Difficulties in the perception of Mandarin Chinese vowels [ɤ] and [ɿ]/[ʅ] by Hungarian learners of Mandarin

Libo Fan

University of Szeged

Abstract

This study investigates how Hungarian learners of Mandarin Chinese identify the vowels [ɤ], [ɿ], and [ʅ], which are absent from Hungarian phonology. It focuses on how learners distinguish the mid vowel [ɤ] from high vowels [ɿ]/[ʅ] at different stages of learning and explores the factors influencing their perception. Two main variables are considered: the quality of input (native vs. non-native Chinese teachers) and the quantity of input (beginner vs. intermediate learners). Consonant context as a variable also appears in the analysis, along with qualitative formant measurements on the samples of the perception test, investigating and establishing a possible explanation for the results. Participants included 21 beginners and 10 intermediate learners. Beginners were divided into three subgroups based on teacher type: native Chinese, Hungarian L2 speakers of Chinese, and both. Intermediate learners were taught by both types of teachers. An X(AB) perceptual identification test was used to investigate the perception of Chinese vowels [ɤ] and [ɿ]/[ʅ] among Hungarian learners of Chinese. Results showed that [ɤ] was identified less accurately than the high vowels. Learners taught by a native speaker performed better, highlighting the importance of input quality. Surprisingly, intermediate learners did not outperform beginners, which may be due to orthographic interference and fossilisation of pronunciation skills rising from the elimination of pronunciation training in advanced classes. Overall, the study suggests that both teacher background and the writing system affect the perceptual identification.

1. Introduction

Mandarin Chinese (hereafter Chinese) has been the official language of China for a few decades. It is used in schools and universities, and on national radio and television broadcasts (Duanmu, 2007:4). An increasing number of Hungarian speakers learn Chinese from year to year (Simay et al., 2020), which raises questions on the similarities and differences between the two languages, with

Email address: flbwyf@gmail.com (Libo Fan)

Beszédtudomány – Speech Science 2025/2. 82–111. DOI: 10.15775/Besztud.2025.2.82-111

the aim to improve and adapt the teaching methods to the specific requirements of the target learners. The present study explores one of these questions: the perception, specifically the identification, of Chinese vowels [ɤ]/[ɿ]/[ʅ], that are neither part of the vowel phoneme system, nor appear as allophones in the speakers' mother tongue. In general, our goal is to describe the possible difficulties of Chinese learners with Hungarian as their mother tongue, and to discuss the possible explanations for them, in order to make the results usable in the future improvement of teaching methodology.

1.1. Rationales of the question of [ɤ] and [ɿ]/[ʅ] perception

An understanding of how learners acquire a new phonological system must account for both the linguistic differences between the native and target language, and the universal facts of phonology (Gass et al., 2013). When learning non-native languages, the influence of previous linguistic experience is particularly significant (Strange, 1995). Contrastive Analysis Hypothesis (CAH) assumed that learners tend to transfer the patterns of their native language structure to the foreign language, and it is also assumed that this is the major source of difficulty or ease in learning the structure of a foreign language. Similar structures are assumed to be easy to learn, while different ones are considered to be difficult (Lado, 1957:59). We treat all non-native languages as L2 languages in this study. Hungarian is L1 and the phrase "second language" L2 of Hungarian learners refers to Chinese in the present research; in other words, Chinese is L2, and Hungarian learners of Chinese are analysed. We have to note, however, that Chinese is the 3rd or 4th language for most speakers, since English and/or German (and often another foreign language) is compulsory in primary and secondary education in Hungary.

Eckman (1977) proposed the Markedness Differential Hypothesis (MDH), grounded in a phonological theory of markedness. According to this proposal, the most difficult structures to learn are those being both different and more marked at the same time compared to the corresponding native language structure. Chinese vowels [ɤ]/[ɿ]/[ʅ] that are the focus of the present study neither

exist in Hungarian, nor are common in natural languages. Therefore, these vowels can be considered difficult for Hungarian learners of Chinese.

In the present experiment, there are several reasons to analyse the identification between [ɤ] and [ɿ]/[ʅ]. First, the Chinese [ɤ] and [ɿ]/[ʅ] vowels, which are not part of the Hungarian vowel system, are allophones of vowel phonemes in Chinese /ə/ and /i/, respectively, and appear in the same consonantal contexts in Chinese. Second, based on Fan's questionnaire results (Fan, 2024), Hungarian learners of Chinese do not consider Chinese vowels [ɤ]/[ɿ]/[ʅ] to be difficult, while Chinese teachers reported [ɤ] causing various problems to the learners, and eight out of twenty participating teachers claimed that their students also face difficulties learning [ɿ]/[ʅ] vowels.

The Speech Learning Model (SLM) proposed by Flege (1995), the revised SLM (SLM-r) proposed by Flege & Bohn (2021), and the Perception Assimilation Model (PAM) stated by Best (1995) and Best et al. (2001) suggest that segments that are not part of the language learners' mother tongue would be more difficult to learn, as the students need to acquire both their perception and production. The production of the segments in question was addressed by Juhász (2020). Her results showed that the L2-learners' pronunciation of [ɤ] was significantly different from that of the Chinese native speakers, but [ɿ] and [ʅ] did not show any significant difference. The present research aims to broaden the scope of the investigation related to this issue, focusing on the perception of the Chinese L2 sounds at hand.

1.2. Chinese and Hungarian vowels: phonological and phonetic aspects: with focus on [ɤ] and [ɿ]/[ʅ]

As mentioned before, the L1 phonological system directly affects the acquisition of L2 speech sounds (SLM, PAM). Thus, in the next section, L1 and L2 (i.e., Hungarian and Chinese) vowel systems are compared to discover the problematic aspects of the [ɤ] and [ɿ]/[ʅ] speech sounds. The Hungarian vowel system includes 14 phonemes /i, iː, u, uː, y, yː, ø, øː, ɑ, aː, o, oː, ɛ, eː/ that do not have allophonic variation. The Chinese vowel system includes five phonemes

/a, ə, i, u, y/ with 9 contextual allophones altogether (Figure 1). Zhu's (2010) phonological-phonetic theory is applied in my research, as the present study only focuses on monophthongs in open syllables. [e] in Figure 1 only occurs in diphthongs, while other vowel variants, appearing in diphthongs and triphthongs, are not shown in the present study either. As the present study focuses on the perception of mid [ɤ] and high [ɿ]/[ʅ] speech sounds, we describe these in more detail.

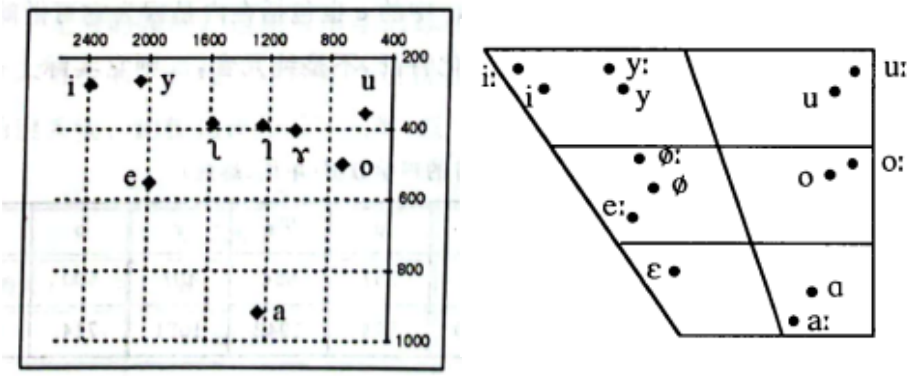

Figure 1: Mandarin vowels (left) (Zhu, 2010: 268); and Hungarian vowels (right) (Szende, 1994).

The phonological category and behaviour of the two apical segments [ɿ] and [ʅ] are subject to debate. Even though their acoustic structure is characterized by a formant structure indexing their vocalic nature from a purely phonetic viewpoint, some scholars regard them as syllabic fricatives, hence the segments are transcribed with the following IPA graphemes (even though they are the same speech sounds): [z] and [z̩] (Chao, 1934; Hartman, 1944; Pulleyblank, 1984; Lin, 1989; Wiese, 1997; Duanmu, 2000, 2007. cf: Lee-Kim, 2014). The second proposal is the approximant account, wherein [ɿ] and [ʅ] are written as [ɹ] and [ɻ], respectively (Lee-Kim, 2014). In the present study, we adhere to the traditional Chinese phonological analysis, i.e. we treat them as apical vowels (Cheng, 1966, 1973; Lee & Zee, 2001), since [ɿ] and [ʅ] are in complementary distribution with the high front vowel [i]. [ɿ] only occurs after the non-retroflex

consonants [ts, tsʰ, s], while [ɿ] only appears after the retroflex segments [tʂ, tʂʰ, ʂ, ɻ], and [i] occurs in other environments. This speech sound variation based on the preceding consonant can be summed up as contextual allophony with a clear-cut complementary distribution (Duanmu, 2000). Based on X-ray images taken by Zhou & Wu (1963), the tongue tip/blade gesture of [ɿ]/[ʅ] inherited from the preceding dental and retroflex consonants remains nearly unchanged for the following voiced period. The retroflex or non-retroflex feature of the preceding consonant spreads onto the vowel, [ʅ] is called the retroflex apical vowel, while [ɿ] is called the dental apical vowel (Cheng, 1973:13). Zhang (2003) performed an identification test and a discrimination test for isolated synthesized [ɿ] and [ʅ] stimuli. The results of the discrimination test show that the perception of [ɿ] and [ʅ] by native Chinese speakers is non-categorical, compared to the extreme categorical perception of stops. The identification test shows that the patterns of the first two formants determine the categorical boundary for phonemic distinction between [ɿ] and [ʅ], but F3 is also a potent cue for identifying the two apical vowels. Zhang states that maybe the listeners use some cues other than those manipulated in the identification of the stimuli as phonemes for discriminating the stimuli (Liberman et al., 1961).

The question of [ɤ] is generally clear. It is a mid back, unrounded vowel (Duanmu, 2007:37; Lin, 2007:73; Chen et al., 2019) and is generally regarded as a contextual allophone of the mid central vowel /ə/ in the nucleus of open consonant-vowel (CV) syllables. [ɤ] enjoys a more expansive phonotactic context, it does not only occur after [ts, tsʰ, s] and [tʂ, tʂʰ, ʂ, ɻ], but also some other consonants like [k, kʰ, x, t, tʰ, n, l]. Therefore, the mid vowel [ɤ] and the apical vowels [ɿ]/[ʅ] have a different status in the Chinese vowel system. [ɿ] and [ʅ] occur in limited contexts, [ɤ] occurs in more contexts. For the present study, it is important that [ɿ]/[ʅ] and [ɤ] share their consonantal contexts: [ɤ] can appear in all syllables where the /i/ allophones in question can.

Despite being allophones of the same vowel /i/, [ɿ] and [ʅ] segments are produced at different places of articulation (post-alveolar & dental), hence they are denoted by two different IPA-symbols and considered two distinct segments for

L2 learners to acquire. However, these vowels are not perceptually differentiated by Chinese natives (as it is found by Zhang, 2003), resulting from the fact that that they can be predicted by the consonantal context because they are allophones appearing in complementary distribution. Therefore, we will consider the apical vowels as one category within the identification task, and the test only contrast the following two pairs: [ɿ] with [ɤ] and [ɻ] with [ɤ]. Precisely the present research only focuses on the syllables starting with consonants [tʂ, tʂ^{h}, ʂ, ʐ, ts, ts^{h}, s], which can precede both the /i/ and /ɤ/ allophones in question, and ending with vowels [ɤ, ɿ, ɻ] (Table 2).

1.3. Orthographic considerations of non-native perception

Wang (2001) studied the vowel perception of Japanese and Korean learners of Chinese. She chose the identification of two or three vowels specifically for the relation of the vowel system of the given mother tongue and Chinese in a similar (X(AB)/X(ABC)) identification test as in our study. The results of Wang's research (2001) also raise the question of the perception of [ɻ] by listeners with Korean and Japanese as their mother tongue. This vowel does not exist in these two languages. When [ɻ] was played, they had to choose an option from A[ɤ]/B[ɻ]. The Korean students chose [ɤ] nearly 3 times more than [ɻ]. On the contrary, Japanese students did not choose [ɤ] at all. Wang pointed out that this might be affected by the orthographic symbol <i>, since in Chinese, <i> represents [i].

L2 orthographic input has been confirmed to show effects on the acquisition of non-native phonological and phonetic patterns. L2 orthographic input interacts with the acoustic input, influencing L2 learners' mental representations of L2 phonology, and orthography-induced pronunciations may be part of the acoustic input for instructed learners (Bassetti, 2008). Erdener & Burnham (2005) found that, while all L2 learners were more capable of repeating L2 words when they saw graphemes of the words, the effect was stronger or weaker depending on the level of phonological transparency of both L1 (native language) and L2 orthographies. It is found that native users of transparent

L1 writing systems are more negatively affected by a less transparent L2 orthography. In our case, the Hungarian writing system is phonologically highly transparent, whereas Pinyin, which is used in language teaching as a base for Chinese, is more opaque. Pinyin is the transcription of Chinese characters using the graphemes of the Latin alphabet. < > is used to represent graphemes in the present study. In Chinese, the grapheme <e> represents /ə/, therefore also its allophones: [ɛ], [ə], [ɤ], [e], [o] (Xu, 1980:33, cf: Duanmu, 2007:37). And the grapheme <i> represents /i/, therefore [j], [i], [ɿ] and [ʅ]. However, in Hungarian, <e> represents /ɛ/, therefore [ɛ], and <i> represents /i/, and therefore its allophones. Thus, in our case, the grapheme <i> denotes two different segments in Chinese (the apical vowels [ɿ] and [ʅ]). As we can see in terms of orthography, these apical vowels are differentiated from the mid back vowel because they are denoted by different graphemes, i.e., [ɿ] and [ʅ] are denoted by the grapheme <i> and the mid vowel [ɤ] is denoted by <e>. However, while <i> in Hungarian also denotes /i/, but there is no allophonic variation, <e> denotes [ɛ] also without allophonic variation. Thus, in L2 Chinese, learners are required to associate different speech sounds with the same grapheme, by recognizing the determining role of the onset consonant, which might pose difficulty since the segment-to-grapheme link in the L1 is almost exclusive and apparent and not characterized by this variation. The present study also considers the phonetic context. Based on PAM (Best, 1995), the consonants surrounding a vowel affect how that vowel will be perceptually assimilated. Flege & Bohn (2021) also state that it is important to note that the context in which input is assessed may also influence how well the input is consolidated and thus indirectly influence speech learning.

Finally, the SLM also proposed that L2 learners gradually "discern" L1-L2 phonetic differences as they gain experience using L2 in daily life, and that the accumulation of detailed phonetic information with increasing exposure to statistically defined input distributions for L2 sounds will lead to the formation of new phonetic categories for certain L2 sounds (Flege, 1995). Furthermore, in SLM-r, Flege & Bohn (2021) state that the quality of input has been largely

ignored in L2 speech research even though it may well determine the extent to which L2 learners differ from native speaker. In the present study, the learner's experience with the quantitative input (two groups with different L2 experience) and the qualitative dimension (students with teachers of different L1, Hungarian and Chinese) are considered as well.

The relationship between perception and production is that many L2 production errors have a perceptual basis. Flege (1995) suggested that L2 production accuracy is limited by perceptual accuracy. And the PAM also holds that the pronunciation difficulty encountered by L2 learners is determined by perceptual limitations (Best, 1995; Best et al., 2001). However, the SLM-r (revised Speech Learning Model) assumes that L2 segmental production and perception coevolve without precedence (Flege & Bohn, 2021).

1.4. Formant values of [ɤ] and [ɿ]/[ʅ]

The [ɤ] and [ɿ]/[ʅ] vowels in question were studied in the pronunciation of native speakers and language learners of Chinese by Fan (submitted). We assume that there is a direct link between the characteristics of Chinese vowel formants and the perceptual traits of Hungarian L1 speakers. The production results of a native Chinese speaker showed that the first formants are lower in the two apical vowels [ɿ]/[ʅ], while the second formants are higher than the mid vowel [ɤ]. The second and third formants of the retroflex high vowel [ʅ] are closer to each other, mainly due to a shift in both values.

Zhu's (2010) data (Figure 1, left) show no difference in the first formant between the two high vowels, while Fan's (submitted) do.

We also did a qualitative comparison between acoustic values of Chinese [ɤ], [ɿ], [ʅ] and the Hungarian perceptual vowel map/space (Figure 3).

We can see the Hungarian perceptual vowel map/space and the Chinese acoustic data. Comparing Figure 2 and Figure 3, we can see that the acoustic measurements of [ɤ] and [ɿ]/[ʅ] are mapped onto the same Hungarian perceptual category of <ö/ő> ø. And it can be seen that the three analysed vowels [ɤ] and [ɿ]/[ʅ] are quite close to each other in Figure 1 as well. Hungarian learners

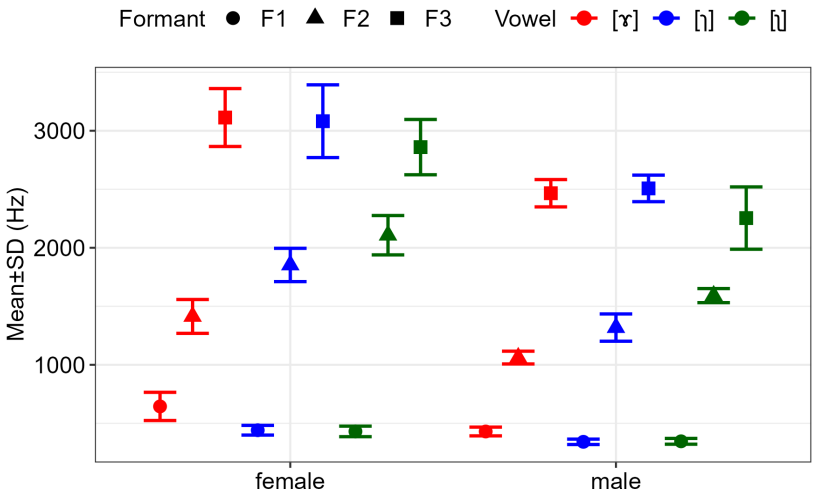

Figure 2: Formants of the high and mid vowels (Fan, submitted).

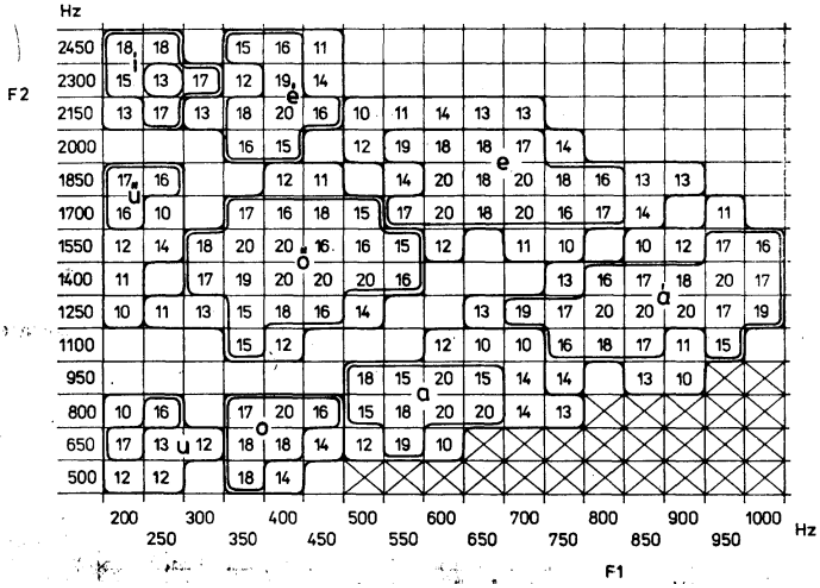

Figure 3: Plot of Hungarian vowels (Kiss, 1985).

of Chinese produce [ɤ] in a more acoustically palatalized way (Juhász, 2020). These facts might result in interference in the perception/identification of the

vowels, which is the motivation for designing a two-alternative forced-choice test in 2.2.

1.5. Hypotheses

Based on the theoretical considerations, the production study by Juhász (2020), and Fan's (2024) questionnaire and interview results, the following hypotheses were formulated along the two variables investigated in this perceptual analysis (i.e., quality and quantity of the L2 input):

i) The correct identification of [ɤ] is lower than that of the apical vowels [ɿ] and [ʅ], as their production was found to be native-like, suggesting a possibly more stable perception.

ii) The perception of the attested vowels may not show a difference between advanced learners and beginners, as the ratio of pronunciation and perception teaching is lower in the case of the former, and they have been more exposed to Pinyin.

iii) The learner groups having a native speaker as at least one of their teachers will have higher correct response ratios, as they experience native pronunciation in a higher ratio.

iv) Identifying [ɤ] and [ɿ] vowels after non-retroflex [ts, ts^h, s] could induce a lower identifying rate than [ɤ] and [ʅ] identification after retroflex [tʂ, $tʂ^h$, ʂ, ɻ], as the F2 and F3 show a larger difference between [ɤ] and [ʅ].

2. Methods

2.1. Subjects

31 native Hungarian speakers participated in a discrimination task (Table 1). All participants were born and raised in Hungary. They were divided into four groups: The advanced group included 10 subjects who had been learning Chinese for five semesters. During the first two semesters, they were taught online by a native Chinese and a native Hungarian L2-speaker of Chinese, and

from the third semester, they started attending physical classes with the same teachers. The rest of the participants (21) were beginners, who had been learning Chinese for one semester. The beginners were divided into three groups: 7 of them were taught exclusively by a native Chinese speaker, 6 of them by a native Hungarian L2-speaker of Chinese, and 8 of them by both a native Hungarian L2-speaker of Chinese and a native Chinese speaker. The native Chinese teacher was consistent across all three groups that included a native teacher.

No participant reported any hearing problems. One student in the advanced group had spent half a year in Taiwan as an exchange student studying economics, but she did not learn Chinese. The remaining participants had never been to China before. All participants use Hungarian as their primary language in daily life. They all could also speak English. Some participants had learned other languages as well. None of the sounds included in the task ([ɤ]/[ɿ]/[ʅ]) are part of the English sound system either, thus in the present study, the interference of English is not considered.

All the participants were university students aged between 18 and 22 years.

Table 1: Participants.

	teacher(s)	**no. of students**	**group name**	**no. of semesters**
beginner groups	native Chinese speaker	7	BegChi	1
	native Hungarian speaker	6	BegHu	1
	joint teachers	8	Begmix	1
advanced group	joint teachers	10	Intmix	5

2.2. Stimuli

All stimuli consisted of CV-syllables ending in [ɤ], [ɿ] or [ʅ], with the high vowels depending on the expected vowel after the onset (Table 2). Each stimulus represents a lexical item in Chinese. To exclusively attest vowel differentiation and exclude the possible effect of the tones, the present study introduces only

the results for items in falling tone 4. Tone was limited to tone 4 because all 14 test items are meaningful words in Chinese with tone 4, while not all of them exist with other tones. The vowels [ɿ] or [ʅ] appear after seven consonant onsets [ts, ts^h, s, tʂ, $tʂ^h$, ʂ, ɻ], hence 14 items are analysed in this study. Besides the 14 stimuli recorded for the present study, 114 further items (test items for different vowel allophones and tones and distractors) were played three times each (N = 384) in randomized order. The actual test was preceded by a short training period with 3 non-test items.

The test items were recorded by 6 native speakers in a sound-treated room using a head-mounted microphone via Speech Recorder (Christoph & Klaus, 2004). After recording the sounds, a professional native Chinese voice actor was asked to judge the six speakers' pronunciation. The voice actor had passed a Mandarin Chinese proficiency test, was born in Hebei and finished his university in Beijing, and he had dubbed Chinese textbooks as well, therefore he can be accounted as a high proficiency speaker with a well based judgmental reliability. The speaker whose reading was used for the perception test was chosen based on the scores given by the native actor. The one closest to the ideal Chinese pronunciation was selected. The male speaker is from Mainland China, and he had been studying in Hungary for 3 years at the time of recording. During these three years, he was mostly in a Chinese-speaking environment, except for attending classes at the university.

Table 2: Stimuli.

non-retroflex		**retroflex**	
Pinyin	**IPA**	**Pinyin**	**IPA**
<zè> – <zì>	[tsɤ] – [tsɿ]	<zhè> – <zhì>	[tʂɤ] – [tʂʅ]
<cè> – <cì>	[ts^hɤ] – [ts^hɿ]	<chè> – <chì>	[$tʂ^h$ɤ] – [$tʂ^h$ʅ]
<sè> – <sì>	[sɤ] – [sɿ]	<shè> – <shì>	[ʂɤ] – [ʂʅ]
		<rè> – <rì>	[ɻɤ] – [ɻʅ]

In Table 2, the retroflex feature indexes the consonant context, since [tʂ, tʂʰ, ʂ, ɻ] are retroflex sounds.

2.3. Experiment design

The present research examined identification of the vowels discussed above in a two - alternative forced-choice (2AFC) test by adult native Hungarian speakers. The perception test was administered in a quiet room through headphones using Praat ExperimentMFC (Boersma & D., 2022).

The subjects were instructed to select the word they heard as soon as possible after listening to the stimulus (Table 2). After the subject listened to a single stimulus, they had to click one of two responses. In other words, the participant had to select one from two possible answers displayed in Pinyin: <zè>/<zì>, <cè>/<cì>, <sè>/ <sì>, <zhè>/<zhì>, <chè>/<chì>, <shè>/<shì> and <rè>/<rì>. This means that the participants did not have to decide between the two high vowels, but between the mid and one of the high vowels. The response and the reaction time were recorded. The listeners also had to judge how sure they were in their answer on a 1 to 5 scale where 1 meant absolutely unsure, 5 meant absolutely sure. The present study will not analyse the goodness responses.

2.4. Statistics

The answers were analysed in R (R Core Team, 2022). The correctness of the answers was analysed using Binomial Generalized Linear Mixed Models (BGLMM), and for the reaction times Generalised mixed effects linear models were run for logistic distribution (GLMM, lme4: Bates et al., 2015; lmerTest packages: Kuznetsova et al., 2017). A Tukey post hoc test was administered to attest the effects of the interactions (emmeans package: Lenth, 2021). The models were built in a top-down selection method: the simplest model was chosen and that was still not significantly different from the possible largest, converging model.

The correctness of the answer was set as the dependent variable in the BGLMMs, while reaction time was the dependent variable in the GLMMs. The factors were the following: phoneme category (i.e., mid or high vowel), learner group (advanced, beginner with Chinese teacher, beginner with Hungarian teacher, beginner with joint teachers), and tongue tip position (retroflex or not). The models including all three factors did not converge, therefore the tongue tip position was eliminated and attested separately. The p-value of the final model was extracted by Anova (car package: Fox & Weisberg, 2019). In order to analyse the possible effect of the retroflex context and the retroflex feature of the vowel on the correctness of the identification, the results for the mid and high vowels were tested separately using two further BGLMMs. The correctness of the answer was set as the dependent variable, and the retroflex feature and the learner group were set as factors. The model selection and the extraction of p-value were run as described above.

All figures were drawn using ggplot2 (Wickham, 2016).

2.5. Formant analysis

The vowel formants of the test items were also measured to be used in the interpretation of the contextual effect in the perception results. The first three formants (F1, F2, F3) of the stimuli in Table 2 were measured in Praat automatically by a script. The vowels were labelled from the start to the end of the F2 in the oscillogram and spectrogram. In the case of [ɻʏ] and [ɻʉ], the spectrogram had to be set to the range of 0–8kHz in order to let the higher frequency frication appear in the view range. In the case of these sequences, the loss of this higher frequency frication was considered the start of the vowel. The formant range was set to 5 kHz in general, except in the case of the four [ʅ] vowels, where the F2 and F3 fall close to each other leading to mis-measurements. After manual checking, the formant range was set to 4.5 kHz in these four cases. The further settings were left as standard (5 formants for male voice). The measurements were taken at the mid 40 ms of the vowels.

3. Results

3.1. Reaction times

Reaction times between the correct and incorrect answers did not show any differences (Figure **??**, left). Reaction times for the correct answers showed some tendencies: reaction times for the correct answers for <i> [ɪ]/[ɪ] were shorter than for <e> [ɤ] (Figure **??**, right). Accordingly, the best fitting Generalised Mixed Effects Logistic Regression included the interaction of the vowel and the group, a random slope on the vowels by learners, and a random intercept by the stimuli. However, the results did not indicate significant differences by any of the factors or their interaction. As there was no significant difference, reaction time is not included in the latter results.

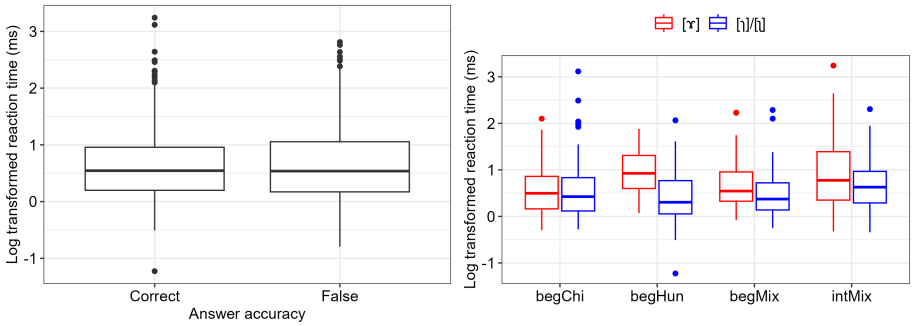

Figure 4: The reaction times (log-transformed, ms). Left: Reaction times for the correct vs. false answers, right: Reaction times for the correct answers [ɤ] vs [ɪ]/[ɪ] among the learner groups.

3.2. Accuracy rate of [ɤ] and [ɪ]/[ɪ]

The accuracy rate was calculated speaker-by-speaker for the mid and the high vowels separately. The ratio of correct answers revealed a noticeable difference in performance when identifying the two vowel phonemes. The results were grouped by the vowels and the groups. The mean value is not used in the present research because the data are not normally distributed, thus the median and the interquartile range are used for description. The best fitting

model was the one that included the interaction of the vowel and learner group with a random intercept by learners.

Based on the statistical results of the generalized linear mixed model, there is a significant difference between the accuracy rates of <e> [ɤ] and <i> [ɿ]/[ʅ] (results for the vowel factor: $\chi^2(1, 1302) = 64.99$, $p < 0.001$) (Figure 5). In addition, the highest accuracy rate for <i> [ɿ]/[ʅ] is 0.76, suggesting that the apical vowels also cause some problems. The median of <e> [ɤ] from different groups is lower than that of <i> [ɿ]/[ʅ] (Figure 6). The accuracy rates of <e> [ɤ] of begChi, begHun, begmix and intmix are 0.60, 0.30, 0.64 and 0.58, respectively. The median range of <e> [ɤ] is from 0.30 to 0.64, but the range of <i> [ɿ]/[ʅ] is from 0.67 to 0.76. This result is in agreement with Juhász's production test (2020).

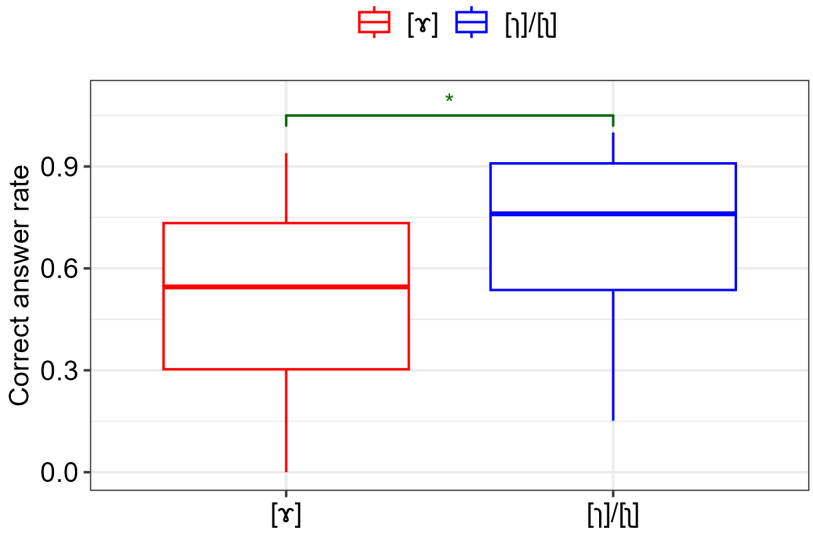

Figure 5: The rate of answer correctness for the mid vs. high vowel regardless of learner group.

The interaction of the two factors was also significant ($\chi^2(3, 1302) = 13.38$, $p = 0.004$) (Figure 6). According to the Tukey post hoc test, there is a significant difference (i) between the begChi and begHun, and also (ii) between

the begmix and begHun groups' results in the case of <e> [ɤ]. This means that among the beginner groups, the accuracy rate may be influenced by the teachers. The begChi and begmix (with a native Chinese teacher) groups have significantly higher accuracy rates than the begHun group does (with a native Hungarian teacher) for [ɤ]. Comparatively, the accuracy rate of <e> [ɤ] in the begHun group is 0.30, while it is 0.60 and 0.64 in the begChi and begmix groups, respectively. As for apical vowels, the accuracy rates of the begChi and begmix (with a native Chinese teacher) groups were also higher than in the begHun group (with a native Hungarian teacher). Comparatively, the accuracy rate of <i> [ɿ]/[ʅ] in the begHun group is 0.67, while it is 0.77 and 0.75 in the begChi and begmix groups, respectively.

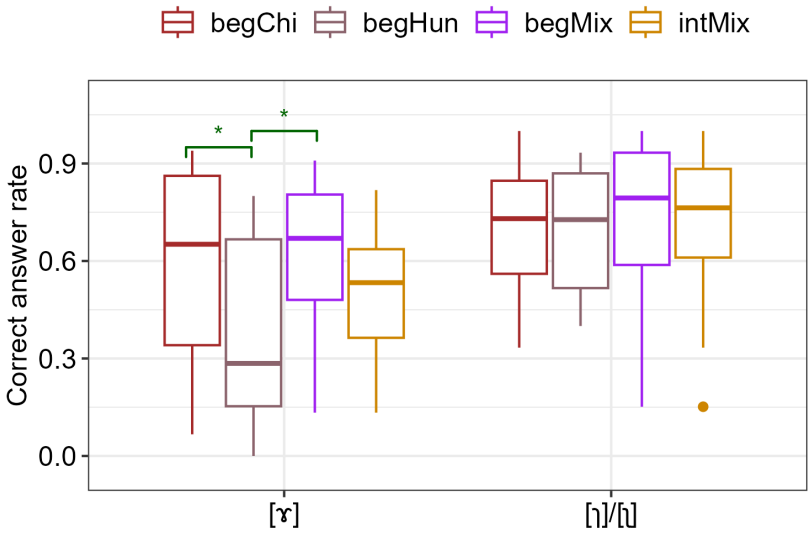

Figure 6: Correct answer rate for the vowels within the learner groups.

The advanced (intmix) group's results did not show significant differences from any of the beginner groups, which means that their results were not significantly better despite having more experience with the language.

3.3. Phonetic context for vowel perception

The possible influence of tongue tip position (retroflex or not) is shown in Figure 7 and Figure 8. We have to emphasize that the feature of being retroflex or not is intrinsic for the high vowels, i.e., these apical vowels are distinguished by this specific feature, while in the case of the mid vowel, this appears only as a contextual coarticulatory effect and thus is not supposed to appear throughout the entire duration of the vowel. Based on the ultrasound study of Lee-Kim (2014), both the tongue tip raising and tongue back retraction of [ɻ] and [ʅ] are maintained throughout the entire syllable. The accuracy rate of identifying the non-retroflex vowel [ɻ] is higher than that of the retroflex vowel [ʅ] (Figure 7). While this seems to be a difference within the groups with a native teacher (Figure 8), the best fitting generalized linear mixed model was the one that included only the factor of retroflection, but not the group, and included a random intercept by the learner. There was a significant difference between the retroflex and non-retroflex vowel ($\chi^2(1, 651) = 9.536$, $p = 0.002$).

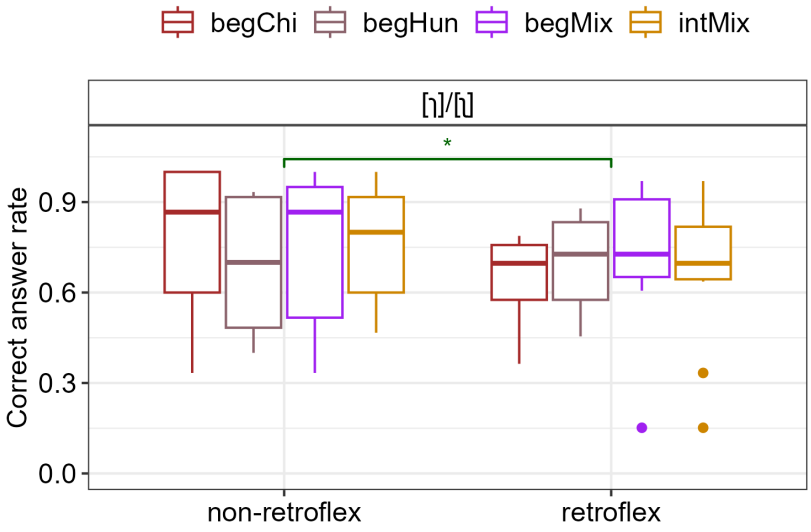

Figure 7: Correct answer rate of [ɻ]/[ʅ].

In the case of the mid vowel (Figure 8), the best fitting generalized linear mixed model was the one that included the retroflex and group factors with their interaction and random intercept by the learner. There was a significant difference between the groups ($\chi^2(3, 651) = 8.081$, $p = 0.044$), and also the interaction of the group and retroflection was significant ($\chi^2(3, 651) = 9.081$, $p = 0.028$), but no significant difference was found between the vowels. Based on the Tukey post hoc test, there is a significant difference between the begChi and begHun groups, and between the begHun and begmix groups, where the begHun group has lower correct answer rates. Although the effect of the group factor was significant meaning that there is a general difference between the begHun and the other two beginner groups, the post hoc test for the interaction of the two factors made it clear that if the contextual effect is considered, the lower perception rate in the begHun group from the other two beginner groups appears in the vowels following a retroflex consonant, while the difference does not reach the level of significance in the non-retroflex context.

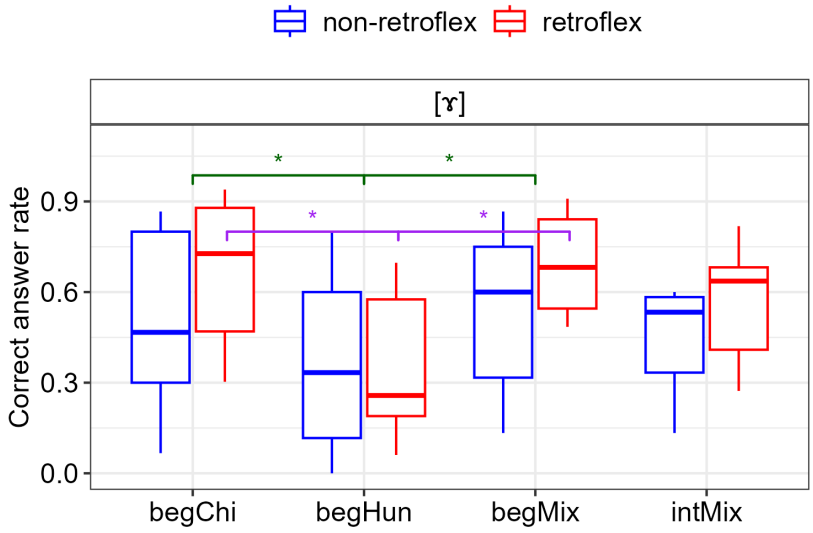

Figure 8: Correct answer rate of [ɤ].

3.4. Formant analysis

The mixed results for the contextual effect on the perception raised the question of how the formants of the mid vowel change between retroflex and non-retroflex contexts as compared to the high vowels.

Figure 9 shows the formant frequencies for the test items, and Table 3 shows the mean values. The present data lies in 14 items, and the data is only from one person, therefore we do not intend to draw general conclusions. However, in these specific data, we can see that the F2 and F3 values get closer in a retroflex context/pronunciation. What is more important for the results above, is that the formant values of the mid and high vowel are closer in a non-retroflex context, and further apart following a retroflex consonant. The larger difference must appear due to the inherent retroflex feature in the high vowel. While not willing to draw large conclusions in general for Chinese vowels, for the present data we can say the following. The formant values of the mid and high vowels lay considerably closer to each other in the present stimuli in the non-retroflex scenario compared to the retroflex scenario, in correspondence with Lee & Zee (2001). We assume that phonetic context may influence these vowels' perception by Hungarian learners of Chinese. As the results in 3.3 showed, the correct identification for [ɤ] shows higher differences across the learner groups in the retroflex context than in the non-retroflex one. However, it is the other way for the /i/ allophones. The retroflex context resulted in somewhat higher correct answer rates for the beginner groups taught at least partially by a native speaker, while the third beginner group had lower accuracy rate in this context. The perception of the high vowel was significantly lower when it was retroflex in general, regardless of the speaker groups. At the present point, we can only draw the conclusion that the effect of the context should be addressed in studies more focused on this specific question.

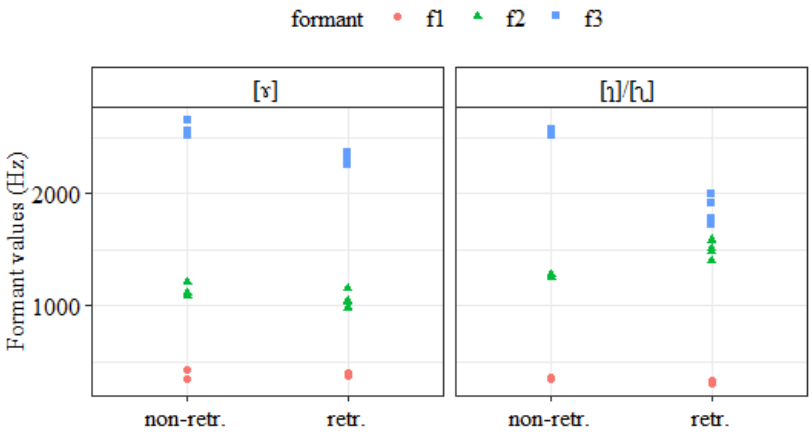

Figure 9: The mean formant frequencies (Hz) and their ratios in the test items.

Table 3: The mean formant frequencies (Hz) and their ratios in the test items.

	[ɤ], non-retroflex	**[ɤ], retroflex**	**[ɿ]**	**[ʅ]**
F1 (Hz)	403	384	350	321
F2 (Hz)	1120	1048	1259	1495
F3 (Hz)	2564	2315	2547	1852
F1/F2 ratio	0.34	0.34	0.28	0.21
F2/F3 ratio	0.44	0.45	0.49	0.81

4. Discussion

4.1. Effects of native language and markedness on the identification of Chinese vowels

The perception test's results showed that the accuracy rates of the mid vowel were lower in all listener group than that of the high vowels, which suggests that <e> [ɤ] is more difficult than <i> [ɿ]/[ʅ] for Hungarian learners of Chinese. These results are in correspondence with Juhász's (2020) results in the production domain. We assume that besides the effect of the difference in the vowel system of the target and the mother language – as suggested by CAH

and MDH –, there must be further factors influencing the acquisition of these segments. Orthographic interference may be raised as a possible reason.

[ʏ] is more difficult than [ɿ] and [ʅ] for Hungarian listeners, and [ʏ] is often perceived as [ɿ]/[ʅ]. The perception result corresponds with the production result of Juhász (2020): the formant values of [ʏ] are significantly different from native Chinese, and Hungarian learners produced the [ʏ] with higher F2 values than Chinese native speakers. There may be a connection between the perception and the production of speech sounds; however, perception and production may also develop mutually and synchronously. This is stated in the L2-learning models (Flege & Bohn, 2021).

Hungarian students in the present study confounded [ʏ] and [ʅ], which was found for Korean students (Wang, 2001) as well. In terms of sound value, [ʏ] in Chinese and [u] in Japanese are similar. However, Japanese students were more likely to perceive [ʏ] as [a]. [ʏ] in Chinese and [ø]/[øː] in Hungarian are similar, but the production result only partly suggested that Hungarian students' production is close to [ø]/[øː]. But the affirmed result is that [ʏ] is realized with higher F2 by Hungarian speakers (Juhász, 2020) and [ʏ] is proven to be confounded with [ʅ]/[ɿ] by Hungarian students in the present study. Compared to the Korean and Hungarian students, Japanese students did not perceive [ʅ] as [ʏ] at all in the selection of "[ʏ]/[ʅ] (maybe the students regard [ʅ] as [i])". Therefore, it seems that predicting and explaining the perceived difficulties of [ʏ]/[ʅ] based solely on acoustic patterns between the native language and the target language is not enough.

4.2. Effects of linguistic experience on the identification of Chinese vowels

The advanced group did not have better results than the beginner groups in our mid-high differentiation task, which is in agreement with Juhász's (2020) production test results.

As discussed in 1.3, orthographic input plays an important role in second language acquisition. More experienced, i.e. advanced learners of Chinese spend more time receiving more orthographic input, while the amount of focused pro-

nunciation (and thus perception) training decreases. Generally, Chinese pronunciation practice is in the first semester, the proper production of the Chinese segments is more highlighted and emphasized in the first year of studying, i.e., Chinese teachers are more careful and articulate more effectively to differentiate these segments. However, as time goes by, everyday communication does not require this efficient distinction between these speech sounds, thus if the mental discrimination of the categories is not established in the beginning in the L2 learner's mind, it is likely that advanced learners will face difficulties when trying to tell them apart – because "high-quality and well differentiated" input to help them discriminate these sounds becomes more and more absent. In other words: this result suggests that if the correct perceptual discrimination is not founded in the beginning of L2 acquisition, then the lack of distinction in the L2 learner's mind persists and fossilises, and may probably deteriorate as well (but this is just a hypothesis which should be addressed in another analysis). Thus, Chinese learners' mental representations of Chinese phonology may be negatively influenced by more Chinese orthographic input. Hungarian advanced learners of Chinese get more orthographic input than beginners, and Pinyin is more opaque than Hungarian orthography – as mentioned above.

4.3. Effects of consonant context on the identification of Chinese vowels

Based on Figure 7, [ɤ] and [ɿ] are situated closer in terms of their formant values compared to the distance between [ɤ] and [ʅ]. We expect that identifying [ɤ] and [ɿ] after non-retroflex [ts, ts^h, s] could induce a lower correctness rate, compared to [ɤ] and [ʅ] after retroflexes [tʂ, $tʂ^h$, ʂ, ɻ], because their second and third formants are closer in these contexts. From the results of the present study, it is only proved that [ɤ] after non-retroflexes [ts, ts^h, s] induces a lower identifying rate than [ɤ] after retroflex [tʂ, $tʂ^h$, ʂ, ɻ] in groups of Begmix, BegChi and intMix, whereas [tsɿ, $ts^hɿ$, sɿ] does not induce a lower identifying rate than [tʂʅ, $tʂ^hʅ$, ʂʅ, ɻʅ] in groups of Begmix, BegChi and intMix. This may be caused by the inherent difficulty for retroflex perception (Tabain et al., 2020). In the syllables [tʂʅ, $tʂ^hʅ$, ʂʅ, ɻʅ], both consonant contexts and the apical vowel

are retroflex, which is perhaps the major contributor to their difficulty. The magnitude of the acoustic change is less apparent (as compared to a dental C + velar V sequence), which might pose problems since there is no dynamic formant change to be used as an acoustic cue to anchor the vowel. The accuracy rate of the non-retroflex vowel [ɿ] is higher than that of the retroflex vowel [ʅ], implying that the retroflex vowel [ʅ] might be more difficult than the non-retroflex vowel [ɿ].

4.4. Effects of a Chinese teacher on the identification of Chinese vowels

The present results showed that experience with a native language teacher may have an effect on Hungarian listeners' perception. The accuracy rate was higher in the beginner groups with a native Chinese teacher than the group without one, which also implies that Hungarian teachers of Chinese have an accent and possibly also have difficulties in discriminating these sounds in production. However, based on the results of the phonetic context for vowel perception, it is interesting that BegHun is in reverse to the other groups with a native Chinese teacher, as seen in Figure 6: [ɤ] after non-retroflex [ts, ts^h, s] induces a higher identifying rate than [ɤ] after retroflex [tʂ, $tʂ^h$, ʂ, ʐ] in the BegHun group. In addition, [tsɿ, $ts^hɿ$, sɿ] also induces a higher identifying rate than [tʂʅ, $tʂ^hʅ$, ʂʅ, ʐʅ] in the BegHun group. This result suggests that experience with a native speaker through merely a native Chinese teacher may already have an impact on identification.

5. Conclusion

This study investigated the perception of Mandarin Chinese vowels [ɤ], [ɿ], and [ʅ] by Hungarian learners, drawing on empirical results and theoretical models of L2 phonological acquisition. The data clearly show that [ɤ] is significantly more difficult for learners to identify than the apical vowels [ɿ]/[ʅ], a finding aligned with previous production studies and perception-based research.

While Contrastive Analysis Hypothesis (CAH) and Markedness Differential Hypothesis (MDH) offer general predictions about difficulty in learning unfa-

miliar and marked L2 sounds, they fall short of explaining the present findings. All three vowels in question are both absent from Hungarian and relatively marked in terms of phonetic rarity, yet learners showed a clear asymmetry in performance. This suggests that markedness alone cannot predict perceptual difficulty in this context.

Another finding was that the accuracy rate in the advanced, intMix group was not proven to be significantly higher than in the beginner groups, moreover, in some aspects it showed a tendency of lower values. Our interpretation – as explained in the Discussion – is that their phonetic memory is not kept awake by focusing on the phonetic-phonemic level only during the first semester; if accurate perceptual discrimination is not established at the onset of L2 acquisition, the inability to distinguish sounds may persist, become fossilised, and potentially deteriorate further over time. Tusor (2016) states that Hungarian learners of Chinese always rely on Pinyin transcription when they are studying Chinese characters and pronunciation. The students who are not made aware that the sounds written in Pinyin are not the same as the sounds represented by the letters of the English and Hungarian alphabets will certainly tend to pronounce these speech sounds incorrectly. The incorrect pronunciation may persist and become fossilised even after abandoning Pinyin.

Furthermore, the study confirms that input quality, as predicted by SLM-r (Flege & Bohn, 2021), is a key determinant in L2 perception: learners taught by native Chinese instructors had significantly higher identification accuracy, particularly for the problematic [ʏ]. This underlines the pedagogical importance of early exposure to native input, especially for phonologically subtle or unfamiliar segments.

In summary, these findings highlight the importance of early, high-quality phonetic training, ongoing perceptual practice, and a critical approach to orthographic input in teaching Chinese to Hungarian learners. Future work should further explore how these variables interact longitudinally and whether targeted interventions can mitigate fossilisation and orthography-driven misperception.

References

Bassetti, B. (2008). Orthographic input and second language phonology. In T. Piske, & M. YoungScholten (Eds.), *Input Matters in SLA* (p. 191–206). Clevedon, UK: Multilingual Matters.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Best, C. (1995). A direct realist view of cross-language speech perception. speech perception and linguistic experience. In Strange (Ed.), *Speech perception and linguistic experience: theoretical and methodological issues* (p. 171–204). Timonium: New York Press.

Best, C., McRoberts, G., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*, 775–794.

Boersma, P. W., & D. (2022). Praat: doing phonetics by computer [Computer program].

Chao, Y. (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philosophy, Academia Sinica*, *4*, 363–397.

Chen, Y., Zhang, J., Sieg, J., & Chen, Y. (2019). Is [ɤ] in mandarin a transitional vowel? — evidence from tongue movement by ultrasound imaging. *Journal of Chinese Linguistics*, *47*, 371–405.

Cheng, C. (1973). *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.

Cheng, R. (1966). Mandarin phonological structure. *Journal of Linguistics*, *2*, 135–158.

Christoph, D., & Klaus, J. (2004). Speech recorder - a universal platform independent multi-channel audio recording software. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04*. Lisbon, Portugal: European Language Resources Association (ELRA.

Duanmu, S. (2000). *The phonology of standard Chinese*. New York: Oxford University Press.

Duanmu, S. (2007). *The Phonology of Standard Chinese*. (2nd ed.). Oxford: Oxford University Press.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*, 315–330.

Erdener, V., & Burnham, D. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, *55*, 191–228.

Fan, L. (2024). Difficulties of chinese vowel finals: A study on hungarian learners and teachers. In P. Sz. Simon, & L. A (Eds.), *15th International Conference of J. Selye University. Language and Literacy Section. Conference Proceedings* (p. 23–41). Komárno, Slovakia.

Fan, L. (submitted). Perception and production of chinese vowel finals by hungarian learners – some relevant difficulties.

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

Flege, J., & Bohn, O. (2021). The revised speech learning model (slm-r. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (p. 3–83). Cambridge: Cambridge University Press.

Fox, J., & Weisberg, S. (2019). An r companion to applied regression.

Gass, S., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An introductory course*. New York: Routledge.

Hartman, L. (1944). The segmental phonemes of the peiping dialect. *Language*, *20*, 28–42.

Juhász, K. (2020). A mandarin illabiális veláris magánhangzó [ɤ], illetve az alveoláris [ɹ] és posztalveoláris [ɻ] approximánsok produkciója kínaiul tanuló magyarok körében. *Alkalmazott nyelvtudomány*, *20*.

Kiss, G. (1985). A magyar magánhangzók első két formánsának meghatározása szintetizált hangmintákat felhasználó percepciós kísérlet segítségével. Nyelvtudomány Közlemények,.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Lado, R. (1957). Language across cultures. Ann Arbor:.

Lee, W., & Zee, E. (2001). An acoustical analysis of the vowels in beijing mandarin. In P. Dalsgaard, B. Lindberg, H. Benner, & Z.-H. Tan (Eds.), *7th European Conference on Speech Communication and Technology (EUROSPEECH 2001 Scandinavia* (p. 643–646). Aalborg: International Speech Communication Association.

Lee-Kim, S.-I. (2014). Revisiting mandarin 'apical vowels': An articulatory and acoustic study. *Journal of the International Phonetic Association*, *44*, 261–282.

Lenth, R. (2021). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, *1*, 5–1.

Liberman, A., Harris, K., Eimas, P., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech*, *4*, 175–195.

Lin, Y.-H. (1989). Autosegmental treatment of segmental processes in chinese phonology.

Lin, Y.-H. (2007). *The Sounds of Chinese with Audio CD* volume 1. Cambridge: Cambridge University Press.

Pulleyblank, E. (1984). *Middle Chinese: A study in historical phonology*. Vancouver, BC: University of British Columbia Press.

R Core Team (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Simay, A., Fan, L., & Szemle, K. (2020). A kínai nyelv magyarországi tanításának rövid története és jelene (brief historical overview and present state of chinese studies in hungary.

Strange, W. (1995). *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium, MD: York Press.

Szende, T. (1994). Hungarian. *Journal of the International Phonetic Association*, *24*, 91–94.

Tabain, M., Butcher, A., Breen, G., & Beare, R. (2020). A formant study of the alveolar versus retroflex contrast in three central australian languages: Stop, nasal, and lateral manners of articulation. *The Journal of the Acoustical Society of America*, *147*, 2745–2765.

Tusor, N. (2016). A kínaiul tanuló magyar anyanyelvűek tipikus kiejtési hibái.

Wang, Y. (2001). A preliminary investigation on the perception of high vowels in mandarin chinese by korean and japanese students. *Language teaching and linguistic studies*, *6*, 8–17.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wiese, R. (1997). Underspecification and the description of chinese vowels. In J. Wang., & N. Smith (Eds.), *Studies in Chinese phonology* (p. 219–249). Berlin: Mouton de Gruyter.

Xu, S.-R. (1980). *Phonology of Standard Chinese (*普通话语音知识*)*. Beijing: Wenzi Gaige Chubanshe.

Zhang, Y. (2003). The influence of acoustic properties on perception of apical vowels in beijing mandarin. In *In. Proceedings of the Sixth National Conference on Modern Phonetics* (p. 109–114).

Zhou, D., & Wu, J. (1963). *Putonghua fayin tupu [Articulatory diagrams of Standard Chinese*. Beijing: Shangwu yinshuguan.

Zhu, X. (2010). *Phonetics (*语音学*)*. Beijing: The Commercial Press.