Pronunciation of Novel Phonemes in Mandarin Chinese for L2 Speakers

Rebecca Gips

Abstract

Taking on any second language (L2) can be a challenge as learners are faced with the task of learning new vocabulary, structures, and even sounds. Pronunciation specifically can cause issues for learners when they encounter sounds (i.e., phonemes) that do not exist in their first language(s) (L1). The Perceptual Assimilation Model states that speakers will assimilate unfamiliar phonemes to those they are more familiar with in their L1, and the Revised Speech Learning Model states that speakers have the ability to create new phonetic categories for unfamiliar phonemes in an L2, just as they do when learning an L1 in childhood. For this study, L1 speakers of English learning L2 Mandarin Chinese were asked to pronounce phonemes that do not occur in English to determine whether these learners replace these phonemes with ones that occur in English, or if they accommodate their production to the standard Mandarin pronunciation by adding a new phonetic category. By analyzing formant data from intermediate-level Mandarin Chinese speakers' speech, multiple compensation strategies were displayed, employing tactics from multiple documented models of phonological acquisition.

DOI: 10.5195/pur.2025.104

1: Introduction

Research shows that L1 English learners of L2 Mandarin face many struggles during the acquisition process. One study showed that "learners have difficulty with the pronunciation of … consonants, vowels, and tones in Mandarin Chinese" (Jiang and Cohen 2018, 25). The current study investigates a specific difficulty in learning Mandarin: lack of phoneme overlap. Multiple models of acquisition are posited for second language learning; one advocates that new phonetic categories can be created for L2 learners, whereas another posits that sounds in the L2 that do not exist in the learner's L1 will be replaced with sounds that do. When there exists a phoneme that the learner does not have in their L1, only one of the two models will be used. Although reasons why one model would be used over the other will not be explored in this paper, it is possible to use speech analysis evidence to support the usage of one model over the other. I chose to analyze the high front tense rounded [y] phoneme, which exists in Chinese but not in English, to see if learners accommodate their speech to include this phoneme or assimilate it to an articulatorily similar one that they already know how to pronounce.

Section 3 of this paper first explores the described models of phonological acquisition in more detail. Then, the phonological inventory of Chinese is explained for readers to gain a clear sense of the phonemes that are analyzed in this paper. The next part explains why one phoneme might be chosen over another in the case that the learners don't create a new phonetic category for the [y] sound. In the final part of the background, ideal spectrograms are given, both in visual and quantified forms, to use as a basis for comparison for the ideal versus actually realized speech of learners. Following the background, section 4 poses the research question and hypothesis regarding assimilation to a different inventory versus replacement using sounds in L1 inventory. Following that, the methodology and participant description is given in section 5. Section 6 presents the data, represented as formant measurements in a chart, and section 7 explains the findings. Section 8 briefly discusses the limitations of the study, and results are summarized in section 9.

2: Background

2.1: Models of Phonological Acquisition

In discussing a speaker's ability to perceive and produce unfamiliar phonemes, it is necessary to discuss theories of phonological acquisition. There are two main theories of L2 phonological acquisition that I will discuss in this paper. First is the Perceptual Assimilation Model, or PAM. PAM "posits that, when listening to an unfamiliar nonnative phone (phonetic segment), naïve listeners are likely, due to their native language experience, to perceptually assimilate the nonnative phone to the most articulatorily-similar native phoneme" (Best and Tyler 2007, p. 20). This means that if there are phonemes that do not exist in the L1, the L2 learner will try to choose the phoneme that sounds or is produced most similarly to what exists in the L1. When vowels "are perceived to differ notably from any L1 vowels", larger perceptual learning differences occur than when vowels are perceived to be similar or identical to L1 vowels (Best and Tyler 2007, p. 22).

The second model of phonological acquisition is called the Revised Speech Learning Model, or the SLM-r. It is based on the Speech Learning Model, or SLM, which posits that learning to perceive phonemes in a speaker's L2 uses the same mechanisms as are used when acquiring an L1 in childhood. These mechanisms include the ability to create new phonetic categories for the L2 which the speaker is unfamiliar with (Flege and Bohn 2021, p. 11). The SLM-r, in addition to supporting the ideas of the SLM, accounts for age-based differences in language learning and how stage in life has an impact on this process. This updated model posits that non-native production and perception are different from native not because of cognitive decline later in life, but instead because different learning methods are used for acquiring an L2 than are used for acquiring L1 in childhood. In fact, differences in outcomes for learning come from substituting L1 sounds in the L2, preexisting phonetic categories interfering with new ones, and differences in input for L2 learners versus monolingual native speakers (Flege and Bohn 2021, p. 23). Therefore, the SLM-r posits that there are no cognitive differences in learning an L1 versus an L2, and that phonetic categories that only exist in the L2 can be acquired, although maybe not by the same means that they were acquired for the L1. Therefore, learners should be able to perceive and produce phonemes which do not exist in their L1.

It is well documented that a person's L1 has an impact on their acquisition of an L2. Learners of a second language will try to apply what they know about language in general to their target language, even if the principles do not apply. This concept is called *transfer*, which impacts production and perception in an L2. Speakers tend to formulate in their native language, and then translate into their target language, which can result in non-native or non-standard sounding speech. This is called *negative transfer*, where the L1 negatively impacts the production of the L2 (Cheng 2023, p. 1236-1237). Specifically, L1 can provide challenges for pronunciation in a second language. There also exists a hierarchy of difficulty for acquiring phonemes in the L2. This is because second language learners might default to habits from their L1, especially for sounds that do not exist in their L1 (Liu 2011). A mechanism such as this one would be in accordance with PAM because speakers produce phonemes that are similar to those in their L1, rather than assimilating to the phonological inventory of the L2.

Therefore, speakers can either assimilate their speech to the L2 inventory, according to the SLM-r model, or use phonemes from their native language in place of unfamiliar L2 phonemes, in accordance with PAM. Creating new phonetic categories is a process in line with the SLM-r model, while speakers using PAM use existing phonetic categories. Because L1 can have an impact on the acquisition of an L2, predictions that PAM will be used might be borne due to the difficulty to assimilate to a new phonological inventory.

2.2: Phonological Inventory of Mandarin Chinese

2.2.1: Consonants

Mandarin has 22 consonants (see Table 1). In places where English might have a voiced/unvoiced pair for a certain place and manner of articulation, Mandarin has an aspirated/unaspirated pair. This pair occurs for bilabial plosives (p, p^h), alveolar plosives (t, t^h), and velar plosives (k, k^h). There are three additional pairs of aspirated/unaspirated phonemes, which are affricates and do not occur in English: alveolar (ts, ts^h), alveopalatal (tɕ, $tɕ^h$), and retroflex (tʂ, $tʂ^h$). Additionally, there are three fricatives in

Mandarin that do not occur in English: the voiceless alveopalatal [ɕ], voiceless retroflex [ʂ], and voiceless velar [x]. The remaining consonants, voiceless bilabial stop [p], voiceless alveolar stop [t], voiceless velar stop [k], bilabial nasal stop [m], alveolar nasal stop [n], velar nasal stop [ŋ], voiceless labiodental fricative [f], voiceless alveolar fricative [s], voiced retroflex liquid [ɹ], and voiced lateral liquid [l], all co-occur in English and Mandarin. Furthermore, it should be noted that the only word-final consonants in Mandarin are [n] and [ŋ] (American Speech-Language Hearing Association, n.d.).

	Bilabial		**Labio dental**		**Alveolar**		**Alveo-palatal**		**Retroflex**		**Velar**	
Plosive	p	p^h			t	t^h					k	k^h
Nasal		m				n						ŋ
Fricative			f		s		ɕ		ʂ		x	
Affricate					t͡s	t͡s^h	t͡ɕ	t͡ɕ^h	t͡ʂ	t͡ʂ^h		
Liquids									ɹ			
Glides						l						

Table 1: Consonant Inventory in Mandarin

2.2.2: Vowels

Mandarin has 6 vowels, 4 of which also occur in English. The two vowels that do not occur in English are the high front tense rounded [y] and close-mid back tense unrounded [ɣ] (Lee and Zee 2003). In addition, there are eleven diphthongs and four triphthongs, but only the monophthongs are found in Figure 1.

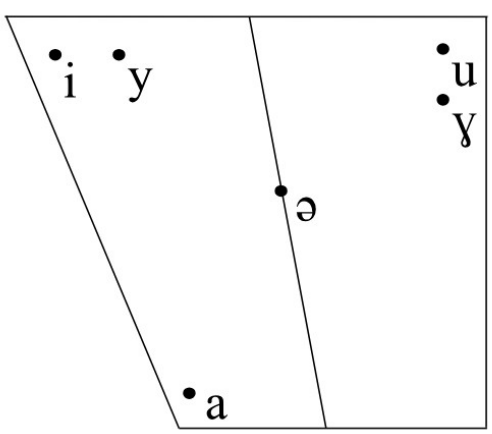


Figure 1: Vowel Inventory in Mandarin

2.3: Phoneme Replacement Choice

The Perceptual Assimilation Model predicts that speakers will use a phoneme from their L1 in the case where they do not recognize an L2 phoneme. In research done on native English speakers learning Quebecois and European French, it was found that they produce the high front tense unrounded [i], high front tense rounded [y], and high back tense rounded [u] “lower in the vowel space (and perhaps more like English vowels) than the native French speakers did”, and therefore did not produce “any of the French vowels at the vowel height used by the native French speakers” (Baker-Smemoe and Smith 2010, p. 731). Although they do not necessarily pronounce the [u] phoneme in place of [y], this finding still demonstrates that language learners have a tendency to produce vowels in a way that is similar to their native pronunciation. Therefore, because English speakers use phonemes most similar to those found in English, this supports the claim of the PAM. But how do speakers choose which phoneme in their L1 to serve as the replacement? First, as previously stated, PAM posits that speakers “assimilate the nonnative phone to the most articulatorily-similar native phoneme” (Best and Tyler 2007, p. 22). This means that for vowel replacement specifically, the replacement phoneme will differ in as few features as possible (backness, height, roundedness, or tenseness of the vowel).

Furthermore, orthography can have an impact on which phoneme is used as the replacement. In two studies, Basetti explored the impact of Pinyin, the romanized transliteration of Chinese, on the acquisition of Mandarin phonology for L1 English speakers. Her findings showed that speakers were more likely to pronounce phonemes when they appeared in the Pinyin forms. For example, the Chinese characters /uei/ and /guei/ are orthographically transcribed in Pinyin as <wei> and <gui>, respectively, but speakers were more likely to pronounce the /e/ in <wei> because of that written form. Therefore, “Pinyin spelling conventions affected learners’ phonological representations of Mandarin” (Bassetti 2006, Bassetti 2007, as cited in Hayes-Harb and Barrios 2021, p. 299). This finding can further help to predict which phoneme speakers may choose as a replacement. When multiple phones are equally articulatorily similar, they might choose the one that has similar representations in Pinyin and English due to their similar orthography. Therefore, because /ü/ has similar orthography to /u/, speakers might be more inclined to say [u], rather than [i], even though they both only differ in one feature from [y].

Therefore, there are multiple reasons why speakers might choose one phoneme over another if they do not create a new phonetic category in accordance with PAM. Most basically, phoneme replacement outcome is affected by what phonemes are articulatorily similar. Speakers will not choose a vowel that differs in many features. Secondly, the way in which a certain phoneme is transcribed will also impact phoneme replacement outcomes because speakers tend to use the orthography for determining the correct pronunciation. These two reasons can both impact which phoneme is used as replacement, but the orthography reason has only been discussed for Mandarin learners, not for French learners.

2.4: Idealized Spectrograms

In order to make a proper comparison between the expected production and the speech actually produced, it is necessary to understand what an ideal spectrogram of the sound in question looks like. The ideal [i], [y], and [u] are shown in the spectrogram in Figure 2.

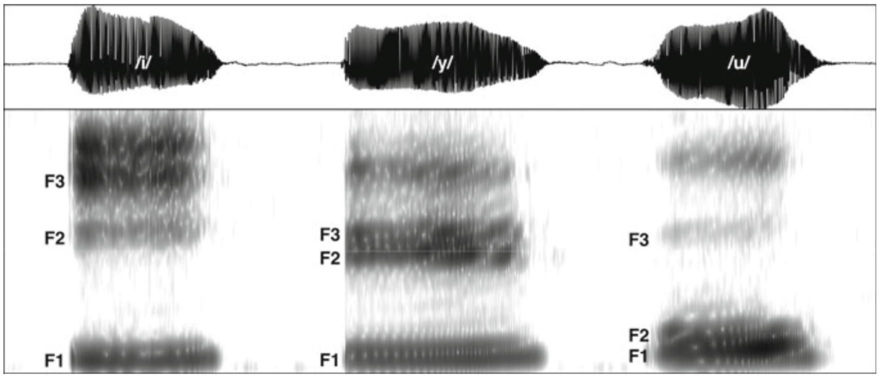


Figure 2: Waveforms and spectrograms of the three high vowels, yí /i/ 'aunt', yú /y/ 'fish', and wú /u/ 'nil'. The first three formants are labeled accordingly (Reproduced from Fon 2020)

According to this figure, the high front tense unrounded [i], high front tense rounded [y], and high back tense rounded [u] all have different formant structures. Because the distance between F1 and F2 is inversely proportional to vowel backness, it makes sense that [u] would have extremely close F1 and F2 formants and that [y] would have distant F1 and F2.

To conduct a proper comparison between the idealized [y] and samples for analysis, I recorded my own samples to collect formant data and HZ measurements. The idealized [y] sound is shown in the spectrogram in Figure 3.

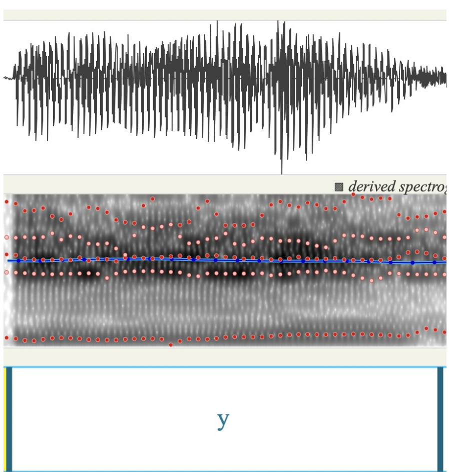


Figure 3: Waveform and spectrogram of the high front tense rounded [y] vowel

This formant structure matches the idealized version from Fon (2020) in shape and distance between F1, F2, and F3. Further, I have measured the averages of each formant. These measurements also mirror the ideal [y] from Fon (2020) in that they show a great distance between F1 and F2, with F2 and F3 being relatively closer to each other. These results have been quantified in the table below, and illustrate that for this phoneme, the distance between F1 and F2 should be around 2000 HZ and the distance between F2 and F3 should be around 500 HZ.

The same steps have been taken with the [u] phoneme and are shown in Figure 4.

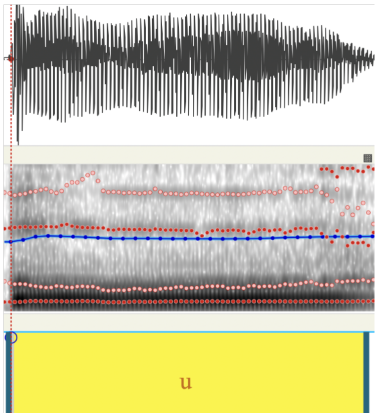


Figure 4: Waveform and spectrogram of the high back tense rounded [u] vowel

This spectrogram also matches the idealized one from Fon (2020), in that F1 and F2 are relatively close, while F2 and F3 are farther apart. These results have been quantified in the table below, and illustrate that for this phoneme, the distance between F1 and F2 should be around 500 HZ and the distance between F2 and F3 should be around 2000 HZ. Refer to Table 2. Note that the idealized recordings come from a 21-year-old female speaker.

	F1	F2	F3
Ideal [y]	401	2409	2913
Ideal [u]	402	896	2790

Table 2: Mean HZ measurements for each formant for ideal high front tense rounded [y] vowel and high back tense rounded [u] vowel

3: Research Question

The topic of investigation for this study concerns a comparison between L2 speakers and idealized phonemes of Mandarin Chinese. The goal of the analysis is to determine whether having English as a native language has an impact on pronunciation of Chinese phonemes that do not occur in English. Specifically, "pronouncing /ü/ appeared to be … problematic", although this difficulty occurred on a continuum and became less prominent the more advanced the speaker (Jiang and Cohen 2018, p. 29). Therefore, the focus of this study will be the high front tense rounded [y] vowel.

The principal question which will guide the analysis in this study is:

To what extent do L2 learners approximate idealized formant values for the [y] vowel, and what is the formant structure of the actual pronunciation?

The hypothesis is that Mandarin learners will not produce Mandarin phonemes in the same way that native Mandarin speakers do. As indicated by the findings of the Baker-Smemoe & Smith study, English speakers are expected to produce phonemes that may sound more similar to English vowels, per the Perceptual Assimilation Model. The specific phoneme being analyzed in this study will be the high front tense rounded [y], which I hypothesize will be replaced by the high back tense rounded [u] for two reasons. Prior research has found support for PAM. First, the [u] phoneme differs only in backness, making it articulatorily similar to [y]. However, [i] also only differs in one feature (roundness). For this reason, researchers have also studied the impact of orthography on phoneme replacement choice. Because [y] is written in Pinyin as /ü/, it has a similar orthography to the more familiar /u/. Therefore, speakers may be more likely to use [u] than [i] as a replacement because of how the phoneme appears in its Pinyin form, as predicted by Bassetti.

4: Methods

The focus of the study is a vowel that exists in Mandarin but does not exist in English, [y]. L2 speakers of Mandarin were asked to pronounce a series of Mandarin words, both containing and not containing the particular vowel.

Four speakers' voices were used for this exploratory study. The speakers are all native English speakers between the ages of 20-26 with 3-4 years experience learning Mandarin in a classroom setting. Some have experience learning other languages, such as Spanish and Russian.

The passage the participants are asked to read is as follows:

绿色的绿。

[ly sʌ dʌ ly]

"Lü" from the phrase 'green'

你好的你。

[ni hɑʊ dʌ ni]

"Ni" from the phrase 'hello'

女性的女。

[ny ʃiŋ dʌ ny]

"Nü" from the phrase 'woman'

速度的速。

[su du dʌ su]

"Su" from the phrase 'speed'

旅行的行。

[ly ʃɪŋ dʌ ʃɪŋ]

"Xing" from the phrase 'travel'

地图的地。

[di tu dʌ di]

"Di" from the phrase 'map'

It contains 5 instances of the [y] vowel for analysis. It additionally contains 4 instances of [i] and 4 instances of [u], for a total of 13 phonemes subject for analysis. This passage aims to see if the speakers pronounce the [y] and [u] phonemes differently.

Following data collection, the samples were manually segmented using Praat (Boersma and Weenink 2024) for the 13 phonemes of interest. After manual segmentation, a Praat script was run on each file that measures the average F1, F2, and F3 HZ for each phoneme. These numbers were taken into an Excel sheet, where the average HZ was calculated for each formant. These averages are reflected in Tables 3 and 4 for each speaker.

5: Data

	F1	F2	F3
Proper [y]	401	2409	2913
Speaker 1 [y]	381	2246	2623
Speaker 2 [y]	364	1822	2437
Speaker 3 [y]	416	1007	3101
Speaker 4 [y]	281	2162	2503

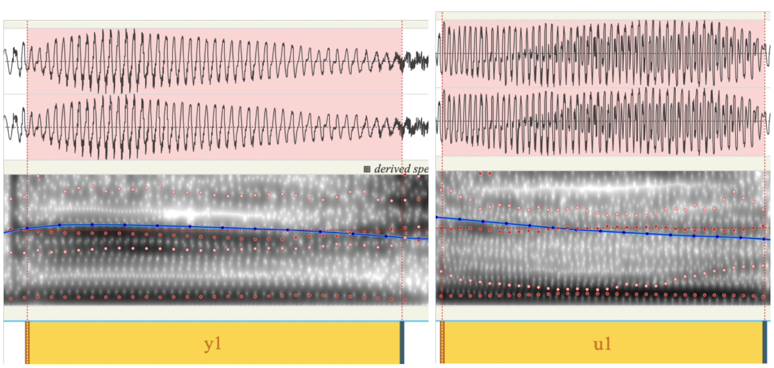
Table 3: Mean HZ measurements for each formant for high front tense rounded [y] vowel over all instances in the utterance

	F1	F2	F3
Proper [u]	402	896	2790
Speaker 1 [u]	406	883	2772
Speaker 2 [u]	587	1957	2657
Speaker 3 [u]	396	981	2961
Speaker 4 [u]	370	1022	2845

Table 4: Mean HZ measurements for each formant for high back tense rounded [u] vowel over all instances in the utterance

6: Analysis

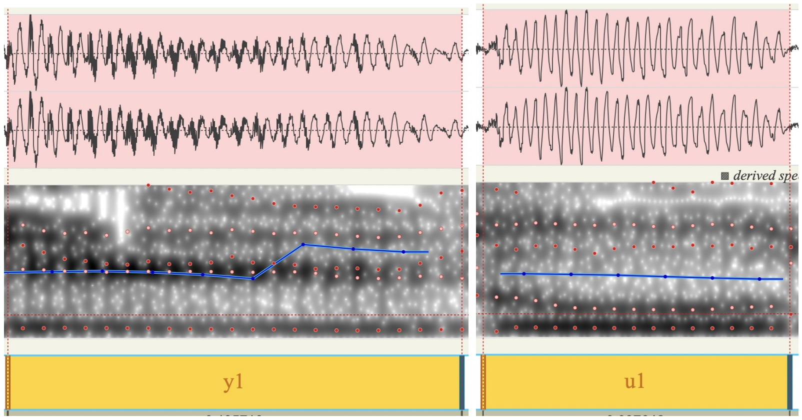
Speakers 1 and 4 had pronunciations that closely modeled the idealized pronunciation. Speaker 1 exhibited nearly perfect pronunciation for both the [y] and [u]. They had formant measurements closest to the ideal pronunciation for both phonemes. The images and measurements in Figures 5 and 6 both clearly demonstrate this conclusion. Firstly for [y], there is a separation between F1 and F2 for [y] of 1,865 HZ and F2 and F3 only have a distance of 377 HZ. For [u], there is a separation between F1 and F2 for [u] of 477 HZ, while F2 and F3 are 1,889 HZ apart. These distances are about what is expected for the idealized pronunciation, and therefore it can be concluded that this speaker has accommodated their speech to the Mandarin phonetic inventory appropriately.



Figures 5 & 6: Speaker 1 [y] and [u]

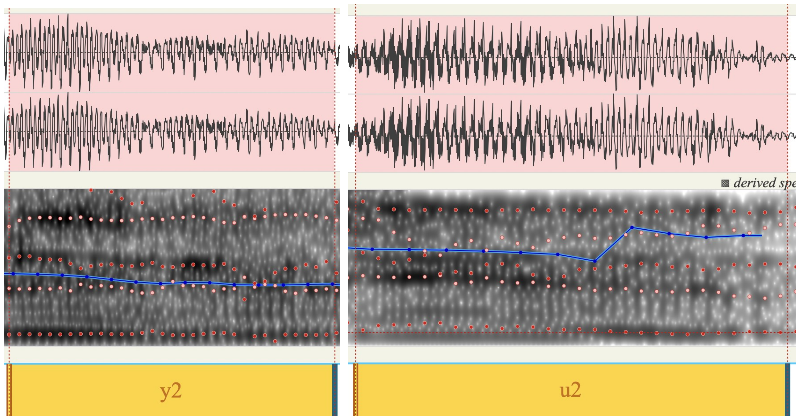
Speaker 4 also showed appropriate pronunciation and formant measurements for a Mandarin inventory. The images and measurements in Figures 7 and 8 support this conclusion. For [y], F1 and F2 have a distance of 1,881 HZ, and F2 and F3 have a distance of 341 HZ. For [u], F1 and F2 have a distance of 652 HZ, and F2 and F3 have a difference of 1,823 HZ. From this data, we can conclude that this speaker has also assimilated to the Mandarin phonetic inventory.

DOI: 10.5195/pur.2025.104



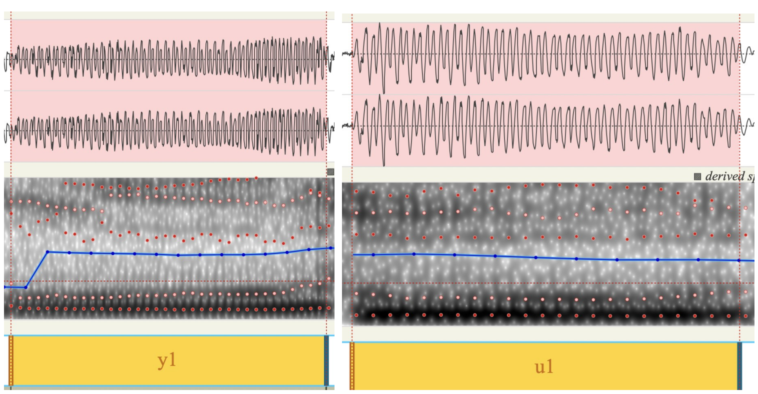
Figures 7 & 8: Speaker 4 [y] and [u]

Speakers 2 and 3 had results that deviated both from the idealized pronunciation, but in different ways from each other. Speaker 2 seems to have an intermediary pronunciation, with formant measurements falling somewhere between the expected frequencies for the idealized phoneme. This speaker had almost equal measurements for distances between F1/F2 formants and F2/F3 measurements for each phoneme, which is not representative of either phoneme. While the expected distance between F1 and F2 is around 2000 HZ for [y] and 500 HZ for [u], this speaker had measurements of 1,458 HZ and 1,370 HZ respectively. Both of these measurements are near the exact middle of 1,250 HZ between what is expected for both differences. This means that for both F1/F2 differences, the pronunciation mirrors neither what is expected of [y] or [u]. However, a different observation can be made about the F2 and F3 differences, where the expected differences are 500 HZ for [y] and 2000 HZ for [u]. Both productions showed HZ differences measuring close to the expected measurement for the [y] phoneme with 615 HZ and 700 HZ for [y] and [u] respectively. This means that for both F2/F3 differences, the pronunciation is similar to the expected distance for the [y] phoneme. Therefore, this speaker neither assimilates their speech to the Mandarin phonetic inventory nor replaces the [y] phone with the [u] phone. Instead, they use a sound that is not characteristic of either phoneme in its ideal state. Further research could be done to see whether this speaker assimilates to the Mandarin phonetic inventory over time by using a longitudinal study. A possible explanation for this is this speaker's language background; they had 4 years of studying Spanish before they started learning Chinese, which could have an additional impact. Further research would need to be conducted on the potential influence of L2 on L3.



Figures 9 & 10: Speaker 2 [y] and [u]

Finally, Speaker 3 did not pronounce the [y] phoneme at all, which is in accordance with the original hypothesis. The formant measurements for [y] and [u] both closely resemble the ideal pronunciation of [u]. The images and measurements in Figures 11 and 12 support this finding. For [y], F1 and F2 have a distance of 591 HZ and F2 and F3 have a distance of 2,094 HZ. For [u], F1 and F2 have a distance of 585 HZ and F2 and F3 have a distance of 1,823 HZ. Because the ideal pronunciation of [u] has an F1/F2 distance of 500 HZ and an F2/F3 distance of 2000 HZ, it is obvious that the pronunciation of both phonemes is like that of the ideal [u].



Figures 11 & 12: Speaker 3 [y] and [u]

DOI: 10.5195/pur.2025.104

The findings for distances for idealized pronunciation and all speakers are summarized in Table 5.

	F1 and F2 Distance		F2 and F3 Distance	
	[y]	[u]	[y]	[u]
Ideal	2000	500	500	2000
Speaker 1	1,865	477	377	1,889
Speaker 2	1,458	1,370	615	700
Speaker 3	591	585	2,094	1,980
Speaker 4	1,881	652	341	1,823

Table 5: Comparison between F1/F2 and F2/F3 differences between mean HZ measurements for each phoneme for each speaker

7: Limitations

In further studies, more speakers would be required to participate to hopefully come to a more general conclusion. A passage that inspires more spontaneous speech would control for possible over-articulation in the speakers. Lastly, a comparison to native Chinese speakers' spontaneous speech could be helpful, because the current comparison to ideal phonemes might not represent the actual realization of these phonemes in standard spoken Mandarin.

8: Conclusion

From the data, it can be shown that the predictions of the SLM were borne out for two speakers, predictions of PAM for one speaker, and neither model predicted the production of the final speaker. Two of the speakers (Speakers 1 and 4) pronounce [y] and [u] in accordance with the expected pronunciation of native Mandarin speakers. This goes against the hypothesis that native English speakers would not use the [y] phoneme. These two speakers' phonological acquisition follows the SLM-r model because they were able to create a new phonetic category for the [y] phoneme.

The hypothesis that these speakers would use the PAM model was only supported by 1 of the speakers (Speaker 3), who used the [u] phoneme in places in the passage where either [y] or [u] occurred. This speaker assimilated unfamiliar phonemes to ones that exist in their L1, which was articulatorily similar in all qualities besides frontedness. Finally, speaker 2 demonstrated neither model of phonological acquisition. An intermediate phoneme was used in place of either [y] or [u]. Further research can be done to determine the phoneme produced by Speaker 2, and whether this speaker would produce the standard phoneme if given more time with the language.

Therefore, the formant structure realized by Chinese language learners differs based on the speaker. Multiple compensation strategies are employed for L2 speakers of Chinese when they are faced with an unfamiliar phoneme, notably [y]. According to this study, which strategy is used is a phenomenon distinct to each speaker. Further research could be done on specific causes that determine which model of phonological acquisition particular speakers use, but factors such as length of study, learning style and environment, language background, and conversational experiences could possibly shape the utterances of these speakers.

References

American Speech-Language Hearing Association. n.d. "Mandarin Phonemic Inventory." American Speech-Language Hearing Association. https://www.asha.org/siteassets/uploadedfiles/multicultural/mandarinphonemicinventory.pdf?srsltid=AfmBOoq83LzqDHW7-mhCk8OpvqVU_mjBVYUQujaug1UOsYCd2rDhB14I.

Baker-Smemoe, Wendy, and Laura C. Smith. 2010. "The Impact of L2 Dialect on Learning French Vowels: Native English Speak English Speakers Learning Que´be´cois and E ers Learning Que´be´cois and European F opean French." *Canadian Modern Language Review* 66:711-738.

Best, Catherine T., and Michael D. Tyler. 2007. "Nonnative and second-language speech perception." *Language Experience in Second Language Speech Learning*. 10.1075/lllt.17.07bes.

Boersma, Paul, and David Weenink. 2024. *Praat: doing phonetics by computer*. 6.4.18. Computer program.

Cheng, Fenglan. 2023. "The Role of First Language in Second Language Acquisition." *Atlantis Press Proceedings of the 2022 4th International Conference on Literature, Art and Human Development (ICLAHD 2022)*, 1236-1243. 10.2991/978-2-494069-97-8_158.

Flege, James E. 1991. "The Interlingual Identification of Spanish and English Vowels: Orthographic Evidence." *Quarterly Journal of Experimental Psychology* 43 (3): 701-731. https://doi.org/10.1080/14640749108400993.

Flege, James E., and Ocke-Schwen Bohn. 2021. "The Revised Speech Learning Model (SLM-r)." *Second Language Speech Learning: Theoretical and Empirical Progress*, 3-83. 10.1017/9781108886901.002.

Fon, Janice. 2020. "The Phonetic Realizations of the Mandarin Phoneme Inventory: The Canonical and the Variants." In *Speech Perception, Production and Acquisition, Multidisciplinary approaches in Chinese languages*, 11-36. N.p.: Springer, Singapore. https://doi.org/10.1007/978-981-15-7606-5_2.

Hayes-Harb, Rachel, and Shannon Barrios. 2021. "The influence of orthography in second language phonological acquisition." *Language Teaching* 54, no. 3 (July): 297-326. https://doi.org/10.1017/S0261444820000658.

Jiang, Xiaoli, and Andrew D. Cohen. 2018. “Learner strategies for dealing with pronunciation issues in Mandarin.” *System* 76:25-37. https://doi.org/10.1016/j.system.2018.04.012.

Lee, Wai-Sum, and Eric Zee. 2003. “Standard Chinese (Beijing).” *Journal of the International Phonetic Association* 33 (1): 109-112. 10.1017/S0025100303001208.

Liu, Qian. 2011. “Factors Influencing Pronunciation Accuracy: L1 Negative Transfer, Task Variables and Individual Aptitude.” *English Language Teaching* 4, no. 4 (December). 10.5539/elt.v4n4p115.