradlow & Alexander, 2007; Masuda & Arai, 2010), there has been insufficient research on both aspects. Furthermore, the available evidence regarding the factors influencing segmental

recognition remains limited.

This study aims to investigate vowel perception in Japanese /CjV/ sequences by Mandarin Chinese listeners in both quiet and multi-speaker babble noise. The goal is to explore possible factors that may affect vowel identification. To address these objectives, the following research questions were posed.

Research Question 1: Do L1 Chinese learners of Japanese distinguish between /u/ and /o/ in /CjV/ sequences under different listening conditions? If so, would non-native listeners demonstrate native-like performance?

Research Question 2: What factors affect Mandarin speaking learners' perception of Japanese /Cju/ and /Cjo/ sequences in quiet and noisy conditions?

2-a: Will Japanese proficiency levels affect learners' ability to perceive /Cju/ and /Cjo/ sequences in quiet and noisy conditions?

2-b: Will phonetic context impact learners' ability to perceive /Cju/ and /Cjo/ sequences in quiet and noisy conditions?

2-c: Will word position affect learners' ability to perceive /Cju/ and /Cjo/ sequences in quiet and noisy conditions?

Research Question 3: How and to what extent does multi-speaker babble noise influence native Japanese and non-native Mandarin listeners' ability in identifying Japanese /Cju/ and /Cjo/ sequences?

3. Methods

3.1 Participants

Sixty-one participants were recruited from Waseda University in Tokyo and two universities[1] in China. The control group comprised nine Japanese native speakers (2 males, 7 females) with a mean age of 21.6 years (*SD* = 2.1). The learner group included fifty-two Mandarin speaking participants aged 18-37 (4 males, 48 females, *mean* = 20.7, *SD* = 3.4), categorized into 'Beginner', 'Intermediate' and 'Advanced' subgroups based on their score in an online Japanese proficiency test[2] and the length of Japanese learning.

The beginner group consisted of first-year university students who had received formal Japanese instruction for approximately eight months at the above universities in China. The intermediate-level group comprised nineteen second or third-year students with 14-48 months of Japanese instruction from the same institutions. The advanced group included thirteen fourth-year and graduate school students with over 46 months of exposure to Japanese. None of the participants had a history of hearing disorder, and all L2 learners had no self-reported experience of studying in Japan. Table 1 shows the char-

acteristics of participants in each group.

Table 1: Characteristics of participants in each group

Subject		**Learner group**		**Control group**
Proficiency	Beginner	Intermediate	Advanced	Native
No. of subjects	20	19	13	9
Mean age	18.3 (SD = 0.6)	20.2 (SD = 0.6)	24.9 (SD = 4.3)	21.8 (SD = 2.1)
Length of study (month)	8 (SD = 0)	28.4 (SD = 13.0)	81.4 (SD = 46.2)	---

3.2 Stimuli

The target stimuli consisted of the nonword disyllables tokens /CaCjV/ or /CjVCa/ with eight Japanese consonants: four plosives ([p], [b], [k], [ɡ]), two affricates ([tɕ], [dʑ]), and two fricatives ([ɕ], [ç]). The /CjV/ sequence occurred in either the initial or final position.

Two stimulus sets were created for two different listening conditions: one in quiet and the other in multi-speaker babble noise. For the noisy condition, a multi-speaker babble noise served as the masker. A 'Signal-to-Noise Ratio (SNR)' of -6dB was chosen, as it has been shown to have a major influence on listeners' performance (Miller, 1947; Festen & Plomp, 1990; Simpson & Cooke, 2005; Van Dommelen & Hazan, 2010). The stimulus set in noise was generated by synthesizing the quiet stimulus set with multi-speaker babble noise with SNR= -6 dB using a Praat plugin, 'Praat Vocal Toolkit' (Corretge, 2012). The total number of tokens per set was 166 (96 stimuli and 70 fillers), resulting in 332 tokens for the entire stimulus set (166 × 2 conditions).

The stimuli were produced in a recording session lasting approximately 30 minutes by one male and one female native Japanese speaker, both from the Saitama Prefecture in Kanto region. They were instructed to read a list of stimuli presented on a computer screen within the carrier sentence (/korewa_____desu/ 'This is_____') three times at a natural speaking rate. All stimuli were read with an accent on the first syllable. Monaural soundtrack recordings were made in a soundproof booth using a dynamic microphone (SONY F-780) connected to an MTR recorder (ZOOM R8) at a sampling rate of 44.1 kHz and 16-bit digitalization.

3.3 Experimental Procedure

A forced-choice identification task was conducted using ExperimentMFC 8 in Praat version 6.0.38 (Boersma & Weenink, 2018), controlled by a computer. The task included a total of 664 trails (332 tokens × 2), and all tokens were presented in a fully randomized order.

Participants, both native Japanese listeners and Mandarin Chinese learners, listened to the trails individually over headphones in a quiet classroom. The stimuli were presented on a computer screen, and participants were instructed to decide which word was more similar to what they had heard. They indicated their choice on the computer interface, with two options represented in *hiragana*. Participants had the option to replay the stimulus tokens once by clicking the 'replay' button. The task was self-paced, and participants moved to the next trial by clicking either of the option buttons. The interface language was Chinese for the Mandarin listeners and Japanese for the native Japanese listeners. The average duration of the identification task was approximately 30 minutes. Reaction time was measured to ensure that each participant clicked the option button after the sound had been played. If multiple negative reaction times were recorded (e.g., -0.31s), the participant was asked to redo the task; otherwise, he/she could not receive compensation. All participants volunteered for the study and received JPY 1,000 (RMB 60) as compensation.

3.4 Data analysis

The performance was quantified as the average percent correct identification rate for each participant and stimulus to facilitate statistical analyses. Welsh's test and analyses of variance (ANOVAs) were conducted on accuracy, with target syllable position (initial and final), voicing of preceding consonants (voiceless and voiced), manner of articulation of preceding consonants (plosive, fricatives and affricate) as within-subject factors. Japanese proficiency levels (beginner, intermediate and advanced) were treated as a between-subject factor. In cases where a significant difference was observed as a main effect, a Bonferroni post-hoc test was conducted for further exploration. To investigate potential factors influencing the perception of /u/ and /o/ in /CjV/ sequences, a test of independence with $\alpha = .05$ as a criterion for significance was employed. All the data analyses were performed using RStudio 1.4.1717.

4. Results

4.1 Vowel identification of native Japanese and non-native Mandarin Chinese listeners in different listening conditions

First, the identification of /u/ and /o/ in /CjV/ sequences by native Japanese and non-native Mandarin Chinese listeners was examined. Percentage identification rates were calculated for each listener in both quiet and noisy conditions. The mean identification rates and distributions for the native and Chinese listeners are presented in Figure 1.

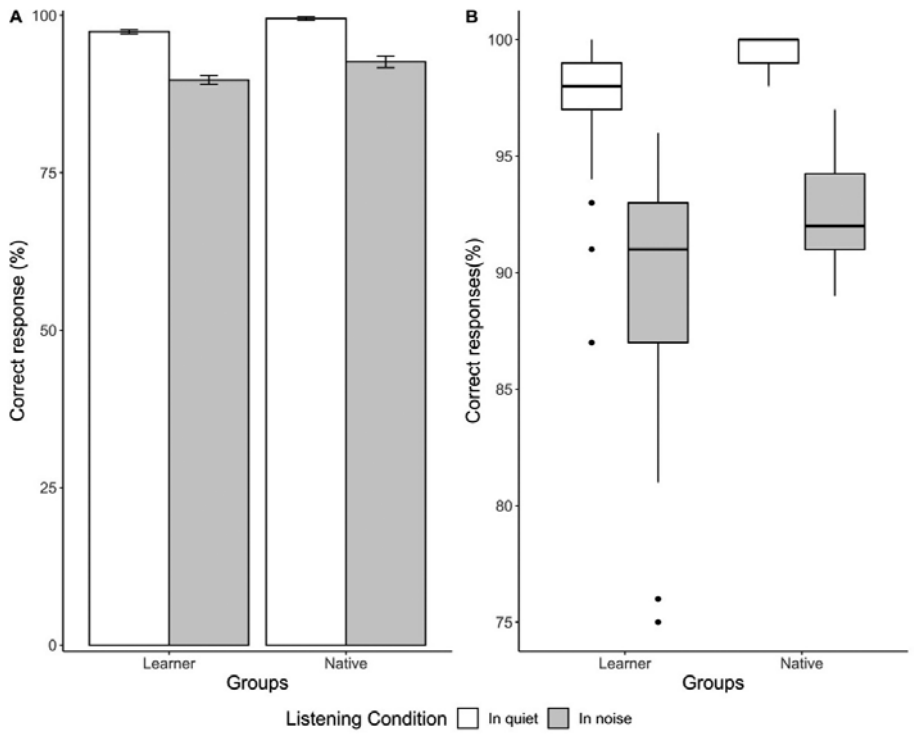

Figure 1. Comparison of the average accuracy (±SE) of learner group (N= 52) and native group (N = 8) in quiet and in noise. The middle line shows the median, and the dots are outliers.

Native listeners' performance in a quiet condition was at a ceiling level (99.5%). Non-native listeners performed well under the quiet condition (97.4%). In noise, although the performance of both native and non-native listeners declined (Native: 92.6%, Learner: 89.7%), the identification rates remained high. The native listeners outperformed the Chinese listeners both in quiet and in the noisy babbles. The results of Welsh's t-test, with L1 as a between-subject factor, revealed that the effect of L1 was highly significant in both the quiet condition [t (33) = -5.01, $p < 0.01$] and in noise [t (16) = -2.52, $p = 0.023$].

4.2 The effect of proficiency

L2 learners' performance in discriminating /u/ and /o/ in /CjV/ sequences was analyzed and plotted to gain insight into the response patterns of the non-native listeners with various Japanese proficiency levels. Figure 2A shows the average vowel identification accuracy under different listening conditions. All L1 Mandarin listeners performed well in quiet (Beginner: 97%, Intermediate: 98%, Advanced: 98%), while their identification accuracy dropped as expected under multi-speaker babble noise. The distribution of average accuracy by learners' proficiency levels in the different listening conditions is shown in

Figure 2B. The small standard errors of the intermediate and advanced groups in quiet indicate that all learners in these groups had relatively equal performance. In contrast, in noise, the comparatively greater standard errors of the beginner and advanced groups suggest that the average accuracy of perception varied greatly between individuals.

Contrary to expectations, the results of the ANOVAs showed no significant differences between the proficiency groups in quiet [F (2, 49) = 1.92, p = .16]. In multi-speaker babble noise, the identification accuracy of all proficiency levels decreased (Beginner: 88%, Intermediate: 91%, Advanced: 91%), but while the identification performance of the beginner level learners was poorer than that of the learners in the other groups, the effect of proficiency was not significant [F (2, 49) = 2.68, p = .08].

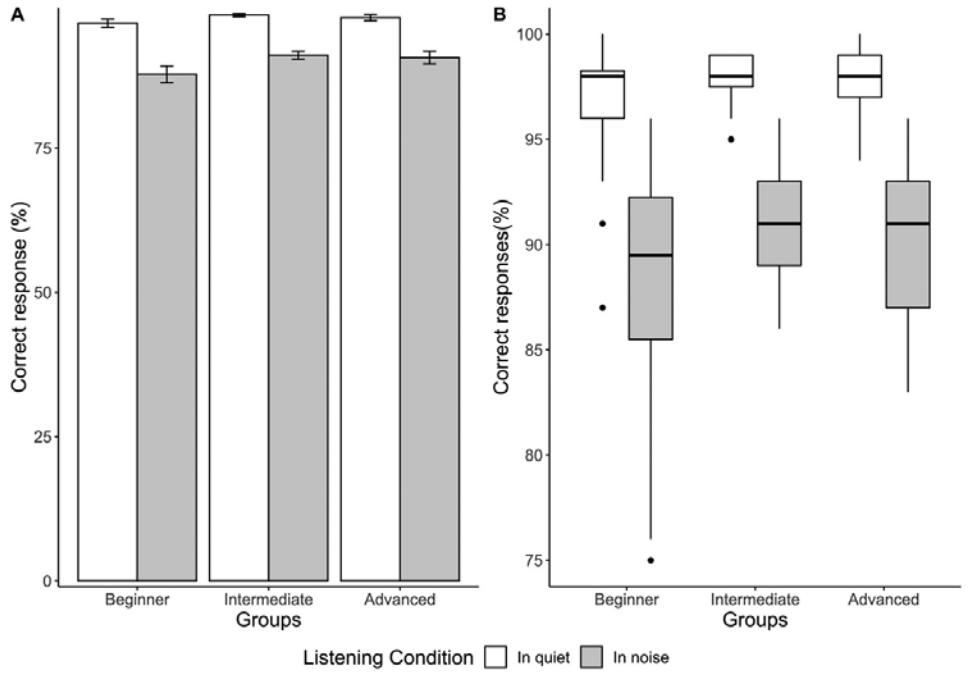

Figure 2. Comparison of the average accuracy (±SE) of the beginner learners (N= 20), intermediate level learners (N = 19) and advanced level learners (N = 13) in different listening conditions and the distribution of the identification performance of three proficiency-level learners in quiet and noisy conditions. The middle line shows the median, and the dots are outliers.

4.3 The effect of phonetic context

The articulatory features of preceding consonants, specifically voicing and manner, were examined for their impact on the identification of /Cju/ and /Cjo/ sequences. Figures 3A and 3B show the average accuracy of voicing and manner of articulation. In the quiet condition, learners' identification scores were consistent, with high accuracy for both

voiceless (97%) and voiced (98%). The independent sample t-test indicated a non-significant difference in voicing in the quiet condition [t (98) = .22, p = .83]. This pattern persisted in the noisy condition, where accuracy for voiceless (90%) and voiced (89%) showed no significant difference, as confirmed by Welsh's t-test under multi-speaker babble noise [t (101) = -.98, p = .33].

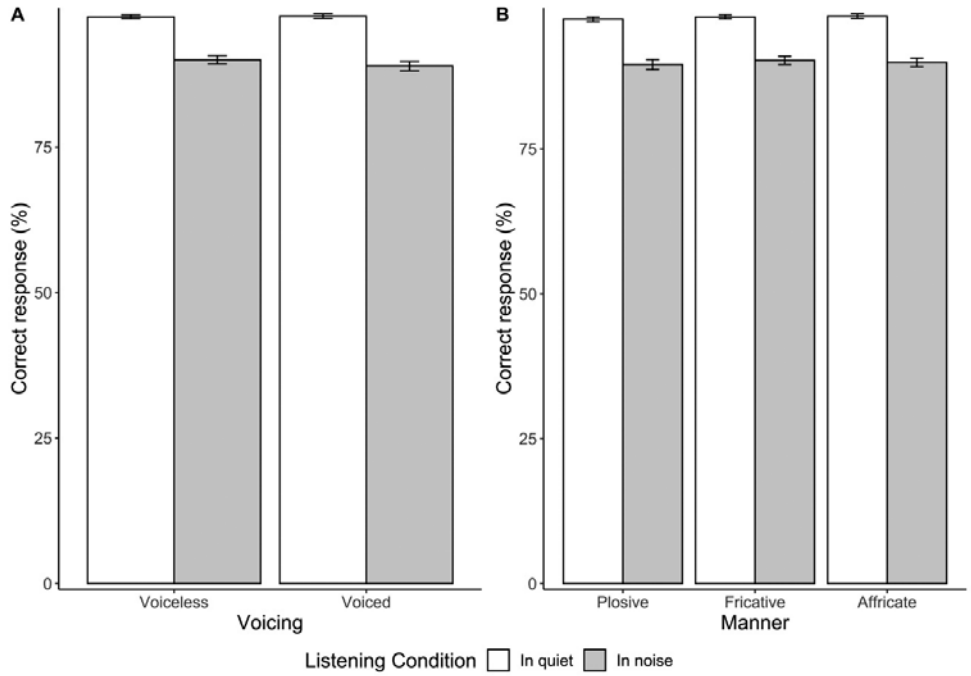

Figure 3. Comparison of the average accuracy (±SE) of Chinese listeners by voicing and manner of articulation in different listening conditions.

The general response patterns of plosives, fricatives and affricates were similar in both quiet (plosive = 97%, fricative= 98%, affricate = 98%) and noisy (plosive = 90%, fricative = 90%, affricate = 90%) conditions. The one-way ANOVA results indicated that there were no significant differences in the manner of articulation, regardless of listening conditions [Quiet: F (2, 153) = .51, p = .6; Noise: F (2, 153) = .22, p = .8]. In essence, this study did not find evidence that phonetic context affects the identification of /u/ and /o/ in *yo-on.*

4.4 The effect of syllable position

To investigate whether syllable position affects the identification of vowels /u/ and /o/ in *yo-on,* an independent sample t-test was conducted. Figure 4A shows the comparison of identification performance in word-initial and final positions under different listening conditions. In quiet conditions, L2 learners demonstrated better performance in the final

position than that in the initial position (initial position: 96%, final position: 98%). However, the accuracy significantly dropped in noise, with the accuracy for the word-initial position stimuli being higher than that for the word-final stimuli (initial position: 93%, final position: 88%). The accuracy distribution for the different syllable positions is shown in Figure 4B. In multi-speaker babble noise, the performance of L2 learners exhibited considerable variability compared to that in the quiet condition.

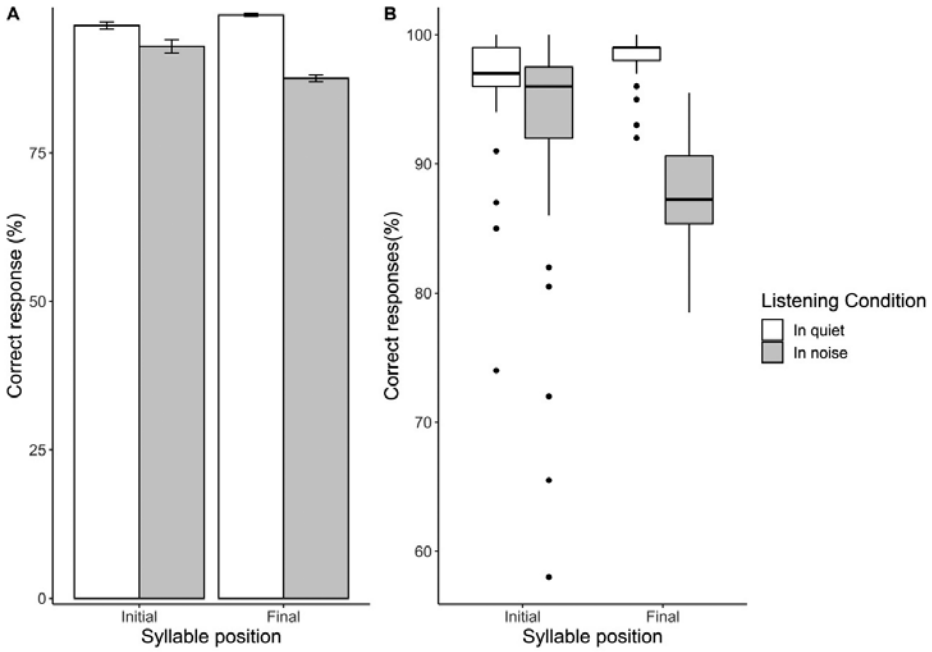

Figure 4. Comparison of the average accuracy (±SE) for the word-initial and word-final position in different listening conditions by Chinese listeners. The distribution of identification accuracy in different syllable positions in quiet and noisy conditions. The middle line shows the median, and the dots are outliers.

Statistical analysis revealed a significant main effect for syllable position in both quiet [t (65) = 2.66, $p = 0.01$] and noise [t (76) = -4.21, $p < 0.01$]. The findings indicate that in noise, learners are more prone to perform poorly when /CjV/ sequences occur in the word-final position. Conversely, the opposite tendency was observed in the quiet condition.

4.5 The effect of noise

To explore the impact of noise on the Mandarin listeners and native Japanese listeners, comparisons were made between the non-native listeners of different proficiency levels and the native Japanese listeners. Figure 5A shows the average accuracy of the four

listener groups in various listening conditions. Overall, multi-speaker babble noise had a detrimental effect on the perception of /CjV/ sequences for both the native and non-native listeners, while both groups demonstrated high accuracy in the quiet condition.

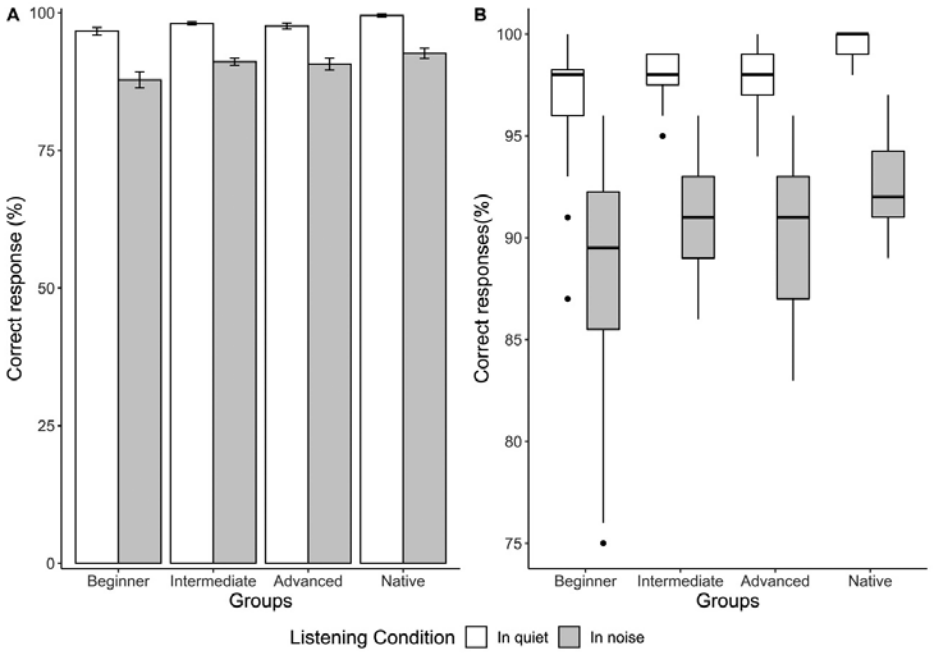

Figure 5. Comparison of the average accuracy (±SE) of Chinese listeners (N = 52) and native listeners (N = 8) in different listening conditions. The distribution of identification accuracy of the identification performance of learners and native listeners in quiet and in noise. The middle line shows the median, and the dots are outliers.

The accuracy distribution of the learners' performances is shown in Figure 5B. In the quiet condition, the intermediate learner group and the native group showed lower mean error, while the performance of the beginner and advanced learners varied more. Conversely, the relatively large errors in noise indicate more variation in individual accuracy performance, especially for beginner and advanced-level learners. T-tests were conducted for the listening conditions (quiet, noise) and groups (learner, native), revealing that both listening conditions [t (85) = -11.1, p < .01)] and groups [t (24) = -2.2, *p* = .04] were significant factors influencing the perception of /Cju/ and /Cjo/.

Although all learner groups performed well in quiet, a comparison was made between learners of different proficiency levels and the native listeners. Additionally, an examination was conducted to determine whether noise adversely affects non-native listeners more than native learners. The results of the one-way ANOVA revealed significant differences between the native group and the three learner groups in quiet [F (56) =

33.52, $p < .01$]. Pairwise mean comparisons showed that while all learners were highly accurate in identifying /Cju/ and /Cjo/ units, they were still far from the native-like level (Beginner: [$p = .01$], Intermediate: [$p = .01$], Advanced: [$p = .02$]). In multi-speaker babble noise, similar results were found for the groups ($F(56) = 2.881$, $p = .03$). However, the results of the post-hoc test revealed that the average accuracy of the learners from the beginner group ($p = .03$) was significantly different from that of the native listeners, while the intermediate and advanced groups were not. This suggests that the intermediate- and advanced-level learners demonstrated native-like speech recognition in multi-speaker babble noise. Nevertheless, for the lower-proficiency learners, native-like perceptions are still difficult to acquire. Adding noise to the stimuli caused a bigger drop in identification rates for the learner groups compared to the native listeners, although the difference was not large (mean quiet-noise difference of Chinese listeners:7.7%; mean quiet-noise difference of native Japanese listeners:6.9%). The results of Welsh's t-test suggested that quiet-noise difference was highly significant between the learner groups and the native group [$t(24) = -4.6$, $p < 0.00$].

5. Discussion

This study investigated the perception of vowels in /CjV/ sequences by native and L1 Mandarin listeners in quiet and multi-speaker babble noise. The first objective was to explore whether L1 Mandarin learners of Japanese encounter challenges in identifying /Cju/ and /Cjo/ sequences under different listening conditions. As noted in the introduction, several lines of evidence indicate that confusion between Japanese /u/ and /o/ in /CjV/ sequences is frequently observed among Chinese learners of Japanese, particularly those at lower proficiency levels (Lin, 1983; Liu, 1983). Consequently, it was expected that beginner-level learners would perform poorly in both quiet and noisy conditions. Contrary to expectations, the findings were inconsistent with previous observations regarding the difficulty in discriminating between /Cju/ and /Cjo/ sequences. While the native listeners outperformed the non-native Mandarin listeners, the Mandarin listeners still demonstrated excellent performance even in the noisy condition. Two potential explanations for these findings exist. First, during the participant selection stage, although efforts were made to consider individual variability, factors such as auditory acuity, motivation, or the amount of L2 input outside the classroom was difficult to control for. Questionnaire items on motivation revealed an overwhelming desire (87%) to sound like a native speaker. Therefore, it is likely that the learners who participated in the experiment were very conscious of identifying /u/ and /o/ in /CjV/ sequences regularly, leading to the high accuracy rates. Furthermore, previous work on L2 speech often compared the phonological

similarities and differences between L1 and L2 to explain L2 difficulties (Flege, 1999; Aoyama et al., 2004). In the Speech Learning Model (SLM), Flege (1988; 1992; 1995) argued that the less similar an L2 category is to an L1 category, the more likely it is to be equated with a new category that is considered easier for L2 learners to acquire, and vice versa. In other words, regardless of learning stage, learners' perceptions are influenced by their L1 phonetic and phonological features. With an awareness of the differences between L1 and L2, learners would make some adjustments to categorize L2 sounds. Accordingly, the successful performance of the Mandarin listeners in the current study may indicate that the participants recognized the dissimilarity between Mandarin /ou/ and Japanese /o/ and then established a novel category /o/ to help them identify /u/ and /o/ in /CjV/ units.

The second research question explored the effects of Japanese proficiency, phonetic context, and syllable position on the identification of /u/ and /o/ in /CjV/ sequences. According to Liu (1983), the confusion of /Cju/ and /Cjo/ sequences was frequently observed among L2 learners with lower proficiency. Consequently, it was hypothesized that beginner-level learners would exhibit less accuracy than highly proficient learners in both listening conditions. However, the results of the identification task did not support this hypothesis. In the quiet condition, all L1 Mandarin listeners, regardless of their Japanese proficiency levels, demonstrated little difficulty in discriminating the /Cju/ and /Cjo/ units. In the noisy condition, a decrease in accuracy was observed for all learner groups. However, statistically, no significant difference was found between the learner groups. This finding aligns with literature indicating that language proficiency level plays an important role in speech perception in noise (Kilman et al., 2014) but contrasts with the findings of Masuda and Arai (2013). Masuda et al. (2013) found that listeners with low and high-level proficiency performed similarly in English consonant identification in quiet and noisy conditions. The results of the current study might be attributed to an unbalanced sample size, as there were more learners with high TOEIC scores than those with low scores. However, further empirical evidence is needed to explain the effect of language proficiency on vowel identification in noise. Additionally, L2 perception in challenging conditions is related to the working memory capacity of the listener (Rönnberg et al., 2008). For learners with low proficiency, the process of decoding L2 speech in noise has been shown to require more effortful than that in a laboratory environment (Kilman et al., 2014). Therefore, it is not surprising that the beginner learners in our study performed more poorly in multi-speaker babble noise. In a post-experiment interview with the native listeners, they expressed confidence in their performance but acknowledged the challenge of the task in noise without appropriate contexts and lexical meanings. The literature on segment perception in noise suggests that nonsense words have fewer

linguistic cues, resulting in lower percent-correct performance (Lewis et al., 2010). Therefore, the similar performance in multi-speaker babble noise between the more proficient learners and native listeners confirmed the findings of previous studies and emphasized the importance of linguistic cues. Furthermore, it was expected that the accuracy of /CjV/ sequences occurring in the word-final position would be lower than in the initial position regardless of listening conditions because segments in the word-final position are less salient and more prone to deletion compared to those in the word-initial position (Smith, 2002). This hypothesis was partially supported: as expected, word-final vowel identification was more demanding than word-initial vowel identification in noise. However, in the quiet condition, accuracy was higher in the word-final position than in the word-initial position. One possible interpretation of this puzzling result is that multi-speaker babble noise had a greater effect on the final syllable than on the initial or middle syllable. However, the amount of data was too limited to be conclusive, so further investigation and convincing evidence is required.

The last research question assessed how multi-speaker babble noise affects native and non-native listeners. The current results align with previous studies, indicating that noise interference has a greater effect on non-native listeners than on native listeners (Mayo et al., 1997; Takata et al., 1990). Both native and non-native groups exhibited significantly poorer performance in the multi-speaker babble noise compared to the quiet condition. However, in the noisy condition, no significant differences were found between the high proficiency level learner group and the native group. In other words, the results indicated that, except for the beginner level listeners, both the native and non-native listeners were similarly affected by the multi-speaker babble noise to a similar extent. Length of Residence (LOR) and age of L2 acquisition have been explored in numerous previous studies (Aoyama et al., 2004; MacKain et al., 1981; Mayo et al., 1997; Masuda, 2016). Since the participants in this experiment had no experience studying in Japan, the effect of LOR on vowel identification in noise may not be relevant to the results. Mayo et al. (1997) suggested that L2 learners who start studying an L2 at an early age have an advantage in recognizing speech in noisy conditions. However, all the L2 learners in the current study started learning Japanese after their puberty. The difference in performance in noise between the beginner level learners and the more proficient learners may be explained by the quality and quantity of the input (Masuda, 2016).

An intriguing finding in this study is that, regardless of listening conditions, intermediate-level learners exhibited better perceptual performance than the advanced-level learners. One possible explanation for this observation is that the curriculum for intermediate-level learners (second- and third-year undergraduate students) included extensive listening exercises and instructions. In contrast, the advanced level learners (fourth-year

undergraduate and graduate students) may have focused more on other aspects of learning such as job hunting, resulting in shorter exposure to Japanese. Consequently, a reducing exposure to Japanese may have contributed to less accurate identification for the advanced-level learners.

Lastly, there was a perceptual difference in noise by the L2 learners between the stimuli recorded by the female speaker and the male speakers [$t(94) = 5.32$, $p < .01$]. A possible explanation for this finding could be that speaker variability poses a greater challenge for non-native listeners than for native listeners. However, since only one male and one female speaker participated in the recording session, future studies should include a more diverse set of stimuli speakers to explore the impact of speaker variability on vowel identification in noise.

6. Conclusion and Future Direction

This study investigated how L1 Mandarin learners of Japanese perceive /Cju/ and /Cjo/ sequences in both quiet and in noisy environments. The results indicate that Mandarin learners of Japanese, across Japanese proficiency levels, did not encounter difficulties in identifying the vowels /u/ and /o/ in /CjV/ sequences in quiet. Consequently, no developmental progression from beginner to advanced level was found. In contrast, the discrimination of /Cju/ and /Cjo/ units was affected by multi-speaker babble noise for all listeners, with no significant proficiency-related impact. In addition to proficiency, syllable position and phonetic context were also considered as factors influencing vowel identification. Perceptual performance varied with the position of the /CjV/ units under the different listening conditions, although the effect of the phonetic context was not observed. This study provides empirical evidence that multi-speaker babble noise has a greater impact on non-native identification than on native identification. The advanced learners demonstrated a similar level of performance to the native speakers. It is speculated that higher proficiency-level listeners may employ similar signal processing strategies to native speakers in noise, although further evidence is needed and will be explored in future studies. The assimilation of Japanese /u/ and /o/ in /CjV/ sequences by Mandarin speakers and the extent to which they assimilate this to Chinese segments will also be subjects of investigation.

Endnotes

1 Learners of Japanese were enrolled at Liaoning University and Dalian University of Foreign Languages in China.

2 Japanese Computerized Adaptive Test' (J-CAT) is an adaptive test which has been verified by many academic institutions and widely used in various universities, such as at the Japanese education center in Waseda University. Based on the interpretation of J-CAT score, in this study, subjects who got a score from 0 to 150 were labeled as 'Beginner' level, from 151 to 250 were 'Intermediate' level and 250 and higher were 'Advanced' level.

References

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(2), 233-250.

Beckman J. N. (1998). *Positional faithfulness*. MA: University of Massachusetts Amherst, Ph.D. dissertation.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech. *Language experience in second language speech learning: In honor of James Emil Flege*, 17, 13.

Bohn, O-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied psycholinguistics*, 11(3), 303-328.

Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer*. Retrieved from http://www.praat.org/

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, 121(4), 2339-2349.

Chen, X. (1962). Modern Japanese grammar (Vol.1), Commercial Press.

Cilibrasi, L. (2016). *Word position effects in speech perception*. University of Reading. Ph.D. dissertation.

Corretge, R. (2012). Praat Vocal Toolkit. Retrieved from http://www.praatvocaltoolkit.com.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, 116(6), 3668-3678.

Duanmu, S. (2000). *The Phonology of Standard Chinese.* Oxford: Oxford University Press.

Endo, M. (1990). The acquisition process of Japanese syllables in reading and writing by Japanese children, *Studies of educational psychology*, 38(2), 108-117.

Elvin, J. & Escudero, P. (2019). Cross-Linguistic Influence in Second Language Speech: Implications for Learning and Teaching. *Cross-Linguistic Influence: From Empirical Evidence to Classroom Practice*. 1-20.

Escudero, P. & Broersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26(4), 551-585.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, 88(4), 1725-1736.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, 84(1), 70-79.

Flege, J. E. (1992). The intelligibility of English vowels spoken by British and Dutch talkers. *Intelligibility in speech disorders: Theory, measurement, and management*, 1, 157-232.

Flege, J. E. (1993). Production and Perception of a novel, second-language phonetic contrast, *The Journal of the Acoustical Society of America*, 93(3), 1589-1608.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 92, 233-277.

Flege, J. E., Bohn, O-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics*, 25(4), 437-470.

Flege, J. E. (1999). The relation between L2 production and perception. *In Proceedings of the XIVth International Congress of Phonetics Sciences*, 1273-1276.

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. The Journal of the Acoustical Society of America, 106(5), 2973-2987.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. *Phonetics and phonology in language comprehension and production: Differences and similarities*, 6, 319-355.

Florentine, M. (1985). Non-native listeners' perception of American-English in noise. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, InterNoise85, Munich GERMANY, 1021-1024, Institute of Noise Control Engineering.

Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., & Brattico, E. (2014). Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: a behavioral and ERP study. *Frontiers in human neuroscience*, 8, 279.

Han, F. (2013). Pronunciation problems of Chinese learners of English. *ORTESOL Journal*, 30, 26-30.

Kilman, L., Zekveld, A., Hällgren, M., & Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Frontiers in psychology*, 5, 1-9.

Kitamura, Y. (1992). The pronunciation of Chinese learners of Japanese: focus on vowels, *Bulletin of Tokai University*, 12, 13-21.

Kondo, M. (2012). The special and general transfers of first language for learners of Japanese in acquisition process, *Bulletin of the Graduate Division of Literature of Waseda University*. 3, 21-43.

Labrune, L. (2012). *The Phonology of Japanese*. Oxford: Oxford university press.

Lecumberri, M. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, 119(4), 2445-2454.

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech communication*, 52, 864-886.

Lewis, D., Hoover, B., Choi, S., & Stelmachowicz, P. (2010). The relationship between speech perception in noise and phonological awareness skills for children with normal hearing. *Ear and Hearing*, 31(6), 761.

Lin, Y. (2007). *The Sounds of Chinese*. Cambridge: Cambridge university Press.

Lin, Z. (1981). The education of Japanese pronunciation during the beginner period: The perceptual confusions and transfer of first language for Chinese beginner level learners of Japanese, *Japanese language education*, 45, 133-144.

Liu, S. (1983). The frequent pronunciation mistakes for Chinese learners of Japanese and solutions, *Japanese language education*, 53, 93-101.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied psycholinguistics*, 2(4), 369-390.

Maddieson, I., & Disner, S. F. (1984). *Patterns of sounds*. Cambridge: Cambridge university Press.

Masuda, H., & Arai, T. (2010). Processing of consonant clusters by Japanese native speakers: Influence of English learning backgrounds. *Acoustical Science and Technology*, 31(5), 320-327.

Masuda, H., & Arai, T. (2013). Identification of English voiceless fricatives in multispeaker babble noise by native Japanese and English listeners: Influence of English proficiency. *Acoustical Science and Technology*, 34(5), 356-360.

Masuda, H. (2016). Misperception patterns of American English consonants by Japanese listeners in reverberant and noisy environments. *Speech Communication*, 79, 74-87.

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of speech, language, and hearing research*, 40(3), 686-693.

McDonough, K., & Trofimovich, P. (2012). How to Use Psycholinguistic Methodologies for Comprehension and Production. *Research Methods in Second Language Acquisition: A Practical Guide*. 1st edited, 117-138.

Miller, G. A. (1947). The masking of speech. *Psychological bulletin*, 44(2), 105.

Mizutani, O., & Otsubo, K. (1971). What parts in speech are difficult for L2 learners of Japanese? : L1 Chinese learners. *Speech and Speech learning*, 189-193.

Nábělek, A. K., & Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native listeners. *The Journal of the Acoustical Society of America*, 75(2), 632-634.

Paradis, M., Hagiwara, H., & Hildebrandt, N. (1985). *Neurolinguistic Aspects of the Japanese Writing System*. Academic Press.

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International journal of audiology*, 47, 99-105.

RStudio Team (2018). *RStudio: Integrated Development for R. RStudio*, Boston, MA; Retrieved from http://www.rstudio.com/

Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-speaker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America*, 118(5), 2775-2778.

Smith, J. (2002). *Phonological augmentation in prominent positions*. University of Massachusetts. Ph.D. dissertation.

Sukegawa, Y. (1993). Analysis of pronunciation patterns produced by different first language Japanese learners based on the results of questionnaire, *Japanese language pronunciation and education: Studies of Japanese language education methods for international students, pronunciation of Japanese language D1 class, Research report in 1992*, 187-222.

Sugiyama, T. (1985). The pronunciation of Japanese: From the aspect of Chinese speech learning, *Japanese language Education*, 55, 97-110.

Takata, Y., & Nábělek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *The Journal of the Acoustical Society of America*, 88(2), 663-666.

Tsukada, K. (2011). The perception of Arabic and Japanese short and long vowels by native speakers of Arabic, Japanese, and Persian. *The Journal of the Acoustical Society of America*, 129(2), 989-998.

Tsurutani, C. (2004). Acquisition of *Yo-on* (Japanese contracted sounds) in L1 and L2 phonology. *Second Language*. 3. 27-47.

Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. *A Sound Approach to Language Matters: In Honor of Ocke-Schwen Bohn*, 607-630.

Vance, T. J. (2008). *The Sounds of Japanese*. Cambridge: Cambridge university press.

Van Dommelen, W. A., & Hazan, V. (2010). Perception of English consonants in noise by native and Norwegian listeners, S*peech Communication*, 52(11-12), 968-979.

Van Wijngaarden, S., Steeneken, H., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *The Journal of the Acoustical Society of America*, 111, 1906-1916.

Wang, N., Harada, T., Siriluck, D., & Min, K. (1987). D-1 Japanese phonetics, *NAFL Institute Japanese teacher training course*, ALC Press.

Watanabe, H. (2011). The pronunciation and instruction issues of Russian learners of Japanese, *Report of Japanese education*, 7, 71-84.

Wiese, R. (1997). Underspecification and the description of Chinese vowels. *Linguistic Models*, 20, 219-250.

Yu, Z. (1985). The similarities and differences between Japanese vowels and Chinese vowels. *Foreign Languages and Their Teaching*, 2 ,1-5.

Zhu, C. (2011). Why does "yu" sound like "yo" when produced by Chinese learners of Japanese? : Rethinking of Japanese vowel /u/. *Speech and Grammar,* 103-122, Tokyo: Kuroshio Publisher.