

Wei William Zhou*

The role of L1 phonology in the perception of L2 semivowels

<https://doi.org/10.1515/jjl-2024-2012>

Received March 3, 2024; accepted May 8, 2024

Abstract: The study compares how L1 Chinese and Japanese speakers perceive L2 English semivowels (/j/ and /w/) preceding high vowels (/i/ and /u/). In Chinese, [j] and [w] serve as phonetic variants of /i/ and /u/, respectively. However, /j/ and /w/ are distinct phonemes in Japanese, although Japanese /w/ lacks the roundness feature found in English /w/. Participants completed experiments with a discrimination task and an identification task. While the discrimination task revealed no differences in /j/ and /w/ perception between the groups, the identification task showed that the Japanese speakers outperformed the Chinese speakers in their perception of /j/, suggesting an L1 “phonemic over phonetic” advantage. However, the Japanese speakers did not outperform the Chinese speakers in their perception of /w/, suggesting that an L2 feature unexploited in the L1 (i.e., roundness) can impede perception. These findings underscore the importance of considering both phonemic status and features in L2 speech perception.

Keywords: Japanese; L2 phonology; Mandarin Chinese; speech perception

1 Introduction

It is widely acknowledged that the first language (L1) plays an important role in the acquisition of second language (L2) phonology (see Archibald 2021; Eckman 2004; Major 2008). L1 influence has also served as a crucial premise for L2 phonology acquisition theories. For example, some of the most influential L2 phonology models assume that the relative ease or difficulty of acquiring certain sounds in an L2 is dependent on whether and how the L2 sounds are mapped onto L1 phonetic categories, see the Perceptual Assimilation Model (PAM, Best 1995; Best and Tyler 2007), the Speech Learning Model (SLM, Flege 1995, 2005; Flege and Bohn 2021; Flege et al. 2021), and the Second Language Linguistic Perception (L2LP) Model (Escudero 2005, 2007; van Leussen and Escudero 2015). While these models primarily focus on

***Corresponding author: Wei William Zhou**, Department of East Asian Languages and Literatures, The Ohio State University, 398 Hagerty Hall, 1775 College Road, Columbus, OH 43210, USA, E-mail: zhou.2874@buckeyemail.osu.edu

segments and phonetic categories, other L2 phonology theories propose that features are the transferred elements during L2 phonology acquisition. These theories emphasize that the ease or difficulty of acquiring L2 sounds depends on how certain features associated with the L2 sounds are exploited in the learners' L1, see Redeployment Hypothesis (Archibald 2005) and Feature Geometry (Brown 2000; Brown and Matthews 1997). In addition to phonetic categories and features, other aspects of L1 phonological grammar, such as phonological rules and phonotactic constraints, can also influence the acquisition of L2 sounds (e.g., Dupoux et al. 1999; Hallé et al. 1998; Kilpatrick et al. 2019).

In a recent study, Zhou and Nakayama (2023) showed that the perception of English semivowels /j/ and /w/ by L1 Japanese L2 English learners cannot simply be attributed to mapping L1-L2 phonetic categories. Instead, phonotactic grammar and features both play an active role. To further shed light on how L1 phonological systems shape L2 phoneme perception, the current study builds upon Zhou and Nakayama's study and compares L1 Chinese and Japanese speakers' perception of English /j/ and /w/. While English, Japanese, and Mandarin Chinese all have the semivowels phonetically, the linguistic profiles of these sounds vary across the three languages. Therefore, examining this acquisition issue can provide fresh insights into the role of L1 phonology, involving phonetic categories, phonemic status, and featural composition in L2 speech perception.

2 Background

2.1 The role of L1 on L2 phonology acquisition

When adult learners acquire L2 sounds, their L1 phonological system can exert great influence, contributing to varying levels of ease or difficulty in acquisition. In L2 phonology models such as the PAM (Best 1995; Best and Tyler 2007) and the SLM (Flege 1995, 2005; Flege and Bohn 2021; Flege et al. 2021), the acquisition of given L2 sounds essentially depends on how these sounds are mapped onto the learners' L1 phonetic categories. In this regard, one straightforward criterion for predicting the relative ease or difficulty of L2 sounds is whether these sounds have corresponding equivalents in the learners' L1. When learners acquire two sounds in their L2 that constitute a phonemic contrast, but the L1 equivalents of these sounds do not form such a contrast in their L1, they may face challenges in acquiring these L2 sounds. For example, in English, /s/ and /ʃ/ are distinctive phonemes found in minimal pairs such as *sip* versus *ship*. However, /s/ and /ʃ/ are allophones of the same phoneme /s/ in Korean, and when L1 Korean speakers learn English, the /s/-/ʃ/ contrast in English can present a challenge (Eckman and Iverson 2013; Eckman et al. 2009; Fox et al. 2009;

Idsardi and Son 2004). The empirical evidence from this line of research aligns well with the fundamental tenets of models such as the PAM and the SLM.

In addition to the L1 transfer assumed at the segmental level, some L2 phonology theories have emphasized the role of features. In the framework of Chomsky and Halle (1968), speech sounds can be grouped according to their distinctive features, including cavity, manner of articulation, source, and prosodic features. Feature-based L2 phonology models largely propose that features (associated with L2 sounds) that are also exploited in the learners' L1 will be acquired more successfully compared to those that are not exploited in the L1 (e.g., Archibald 2005; Brown 1998, 2000). Brown (2000) conducted a series of experiments investigating the perception of English consonant contrasts among L1 Mandarin, Japanese, and Korean speakers of L2 English. The results showed that the L2 learners were more successful in distinguishing the consonant contrasts characterized by features present in their L1 phonological grammar, as opposed to those characterized by features absent in their L1. Research on L2 features has also dealt with suprasegmental features such as quantity. McAllister et al. (2002) examined the acquisition of Swedish vowel quantity by L2 learners of different L1 backgrounds. In Swedish, the same vowel category exhibits a long versus short length contrast, contributing to word meaning differences. Among the L1s spoken by the three groups of L2 learners, Estonian has a similar vowel quantity system while American English and Latin American Spanish do not. However, American English vowels exhibit a tense versus lax distinction (e.g., *beet* vs. *bit*), where tense vowels are typically longer in duration. Through production and perception tasks, the study found that the L1 Estonian speakers showed performance similar to the Swedish controls. While both the L1 English and Spanish speakers exhibited differences from the Swedish controls, the English speakers more closely resembled the native controls than did their Spanish counterparts. These results suggest that the presence (or absence) of a quantity feature in the L1 plays a role in the acquisition of L2 quantity. Similar findings have also been reported in studies by Pajak and Levy (2014), Meister et al. (2015), and Lee and Mok (2018), among others.

This study presents a case of L1 Mandarin Chinese and Japanese speakers acquiring L2 English semivowels, which involves both phonetic status and features.

2.2 Semivowels in English, Chinese, and Japanese

Semivowels are common in world languages (Maddieson 1984). In English, there are two semivowels, which are the palatal /j/ (e.g., yes) and labial-velar /w/ (e.g., was). These semivowels are highly resonant, similar to the vowels /i/ (e.g., tea) and /u/ (e.g., too), respectively. However, they distinguish themselves from vowels by

involving constriction and requiring less energy in their production (Lander and Carmell 1996). Besides acoustic differences, the functions of semivowels and vowels in English differ in that semivowels are used to release a vowel (or diphthong), and they are always placed next to the nucleus in consonant cluster contexts. Vowels, on the other hand, are only used as nuclei (Borden and Harris 1980). Moreover, in English, /j/ and /w/ can occur immediately before the vowels /i/ and /u/, such as in words *year* and *woozy*, and these words are lexically contrastive to words containing only /i/ and /u/, such as in *year* versus *ear*, and *woozy* and *oozy*.

In Japanese, both /j/ and /w/ exist as phonemes. While Japanese /j/ and English /j/ are highly similar, Japanese /w/ differs from English /w/ in an important articulatory feature – lip rounding. Specifically, Japanese /w/ is unrounded while its English counterpart is rounded (Labrune 2012; Okada 1991; Vance 1987, 2008). The two semivowels in Japanese also have their corresponding vowel counterparts /i/ and /u/. Acoustically, Japanese /i/ and /u/ resemble English /i/ and /u/, except that Japanese /u/ is unrounded whereas English /u/ is rounded (Nishi et al. 2008; Tsujimura 2014). Unlike English, Japanese /j/ and /w/ follow different phonotactic rules, where they cannot precede the vowels /i/ and /u/, respectively. In other words, the syllables */ji/ and */wu/ are not legal according to Japanese phonotactic grammar.

In Mandarin Chinese (hereafter referred to as Chinese), the semivowels [j] and [w] also exist. Phonetically, Chinese [j] is extremely similar to its English counterpart /j/. Chinese [w] is also similar to English /w/, with both being rounded. However, Chinese [j] and [w] are not phonemes but phonetic variants of [i] and [u] (Duanmu 2007; Lin 2001; Lin 2007). For example, the Chinese word *yī* ‘one’ can be pronounced as either [ji] or [i], and the word *wū* ‘room’ as [wu] or [u]. Although producing [j] and [w] is not required for words such as *yī* and *wū*, it is important to note that *The Standard Chinese Proficiency Test Guidelines* (Liu 2000), serving as the official guide for standard Chinese assessment, has prescribed that a weak homorganic on-glide should be pronounced before high vowels such as [i] and [u] in standard Chinese. In this regard, a recent study by Chan (2023) showed that native Chinese speakers produced [j] and [w] before the vowels [i] and [u] in their reading of individual words; however, the degree of [j] and [w] articulation varied both across words and speakers.

To summarize, Table 1 presents a crosslinguistic comparison of the two semivowels in the three languages.

2.3 L1 Japanese speakers’ perception of English semivowels

The acquisition of L2 English semivowels by L1 Japanese speakers presents an intricate case for evaluating the influence of the L1 on L2 phonology. Zhou and

Table 1: [j] and [w] in English, Japanese, and Chinese.

	Language	Phonemic status	Feature	Phonotactics
[j]	English	Phonemic /j/	Similar	/ji/
	Japanese	Phonemic /j/		*/ji/
	Chinese	Phonetic variation [ji] = [i]		[ji]
[w]	English	Phonemic /w/	Rounded	wu/
	Japanese	Phonemic /w/	Unrounded	*/wu/
	Chinese	Phonetic variation [wu] = [u]	Rounded	[wu]

Nakayama (2023) conducted a study on the perception of English semivowels by L1 Japanese, advanced L2 learners of English. The phonological contexts tested were /j/ and /w/ before high vowels /i/ and /u/, as well as /i/ and /u/ without the preceding semivowels, mirroring English lexical contrasts found in *yeast* versus *east* and *woof* and *oof*. Through an AX discrimination task and a two-alternative forced choice identification task, the study found that Japanese speakers had much trouble perceiving /j/ and /w/ before /i/ and /u/, respectively. Crosslinguistically, Japanese /j/ is acoustically similar to English /j/, but different from English in its distribution, as Japanese /j/ cannot occur with /i/ due to phonotactic constraints (*/ji/) in contemporary Japanese. In the context of the PAM, it would be assumed that English /j/ is assimilated into the Japanese /j/ category; that is, Japanese speakers would map the phonetics of English /j/ to the Japanese /j/ category. In such a case, Japanese speakers’ perception of English /j/ would be expected to be excellent, irrespective of the phonological context. However, the finding of Japanese speakers’ misperception of /j/ suggested that the L1 phonotactic constraint possibly inhibited their perception of /j/ before /i/. On the other hand, while /w/ is also similar between Japanese and English, Japanese /w/ differs from English /w/ not only in phonotactic rules (Japanese */wu/ vs. English /wu/) but also in the lip-rounding feature. The study found that Japanese speakers’ trouble with /w/ was due to their unfamiliarity with the roundness feature in English /w/, which outweighed the effects of the phonotactic constraint.

The crosslinguistic differences in semivowels in English, Japanese, and Chinese present a fascinating yet complex case for investigating the influence of L1 on the acquisition of L2 phonology. Therefore, the current study extends Zhou and Nakayama’s (2023) study and examines L1 Chinese speakers’ perception of L2 English semivowels, comparing these L2 data with Zhou and Nakayama’s L1 Japanese data. This research provides novel insights into how L1 phonological systems influence the perception of L2 phonemes.

3 Research question and predictions

The current study raises the following research question: How do L1 Chinese phonology and L1 Japanese phonology influence the L2 perception of English semivowels?

The study makes the following predictions regarding the L1 Chinese participants' performance compared to the L1 Japanese participants' performance in Zhou and Nakayama (2023). Overall, the Chinese participants' perception of /j/ and /w/ is expected to be less proficient compared to that of the Japanese participants, given that [j] and [w] occur only as phonetic variants of the vowels /i/ and /u/ in their L1. This phonetic realization may not help the Chinese participants to be sensitive to the phonemic contrast formed by /j/ and /w/ in English. However, it is expected that the gap between the Chinese and the Japanese speakers' perception of /w/ will be smaller due to the lip-rounding feature in English /w/ that is not utilized in Japanese. In other words, for /j/, the Chinese speakers are expected to be less accurate than the Japanese speakers. As for /w/, they are expected to either be less accurate than or on par with the Japanese speakers, because in this case, Chinese speakers face the challenge of the sound being a phonetic variant, while their Japanese counterparts grapple with the unfamiliar roundness feature.

4 Methods

4.1 Participants

Twenty L1 Chinese speakers (6 M and 14 F, average age = 22.05) participated in the study. At the time of the experiment, they were matriculated students in universities in the U.S. and were recruited through advisement. In Zhou and Nakayama (2023), the participants were 22 L1 Japanese speakers (6 M and 16 F, average age = 23.50) who were also degree-seeking students in U.S. universities, and 10 L1 English speakers (4 M, 4 F, and others 2, average age = 20.90). Table 2 shows a language background comparison between the Chinese participants in this study and the Japanese participants in Zhou and Nakayama's study. It includes measures in length of residence

Table 2: Language background comparison between the Chinese and Japanese participants.

Group	Length of residence	Age of arrival	Age of English learning onset	Daily English use
Japanese	2.53 (1.92)	21.68 (3.91)	9.45 (4.25)	72.50 (23.06)
Chinese	2.49 (1.75)	19.85 (3.41)	7.10 (3.43)	55.60 (17.02)

(the amount of time the participants spent living in the U.S., in years), age of arrival (the age at which the participants came to the U.S.), age of English learning onset (the age at which the participants began studying English), and daily English use (the frequency with which participants use English daily in the U.S., as a percentage). The standard deviations of these measures are in parentheses.

Four two-sample *t*-tests were performed to compare the four metrics. The results revealed a significant difference between the two groups in self-reported daily English use, $t(40) = 2.68, p < 0.05$, but no difference in length of residence, $t(40) = 0.07, p = 0.94$, age of arrival, $t(40) = 1.61, p = 0.11$, and age of English learning onset, $t(40) = 1.96, p = 0.06$. Given these results, along with the fact that these two groups of participants had similar English language preparation and readiness for studying in the U.S., the two groups of participants were regarded as experienced speakers of English with comparable functional proficiency.

4.2 Stimuli

The stimuli used were identical to those in Zhou and Nakayama (2023), which were 60 disyllabic English-like nonce words (e.g., critical: *eesha* /i:fə/, *yeesha* /ji:fə/, *oosha* /u:fə/, *woosha* /wu:fə/; filler: *leesha* /li:fə/). A male native speaker of American English produced five tokens of each word in a professional sound studio. Two out of the five tokens were chosen for the study, with their intensity being scaled to 68 dB. The words containing an initial semivowel were, on average, longer than those containing an initial vowel. This duration difference was not normalized, as it would be unnatural for /ji/ and /i/, as well as for /wu/ and /u/, to have the same duration due to the additional segment in /ji/ and /wu/. The average duration of the words containing /i/, /ji/, /u/, and /wu/ in milliseconds are listed in Table 3.

4.3 Procedure

The procedure was also identical to that described in Zhou and Nakayama (2023). The participants completed a consent form, a language background questionnaire, a

Table 3: Duration of the words (in milliseconds).

Word duration				Difference
/ji/	558	/i/	485	+73
/wu/	530	/u/	494	+36

headphone test, and a word familiarization session. During the familiarization session, the participants viewed the orthographic forms of the words and heard each word pronounced once. Then, they completed two perception tasks: an AX discrimination task (120 trials) and a two-alternative forced choice (2AFC) identification task (120 trials). There was a break between these two tasks. In the AX discrimination task, they were asked to judge whether the two words in each trial were the same word or different words (e.g., “same”: *eeshha-eeshha*; “different” *eeshha-yeeshha*). The two words within each trial were separated by a 200-millisecond interstimulus interval. After hearing each trial, they were given 3 s to respond. Each “same” trial always comprised two different tokens of the same word. In the 2AFC identification task, they were asked to listen to and identify each heard word from two provided options. After hearing each word, they were again given 3 s to respond. The experimental procedure lasted approximately 45 min, and all the participants received nominal fees for their participation.

5 Results

5.1 Accuracy in the AX discrimination task

Figure 1 displays the proportions of correct responses for the critical items among different language groups in the discrimination task. The L2 Japanese and L1 English

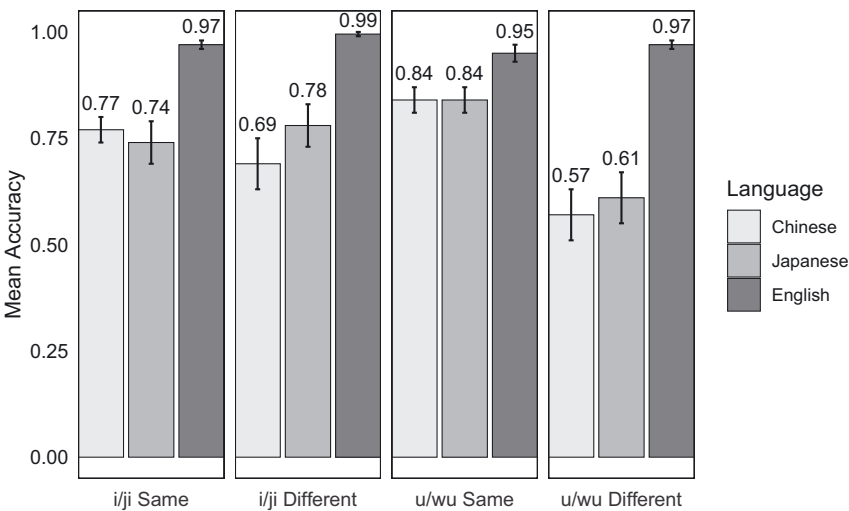


Figure 1: Mean accuracies and standard errors by language background, trial type, and word group in the discrimination task.

control data is from Zhou and Nakayama (2023). The proportions of correct responses for the filler items can be found in Appendix A.

As expected, the L1 English control group achieved high accuracies across the board, suggesting that English speakers took advantage of their L1 sensitivity to distinguishing these sounds, which form phonological contrasts in their language. On the other hand, both groups of L2 participants achieved lower accuracies across different conditions. Notably, both groups of participants appeared to exhibit a strong bias towards the “same” trial types in the “u/wu” word group, as indicated by the higher accuracy rates for the “same” than the “different” trial types. In other words, the participants were more accurate with /u/-/u/ and /wu/-/wu/ pairs than with /u/-/wu/ and /wu/-/u/ pairs. In the “i/ji” word group, the Chinese speakers again showed some “same” bias, whereas their Japanese counterparts did not exhibit such a tendency.

A mixed-design Analysis of Variance (ANOVA) was conducted for the Chinese and Japanese data, predicting response accuracy from three factors: language background (Chinese and Japanese), trial type (“same” and “different”), and word group (“i/ji” and “u/wu”). The model also incorporated an error term accounting for variability among participants within each trial type and word group condition. The model revealed a significant main effect of trial type, $F(1, 40) = 7.20$, $p < 0.05$, and a significant interaction effect between trial type and word group, $F(1, 40) = 21.65$, $p < 0.001$. The effect of language background was not significant, $F(1, 40) = 0.62$, $p = 0.43$. These results indicate that the effect of word group on response accuracy varied depending on the trial type. Post-hoc pairwise comparisons using Tukey’s Honest Significant Difference (HSD) test revealed that both Chinese and Japanese participants’ accuracy was significantly higher for the “same” trial type than for the “different” trial type in the “u/wu” word group (Chinese: Mean difference = 0.27, 95 % CI [0.05, 0.48], $p < 0.01$; Japanese: Mean difference = 0.23, 95 % CI [0.03, 0.44], $p < 0.05$). These results confirmed that both groups of participants exhibited a response bias towards the “same” trial types for the “u/wu” word group but not for the “i/ji” word group.

Based on the statistical analyses, it is evident that a strong response bias is present in the data. To address this and shed more light on the data, d' prime, a sensitivity measure that adjusts for response biases, was also computed for additional analysis in the next section.

5.2 Sensitivity (d') in the AX discrimination task

Rooting from the signal detection theory, d' prime (d') is a sensitivity or discriminability measure assessing one’s ability to distinguish between signal and noise. In the

current task, d' s were computed with the following procedure. First, the hit, miss, correct rejection, and false alarm rates of participants' discrimination responses were calculated. In the context of the current task, "hit" refers to correctly identifying a "different" trial as "different," "miss" refers to incorrectly identifying a "different" trial as "same," "correct rejection" refers to correctly identifying a "same" trial as "same," and "false alarm" refers to incorrectly identifying a "same" trial as "different." Then, the values for $Z(\text{hit}) - Z(\text{false alarm})$ were calculated, where Z is the inverse-normal transformation. To avoid undefined values caused by extreme probabilities from the inverse-normal transformation, a log-linear rule was applied (Hautus 1995). Finally, d' s were identified with the values of $Z(\text{hit}) - Z(\text{false alarm})$ according to the Independent Observation Model (Macmillan and Creelman 2005). By this calculation, a larger d' indicates more successful discrimination. Figure 2 displays the mean d' s and standard errors for the "i/ji" and the "u/wu" word group by the Chinese, Japanese, and English participants.

The d' s by the Chinese and Japanese participants underwent statistical analyses. A two-way, mixed-design ANOVA was conducted to examine the effects of language background (Chinese and Japanese) and word group ("i/ji" and "u/wu") on d' while accounting for within-subjects variability among participants. The model found no significant effects of language background, $F(1, 40) = 1.25$, $p = 0.27$, word group, $F(1, 40) = 0.32$, $p = 0.57$, or their interaction, $F(1, 40) = 0.16$, $p = 0.69$. Overall, these results suggest that Chinese and Japanese learners did not differ in their sensitivity to /j/ and /w/.

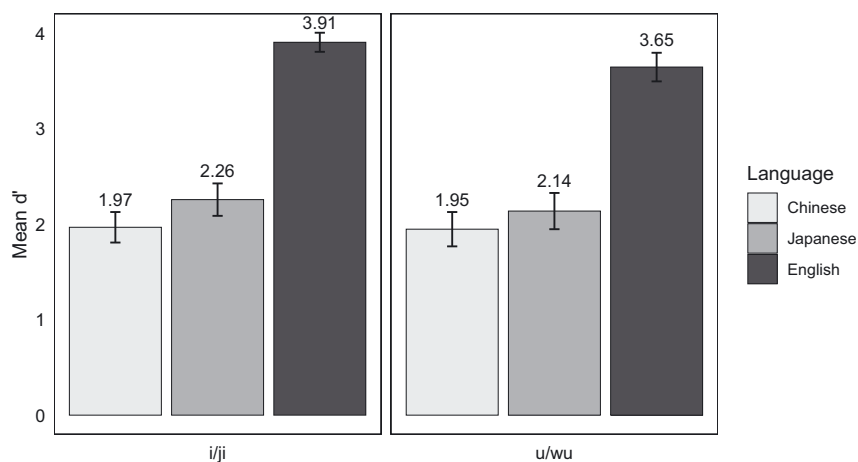


Figure 2: Mean d' s and standard errors by language background and word group in the discrimination task.

Based on the accuracy and d' analyses, it can be concluded that Chinese and Japanese participants did not differ in their discrimination performance. This finding will be discussed along with the results from the identification task in the discussion section.

5.3 Accuracy in the 2AFC identification task

Figure 3 shows the proportions of correct responses for the critical items by the three language groups in the identification task. The L2 Japanese and L1 English control data is from Zhou and Nakayama (2023). The results for the filler items can be found in Appendix B.

The response accuracy data was fit with a two-way, mixed-design ANOVA with language background (Chinese and Japanese) and word group (“i”, “ji”, “u”, and “wu”) as factors. The model also included an error term to account for participant variability within each word group. The model showed a significant main effect of language background, $F(1, 40) = 8.08, p < 0.01$, and a significant main effect of word group, $F(3, 120) = 5.31, p < 0.01$. However, there was no significant interaction between language background and word group, $F(3, 120) = 0.73, p = 0.53$. These results suggest that the Chinese and Japanese participants’ identification performance differed and that participants’ identification performance for the four types of words also differed. Post-hoc comparisons using Tukey’s HSD tests confirmed

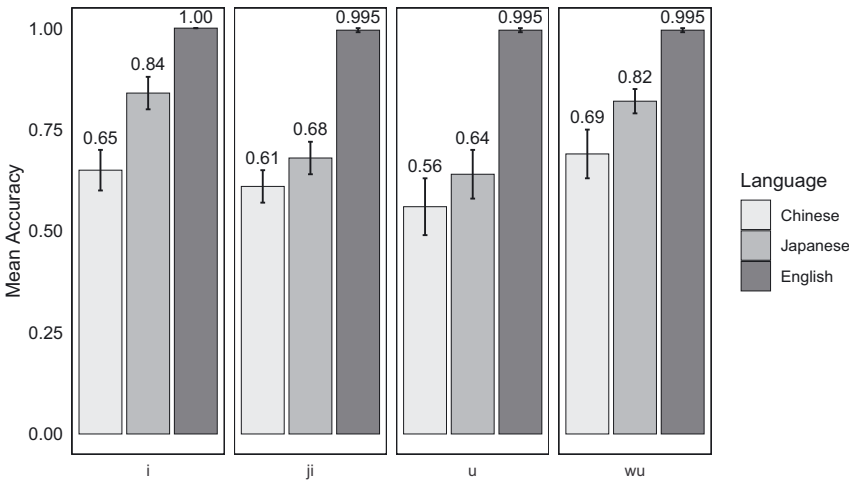


Figure 3: Mean accuracies and standard errors by language background and word group in the identification task.

that (a) in terms of language background, Japanese participants' overall identification performance was significantly better than their Chinese counterparts' (Mean difference = 0.12, 95 % CI [0.05, 0.19], $p < 0.01$); and (b) in terms of word type, participants' identification of /u/ was significantly less accurate than that of /i/ (Mean difference = -0.15, 95 % CI [-0.28, -0.02], $p < 0.05$), and their identification of /wu/ was significantly better than that of /u/ (Mean difference = 0.16, 95 % CI [0.03, 0.29], $p < 0.05$).

Overall, the statistical analyses showed that Japanese participants performed reliably better than the Chinese participants at the identification task. Moreover, both groups of participants had varied performance in identifying different word types. Notably, participants' ability to identify /wu/ was superior to that of /u/. Given the task's design (a two-alternative forced-choice task), the difference in performance between /wu/ and /u/ suggests a word bias. Specifically, the participants were more inclined to perceive a /w/ before /u/ when hearing /u/, but such an inclination was not observed for /j/ before /i/. To gain deeper insights into participants' perceptual sensitivity to the two semivowels, d' was computed for further analysis in the subsequent section.

5.4 Sensitivity (d') in the 2AFC identification task

Sensitivity d' was used to measure the participants' detectability of each semivowel. In the context of this task, "hit" refers to correctly identifying a semivowel before the vowel (detecting the presence of a semivowel), "miss" refers to misidentifying a semivowel as a vowel (failing to detect the presence of a semivowel), "correct rejection" refers to correctly identifying a vowel as a vowel (detecting the absence of a semivowel), and "false alarm" refers to misidentifying a vowel as a semivowel (failing to detect the absence of a semivowel). Figure 4 shows the mean d' s and standard errors by word group and language background in the identification task.

Statistical analyses were performed on the d' s for the Chinese and Japanese learners. A two-way, mixed-effects ANOVA was conducted to assess the effects of language background (Chinese and Japanese) and word group ("i/ji" and "u/wu") on d' . The model incorporated an error term for participants within each word group. The model revealed a significant main effect of language background, $F(1, 40) = 7.30$, $p < 0.05$, suggesting that the perceptual sensitivity of the two groups of participants differed. Neither the main effect of word group, $F(1, 40) = 0.71$, $p = 0.41$, nor the interaction between language background and word group, $F(1, 40) = 2.18$, $p = 0.15$, reached statistical significance. Post-hoc pairwise comparisons using Tukey's HSD tests revealed that the Japanese participants were significantly better than their Chinese peers in perceiving /j/ in the "i/ji" word group (Mean difference = 1.00, 95 % CI

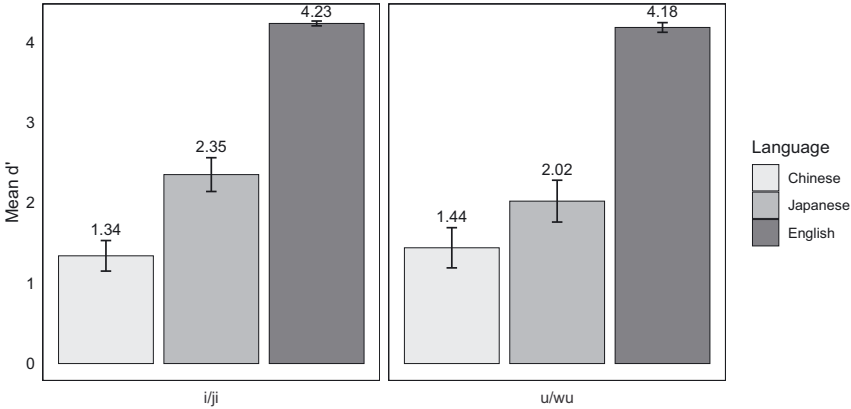


Figure 4: Mean d's and standard errors by language background and word group in the identification task.

[0.15, 1.86], $p < 0.05$); however, the two groups of participants did not differ in perceiving /w/ before /u/ (Mean difference = 0.58, 95 % CI [-0.28, 1.44], $p = 0.29$).

In summary, the accuracy and d' analyses suggest that overall, the Japanese participants were better at perceiving the semivowels than the Chinese participants, and their advantage primarily lies in the semivowel /j/.

6 Discussion

In the current study, L1 Chinese and Japanese speakers' perception of L2 English semivowels /j/ and /w/ were examined and compared through an AX discrimination task and a 2AFC identification task. The discrimination task showed that the Chinese and the Japanese speakers did not differ in their ability to discriminate between /i/ and /ji/, or between /u/ and /wu/. However, the identification task showed that the Japanese speakers outperformed their Chinese counterparts in perceiving /j/, but not /w/. These results provide some direct evidence for our research question, in that L1 Chinese phonology and L1 Japanese phonology influence the perception of L1 English semivowels differently. The phonemic status of /j/ in Japanese facilitated L1 Japanese speakers' perception of English /j/, allowing them to outperform their L1 Chinese peers, for whom [j] only exists as a phonetic variant in their L1. However, Japanese speakers were unable to maintain such a phonemic advantage over their Chinese peers in perceiving English /w/, which possesses a roundness feature not found in their L1 equivalent. These results suggest that both phonemic status and features play a role in shaping L2 speech perception.

Accuracy analysis, coupled with d' analysis, proved to be important in understanding the data in the current study. In the discrimination task, it was found that participants' accuracy was higher for the "same" trials than for the "different" trials in the "u/wu" word group. Considering the task design, it was suspected that this higher accuracy did not correspond to better ability; instead, it suggested a response bias towards the "same" trials. The subsequent d' analysis confirmed this response bias by showing no statistical difference in participant sensitivity between the two word groups. In this task, both the accuracy and d' analyses found no effect of language background, supporting the conclusion that the two groups of participants did not differ from each other in their discrimination. However, in the identification task, the Japanese participants were clearly more capable than the Chinese participants of perceiving /j/. This discrepancy gives rise to the possibility of some task-related artifact. In the discrimination task, there was a noticeable duration difference between the /i/ and the /ji/ words, where the /ji/ words were 73 ms longer than the /i/ words on average. Therefore, it is conceivable that both groups of participants could have used this durational cue for /i/-/ji/ differentiation. The rationale for not reducing the duration of /ji/ words in the stimulus preparation stage was to maintain the naturalness of speech. Specifically, since /ji/ contains one more segment than /i/, it is reasonable for /ji/ to be longer than /i/, just as one would expect the word *year* to have a longer duration than the word *ear* in English. While duration is not a primary cue for differentiating between words such as *year* and *ear*, the durational cue should still be made available for listeners to use. If duration was indeed exploited by the participants in this task, especially by the Chinese participants, to reach a comparable level of performance to their Japanese peers, this would have pedagogical implications for L2 learning. Specifically, L2 instructors and learners could use the durational cue as a starting point while gradually adjusting cue weighting (i.e., spectral cues) to improve semi-vowel processing. This perceptual refinement strategy could be effective in training L2 learners with particularly difficult sounds.

Based on the accuracy data from the identification task, Chinese participants' accuracy rates for the /i/ and the /ji/ words were similar (0.65 vs. 0.61), while the Japanese participants' accuracy for the /i/ words was much higher than for /ji/ words (0.84 vs. 0.68). Given that /i/ exhibits a high acoustic similarity between Japanese and English, the Japanese could have taken advantage of their L1 phonological sensitivity to perceive English /i/, leading to good accuracy. However, even though /j/ is also acoustically similar between Japanese and English, the phonotactic constraint of */ji/ in Japanese could have made /ji/ a less favored candidate in this context. This suggests that L1 phonotactic grammar can shape L2 speech perception (e.g., Dupoux et al. 1999; Hallé et al. 1998; Kilpatrick et al. 2019). On the other hand, although Japanese phonotactic grammar permits /u/ but prohibits */wu/, the Japanese participants much preferred /wu/ over /u/. One explanation, as suggested in Zhou and Nakayama (2023),

to account for this difference is that since the Japanese participants were not accustomed to the roundness feature in English /w/ and /u/, the co-occurrence of two rounded segments, i.e., /wu/, rendered the phonotactic violation unrecognizable to them. Consequently, any detection of roundness, whether it was /u/ or /wu/, would make them more inclined to overgeneralize it to /wu/.

The d' data from the identification task aligned well with the predictions. Crucially, the Chinese participants did not differ in their ability to perceive /j/ and /w/, possibly due to the two sounds both being phonetic variants in their L1. The Japanese participants also showed no difference in their ability to perceive /j/ and /w/, suggesting the possibility of English /j/ and /w/ being assimilated into their L1 phonetic categories (e.g., Best 1995). The comparison between the language groups yielded some interesting results. In the case of /j/, the Japanese participants' performance was reliably better than their Chinese peers, demonstrating a "phonemic over phonetic" advantage. However, in the case of /w/, the Japanese participants did not outperform the Chinese participants. This lack of significant cross-language differences in perceiving /w/ can result from an "even-out" effect.¹ In Japanese, /w/ is phonemic but differs from English by features, while in Chinese [w] is only a phonetic variant but is similar to English [w] in feature. These perceptual advantages and disadvantages across languages result in a smaller performance difference for /w/ compared to the /j/ case, where Japanese speakers benefit from both phonemic and featural advantages. These findings imply that phonemic status and features together shape speech perception. Building upon these findings, more research is still needed to understand how phonemic status and features mediate one another in L2 phonology acquisition.

7 Conclusions

In conclusion, this study examines the perception of L2 English semivowels (/j/ and /w/) preceding high vowels (/i/ and /u/) among L1 Chinese speakers in comparison to L1 Japanese speakers. While Chinese phonology treats [j] and [w] as phonetic variants of /i/ and /u/, respectively, Japanese has /j/ and /w/ as distinct phonemes. However, Japanese /w/ lacks the roundness feature present in English /w/. Through AX discrimination and 2AFC identification tasks, the study found comparable performance by the Chinese and the Japanese participants in /i/-/ji/ discrimination and /u/-/wu/ discrimination. However, the Japanese participants showed better identification of /j/ compared to the Chinese participants, suggesting an L1 phonemic advantage. Notably, the Japanese participants did not outperform the Chinese

1 I thank an anonymous reviewer for raising this point.

participants in their perception of /w/, indicating that the unexploited L2 feature of roundness in the L1 poses some challenges for them. Overall, these findings highlight the importance of phonemic status and features in understanding L2 speech perception.

Acknowledgments: I thank Mineharu Nakayama, Masahiko Minami, Keiko Yukawa, and two *JJL* anonymous reviewers for their valuable feedback. I also extend my gratitude to Yuichi Ono, Atsushi Fujimori, Derek Reiman, and Yuki Hattori for their insightful discussions.

Research funding: The research received support from the Nissen Chemitec America Scholarship, administered by the Department of East Asian Languages and Literatures at The Ohio State University.

**Appendix A: Proportions of correct responses
(standard errors) for the filler items
by language background in the
discrimination task**

Language	li		lu	
	Same	Different (e.g., li-i)	Same	Different (e.g., lu-u)
Chinese	0.89 (0.03)	1.00 (0.00)	0.89 (0.03)	0.98 (0.01)
Japanese	0.90 (0.03)	0.98 (0.01)	0.91 (0.03)	0.99 (0.01)
English	0.98 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.00)

**Appendix B: Proportions of correct responses
(standard errors) for the filler items
by language background in the
identification task**

Language	li	lu
Chinese	0.995 (0.003)	0.99 (0.005)
Japanese	0.98 (0.01)	0.98 (0.01)
English	1.00 (0.00)	1.00 (0.00)

References

- Archibald, John. 2005. Second language phonology as redeployment of L1 phonological knowledge. *The Canadian Journal of Linguistics* 50(1). 285–314.
- Archibald, John. 2021. Ease and difficulty in L2 phonology: A mini-review. *Frontiers in Communication* 6. 1–7.
- Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 171–204. Timonium, MD: York Press.
- Best, Catherine T. & Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Ocke-Schwen Bohn & Murray J. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, 13–34. Amsterdam, The Netherlands: John Benjamins.
- Borden, Gloria J. & Katherine S. Harris. 1980. *Speech science primer: Physiology, acoustics, and perception of speech*. Baltimore, MD: Williams & Wilkins.
- Brown, Cynthia. 1998. The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14(2). 136–193.
- Brown, Cynthia. 2000. The interrelation between speech perception and phonological acquisition from infant to adult. In John Archibald (ed.), *Second language acquisition and linguistic theory*, 4–63. Malden, MA: Blackwell.
- Brown, Cynthia & John Matthews. 1997. The role of feature geometry in the development of phonemic contrasts. In S. J. Hannahs & Martha Young-Scholten (eds.), *Focus on phonological acquisition*, 67–112. Amsterdam, The Netherlands: John Benjamins.
- Chan, Marjorie K. M. 2023. The zero initial in Chinese: A preliminary exploration into D2 and L2 acquisition. In Mineharu Nakayama, Marjorie K. M. Chan & Zhiguo Xie (eds.), *Buckeye East Asian linguistics*, vol. 6, 1–14. Columbus, OH: The Ohio State University.
- Chomsky, Noam & Halle Morris. 1968. *The sound pattern of English*. New York, NY: Harper & Row.
- Duanmu, San. 2007. *The phonology of standard Chinese*, 2nd edn. Oxford, UK: Oxford University Press.
- Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25(6). 1568–1578.
- Eckman, Fred R. 2004. From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition* 26(4). 513–549.
- Eckman, Fred R. & Gregory K. Iverson. 2013. The role of native language phonology in the production of L2 contrasts. *Studies in Second Language Acquisition* 35(1). 67–92.
- Eckman, Fred R., Gregory K. Iverson, Robert A. Fox, Ewa Jacewicz & Soyoung Lee. 2009. Perception and production in the acquisition of L2 phonemic contrasts. In Michael Alan Watkins, Andreia S. Rauber & Barbara O. Baptista (eds.), *Recent research in second language phonetics/phonology: Perception and production*, 81–96. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.
- Escudero, Paola. 2005. *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Utrecht, The Netherlands: LOT.
- Escudero, Paola. 2007. Second-language phonology: The role of perception. In Martha C. Pennington (ed.), *Phonology in context*, 109–134. London, UK: Palgrave Macmillan.
- Flege, James E. 1995. Second language speech learning theory, findings, and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 233–277. Timonium, MD: York Press.

- Flege, James E. 2005. Origins and development of the speech learning model. In *Keynote lecture presented at the 1st ASA workshop on L2 speech learning*. Vancouver, Canada: Simon Fraser University.
- Flege, James E., Katsura Aoyama & Ocke-Schwen Bohn. 2021. The revised speech learning model (SLM-r) applied. In Ratree Wayland (ed.), *Second language speech learning*, 84–118. Cambridge, UK: Cambridge University Press.
- Flege, James E. & Ocke-Schwen Bohn. 2021. The revised speech learning model (SLM-r). In Ratree Wayland (ed.), *Second language speech learning*, 3–83. Cambridge, UK: Cambridge University Press.
- Fox, Robert A., Jacewicz Ewa, Fred R. Eckman, Gregory K. Iverson & Soyoung Lee. 2009. Perception versus production in Korean L2 acquisition of English sibilant fricatives. In The Linguistic Society of Korea (ed.), *Current issues in unity and diversity of languages*, 2662–2680. Seoul, Korea: The Linguistic Society of Korea.
- Hallé, Pierre A., Juan Segui, Uli Frauenfelder & Christine Meunier. 1998. Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance* 24(2). 592–608.
- Hautus, Michael J. 1995. Corrections for extreme proportions and their biasing effects on estimated values of d. *Behavior Research Methods, Instruments, & Computers* 27(1). 46–51.
- Idsardi, William J. & Ah Son Sun. 2004. What do Korean speakers learn about English fricatives? In Susumu Kuno, Ik-Hwan Lee, John Whitman, Joan Maling, Young-Se Kang & Young-joo Kim (eds.), *Harvard studies in Korean linguistics X*, 48–59. Cambridge, MA: Department of Linguistics, Harvard University.
- Kilpatrick, Alexander J., Rikke L. Bundgaard-Nielsen & Brett J. Baker. 2019. Japanese co-occurrence restrictions influence second language perception. *Applied Psycholinguistics* 40(2). 585–611.
- Labrune, Laurence. 2012. *The phonology of Japanese*. New York, NY: Oxford University Press.
- Lander, Terri & Tim Carmell. 1996. Structure of spoken language: Spectrogram reading. <https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/speech.bme.ggi.edu/tutordemos/SpectrogramReading/cse551html/cse551/cse551.html>.
- Lee, Albert & Peggy Mok. 2018. Acquisition of Japanese quantity contrasts by L1 Cantonese speakers. *Second Language Research* 34(4). 419–448.
- Lin, Hua. 2001. *A grammar of Mandarin Chinese*. Munich, Germany: LINCOM EUROPA.
- Lin, Yen-Hwei. 2007. *The sounds of Chinese*. Cambridge, UK: Cambridge University Press.
- Liu, Zhaoxiong. 2000. *Putonghua shuiping ceshi dagang [Standard Mandarin proficiency test guidelines]*, 7th edn. Jilin, China: Jilin Renmin Press.
- Macmillan, Neil A. & C. Douglas Creelman. 2005. *Detection theory: A user's guide*, 2nd edn. Mahwah, NJ: Lawrence Erlbaum Associates.
- Maddieson, Ian. 1984. *Patterns of sounds*. New York, NY: Cambridge University Press.
- Major, Roy C. 2008. Transfer in second language phonology: A review. In Jette G. Hansen Edwards & Mary L. Zampini (eds.), *Phonology and second language acquisition*, vol. 36, 63–94. Amsterdam, The Netherlands: John Benjamins.
- McAllister, Robert, James E. Flege & Piske Thorsten. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30(2). 229–258.
- Meister, Einar, Nemoto Rena & Meister Lya. 2015. Production of Estonian quantity contrasts by Japanese speakers. *Journal of Estonian and Finno-Ugric Linguistics* 6(3). 79–96.
- Nishi, Kanae, Strange Winifred, Reiko Akahane-Yamada, Rieko Kubo & Sonja A. Trent-Brown. 2008. Acoustic and perceptual similarity of Japanese and American English vowels. *Journal of the Acoustical Society of America* 124(1). 576–588.
- Okada, Hideo. 1991. Japanese. *Journal of the International Phonetic Association* 21(2). 94–96.

- Pajak, Bozena & Roger Levy. 2014. The role of abstraction in non-native speech perception. *Journal of Phonetics* 46. 147–160.
- Tsujimura, Natsuko. 2014. *An introduction to Japanese linguistics*, 3rd edn. West Sussex, UK: John Wiley & Sons.
- van Leussen, Jan-Willem & Paola Escudero. 2015. Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology* 6. 1–12.
- Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.
- Vance, Timothy J. 2008. *The sounds of Japanese*. New York, NY: Cambridge University Press.
- Zhou, Wei William & Mineharu Nakayama. 2023. Perception of English semivowels by Japanese L2 English listeners. In Yuichi Ono & Masaharu Shimada (eds.), *Data science in collaboration*, vol. 6, 48–57. Tsukuba, Japan: Media JOHO., Ltd.