University of Szeged

Doctoral School of Linguistics

Theoretical Linguistics PhD Programme

Libo Fan

Perception and production of Chinese vowel finals by Hungarian learners – Some relevant difficulties

Supervisor:

Dr. Huba Bartos

senior research fellow

2025 Szeged

Table of Content

List of Figures

List of Tables

Acknowledgement

I wish to express my sincere and profound gratitude to all those who have supported and guided me throughout my doctoral journey. Their contributions, encouragement, and understanding have been invaluable to the successful completion of this dissertation.

First and foremost, I extend my deepest appreciation to my master's degree supervisor, Professor Li Licheng, for his invaluable mentorship and encouragement. Professor Li not only provided a recommendation letter for my Ph.D. application but also offered continuous guidance and constructive advice during my doctoral studies. His insightful feedback on my thesis has been instrumental in shaping my academic development. I am equally indebted to Professor Li Yuming and Professor Gu Yandong from the Beijing Language and Culture University for their unwavering support, including writing recommendation letters for my Ph.D. application, which paved the way for my academic aspirations.

I am profoundly grateful to the University of Szeged, which provided me with the opportunity to pursue and complete my doctoral studies. I owe my deepest gratitude to my supervisor, Professor Bartos Huba, whose expertise, guidance, and encouragement have been indispensable throughout this process. His unwavering support and patience gave me the confidence to overcome academic challenges and achieve my goals. My sincere thanks also go to Professor Németh T. Enikő, the Head of the Doctoral School, for her steadfast support and assistance during moments of self-doubt and uncertainty. I am further indebted to Professor Ivaskó Lívia, leader of Theoretical Linguistics at the University of Szeged, for her constant encouragement, understanding, and professional advice, which greatly enriched my research.

I would also like to acknowledge the immense support provided by the HUN–REN Hungarian Research Centre for Linguistics (NYTK). I am especially thankful to Gráczi Tekla Etelka, who generously shared her expertise and provided critical guidance on the experimental design, data collection, and analysis of my thesis. Her mentorship has significantly enhanced my academic skills, and her rigorous approach to research continues to inspire me. I am equally grateful to Professor Polgárdi Krisztina, whose expertise in phonetics and phonology provided me with a strong theoretical foundation. Additionally, I wish to express my gratitude to Professor Mády Katalin for introducing me to R programming and to Dr. Juhász Kornélia for her kind and consistent support throughout this academic journey.

My gratitude also extends to Dr. P. Szabó Sándor and Jónás Zsolt from the Ludovika University of Public Service, whose encouragement and collaboration during my teaching tenure strengthened my resolve to pursue doctoral studies. I would also like to express my heartfelt appreciation to the ten native Hungarian teachers and ten native Chinese teachers who participated in my interviews, as well as the students who contributed to my empirical research. Although privacy concerns prevent me from naming them here, their involvement has been vital to the success of this study. I am also grateful to my Ph.D. classmate, Winn Myintzu, for her invaluable collaboration on my first English manuscript.

Last but not least, I wish to acknowledge the enduring love and support of my family. Over the course of these seven years, their unwavering encouragement has been the foundation of my strength and perseverance. My heartfelt thanks go to my husband and children, whose presence and unconditional support have provided me with the motivation to continue striving toward my goals. Their belief in me has been my greatest source of endurance and inspiration.

To all those who have contributed to this journey, I would like to express my deepest and most sincere thanks.

1. Introduction

1.1 General introduction

In the early twentieth century, the standard pronunciation of Standard Chinese was formalized and promoted by the Republic of China under the name Guóyǔ ('national language'). After the establishment of the People's Republic of China in 1949, Standard Chinese was renamed Putōnghuà ('common speech') and redefined as "the common language of China, based on the northern dialects, with the Peking phonological system as its standard pronunciation" (Norman, 1988: 135). Standard Chinese is commonly referred to as Mandarin Chinese, which is used in schools and universities and on national radio and television broadcasts. Since my investigations included in this dissertation are strictly limited to this variety, henceforth I will simply use the term 'Chinese' to refer to Standard/Mandarin Chinese.

Although Mandarin Chinese contains a wide range of speech sound categories that warrant further investigation from the perspective of L2 acquisition by Hungarian learners, this dissertation focuses exclusively on vocalic segments. Compiling the experiments of the dissertation, a multi-step approach was employed. First, a questionnaire was administered to explore students' beliefs of their difficulties with Chinese vowel finals. Following this, interviews were conducted with Chinese language teachers to gather their insights on the challenges their students face. Qualitative interviews provided a better understanding of this topic. Finally, two experimental studies were designed to examine in detail the vowel finals [ɤ], [ɿ] and [ʅ]. The perception and pronunciation of the most difficult vowels in perception research and production research are based on previous research and the Chinese teachers' interview in present research. In particular, the third study and the fourth study in present research examined the mid vowel [ɤ] and the apical vowels [ɿ] and [ʅ]. There were several reasons to analyze the perception and production of [ɤ], [ɿ] and [ʅ]. First, the Chinese [ɤ], [ɿ] and [ʅ] vowels are not part of the Hungarian vowel system, so these sounds are like a "blank slate" or "uncategorized" dimension in Hungarian. In addition, in Chinese [ɤ], [ɿ] and [ʅ] can appear in the same consonantal contexts. The second reason is that the tongue height values of [ɤ], [ɿ] and [ʅ] are very close, thus it is expected that these sounds could cause problems for non-native learners of Chinese. Third, based on the questionnaire results of the first study in present research, Hungarian learners of Chinese do not think that the Chinese vowels [ɤ], [ɿ] and [ʅ] are difficult, since all of them had a mean score which was less than 3.5, but fifteen out of twenty participating Chinese teachers in the second study of the present research mentioned that [ɤ] has different kinds of problems and eight out of twenty participating teachers mentioned that [ɿ] and [ʅ] has different kinds of problems. For example, some learners pronounce [ɤ] as [ɛ]/[e:], [ɿ] and [ʅ] as [i], some students confused [ɤ], [ɿ] and [ʅ]. Fourth, based on my experience as a Chinese language teacher with over 15 years of practice in Hungary, these specific sounds appear to be among the most problematic for Hungarian learners of Chinese. However, this observation is anecdotal and requires systematic empirical validation.

1.2 Importance and growing popularity of learning Chinese in Hungary

In recent years, China-Hungary economic and trade relations have developed rapidly. Due to the strategic geographical location, Hungary has long been an important partner in China's trade relations. Zhou Dongyao, a researcher at the Development Research Center of the State Council of China, noted that Hungary was the first European country to sign an intergovernmental cooperation agreement with China to jointly promote the construction of the "One Belt, One Road" initiative. It was also the first to establish and launch a working group mechanism for the initiative, the first country in Central and Eastern Europe to set up an RMB clearing bank, and the first in the region to issue RMB bonds. Additionally, Hungary was the first Central and Eastern European country to host an office of the China National Tourism Administration and the first European country to establish a bilingual education program for both Chinese and Hungarian speakers.[1] Hungary, as a key country in the Belt and Road Initiative and the largest recipient of Chinese investment in Central and Eastern Europe, has become a model of Sino-European investment cooperation. Chinese companies have invested over $4.5 billion in Hungary, creating more than 15,000 jobs. The country is increasingly attractive to Chinese firms such as the Bank of China, Huawei, and BYD, all of which have established branches in Hungary. In recent years—particularly since the Hungarian government introduced its investment migration policy—there has been a growing number of Chinese immigrants, students, and tourists. This influx has fueled demand in the

[1] 中华人民共和国中央人民政府: http://www.gov.cn/xinwen/2017-12/01/content_5243826.htm

Chinese-language job market, which in turn has indirectly promoted the development of Chinese language education in Hungary.[2]

Chinese teaching was introduced to Hungary in the 1950s (Józsa, 1988), and Chinese education is becoming more and more important among foreign languages in Hungarian language education. In 2011, as a possible second foreign language, Chinese was officially introduced to the Hungarian secondary school graduation examination system (Li et al., 2021). In 2016, Chinese in Hungary was promoted from an optional second foreign language to a possible first foreign language (Li, 2020).

The number of people learning Chinese language in Hungary has experienced a change from single digits to thousands over the decades (Simay & Fan, 2020). During this time, the Chinese government has been actively working to expand its soft power through a variety of initiatives, one of which is the establishment of Confucius Institutes. These institutes' primary mission is to promote the Chinese language and culture abroad. By doing so, they aim to enhance global understanding of China and foster international cultural exchange. In Hungary, to meet the growing demand for Chinese language education, Chinese educational institutions have also expanded to major cities across the country.

The country has established one Sino-Hungarian bilingual school, six Confucius Institutes, and two Confucius Classrooms. Additionally, an increasing number of local Hungarian educational institutions have introduced Chinese language programs. According to the Director of the Confucius Institute at Eötvös Loránd University, by 2020, the number of people learning Chinese language in Hungary had reached 6,259 (Li et al., 2021).

In higher education, Eötvös Loránd University (ELTE), the birthplace of Chinese language education in Hungary, named at that time as Pázmány Péter University, established the Department of Far Eastern Languages and Literature as early as 1923 (Mei, 1990). However, according to the university's official website, the department was formally established in 1924[3]. Subsequently, the university introduced its Chinese language undergraduate program in 1926, making it the first institution in Hungary to offer such a program (however the program was not continuously operated until the 1980s). Following ELTE's lead, Pázmány Péter Catholic University (PPCU) established its Department of Chinese Language in 2012.

[2]中国人民共和国商务部: http://www.mofcom.gov.cn/article/i/jyjl/m/201910/20191002902145.shtml

[3]ELTE Távol-Keleti Intézet：http://tavolkeletiintezet.elte.hu/.

According to data from Hungary's official admissions website, felvi.hu, student enrollment in Chinese undergraduate programs at both ELTE and PPCU has steadily increased between 2019 and 2023. ELTE admitted 48 students in 2019, increasing to 62 in 2023, while PPKE saw an increase from 16 admissions in 2020 to 17 in 2023. Furthermore, Károli Gáspár University of the Reformed Church in Hungary (KGU) became the third institution in Hungary to offer Chinese language programs, enrolling 12 students in 2020 and 11 in 2023.

In addition to these specialized Chinese language programs, several Hungarian universities offer Chinese as an elective or compulsory course. Notably, Ludovika University of Public Service (LUPS), with its Department of China Studies, provides not only Chinese language courses but also subjects related to China's international relations, economy, public administration, and society. Budapest Business School (BBEU) has a long-standing history of offering Chinese language courses, first introduced as electives in 1996, and also hosts a Business Chinese Proficiency Test Center to assess business Chinese proficiency (Simay & Fan, 2020). Corvinus University of Budapest (CUB), Semmelweis University (SE), and the Dharma Gate Buddhist College (DGBC), the six universities hosting Confucius Institutes in Table 1, among others in Hungary also offer Chinese language courses. By 2017, 14 out of 66 universities in Hungary offered Chinese language courses (Li, 2020). These programs not only facilitate academic exchanges between Hungary and China but also help meet the foreign language learning requirements of Hungarian students. According to Hungarian Ministry of Education regulations, university students majoring in non-language disciplines must take a foreign language course if they have not passed an intermediate-level foreign language exam before enrollment.

Chinese language education has also been introduced in Hungarian secondary and elementary schools. The Hungarian - Chinese Bilingual School started its operation in 2004, providing only elementary school education at the beginning, and in September 2016, the school expanded its scale of operation and started to provide secondary school education. The school's best feature is that Chinese is one of the bilingual languages of instruction, which is the only one of its kind among bilingual schools in Central and Eastern Europe. By 2017, 26 out of 1146 secondary schools and 8 out of 2248 elementary schools in Hungary were offering Chinese language classes, reflecting the growing interest in Chinese language acquisition at multiple educational levels (Li, 2020).

Finally, Confucius Institutes and Confucius Classrooms have played an important role in the development of Chinese language education in Hungary. The worldwide network of Confucius Institutes was founded in 2004 by the Center for Language Education and Cooperation (formerly Hanban) of the PRC in collaboration with foreign universities or educational institutions, and these institutes are non-profit educational organizations. Currently, there are six Confucius Institutes and two Confucius Classrooms in Hungary (see Table 1).

Table 1. Six Confucius Institutes (CI) and two Confucius Classrooms (CC).
(based on Fan, L - Myintzu, W., 2022; Simay, A. E.- Fan, L., 2020)

Institution name	Established time	Chinese partner	Nature of Chinese Education
ELTE Confucius Institute (CI)	2006	Beijing Foreign Studies University	academic qualifications + non-academic education
Confucius Institute at University of Szeged (CI)	2012	Shanghai International Studies University	academic qualifications + non-academic education
University of Miskolc Confucius Institute (CI)	2013	Beijing University of Chemical Technology	academic qualifications + non-academic education
Confucius Institute of Traditional Chinese Medicine at University of Pécs (CI)	2015	North China University of Science and Technology	academic qualifications + non-academic education
University of Debrecen Confucius Institute (CI)	2019	Tianjin Foreign Studies University	academic qualifications + non-academic education

Confucius Institute of Hungarian-Chinese Bilingual School (CI)	2024	Capital Normal University	academic qualifications + non-academic education
Confucius Classroom of Hungarian-Chinese Bilingual School (CC)	2017-2024	Capital Normal University	academic qualifications
Confucius Classroom of Kecskemét Bolyai János Gimnázium (CC)	2009	*ELTE Confucius Institute*	academic qualifications
Confucius Classroom of Eötvös József Gimnázium (CC)	2019	*Confucius Institute at University of Szeged*	academic qualifications

The Confucius Institute of Hungarian - Chinese Bilingual School was opened in collaboration between the Hungarian - Chinese Bilingual Primary School and the Capital Normal University in 2024. This institution was a Confucius Classroom from 2017 to 2014. Two Confucius Classrooms listed in Table 1 do not have a Chinese partner; instead, they collaborate with the ELTE Confucius Institute and the Confucius Institute at the University of Szeged, respectively. As comprehensive educational and cultural exchange platforms focusing on Chinese language teaching and cultural dissemination, Confucius Institutes and Confucius Classrooms are primarily staffed by government-funded Chinese language teachers and volunteers, who can select teaching materials based on the characteristics and needs of their students. The Confucius Institute not only offer Chinese courses for cooperating university students, but also for participants outside of university system.

At present, in Hungary, apart from the six Confucius Institutes mentioned above, the number of other language training schools offering Chinese courses in Hungary has also increased, most of which are located in Budapest. These include the Oriental Language School (Because of pandemic, this school is closed now.), Language Solutions (this school only offers online Chinese training for companies), Youle Mandarin, and others. The Oriental Language School (Keleti Nyelvek Iskolája) is the largest and oldest local Chinese language training institution in Hungary offering Chinese courses since 1995. The school employs full-time local Chinese teachers, with the number of teachers fluctuating slightly each year based on student enrollment. And among the other educational institutions mentioned above, the Hungarian-Chinese Bilingual Primary School, ELTE, PPCU, KGU employ both native Chinese teachers and native Hungarian teachers of Chinese. In contrast, some other universities like BCE and LUPS do not have any native Chinese teachers (at the time of writing this). In the six Confucius Institutes and the two Confucius classrooms in Table 1, the teaching staff consists primarily of native Chinese teachers. As a result, some Hungarian learners of Chinese are taught exclusively by native Chinese teachers, others solely by Hungarian teachers of Chinese, while some benefit from instruction by both native Chinese teachers as well as Hungarian teachers of Chinese.

1.3 Motivation and aim of the dissertation (contributing to Chinese as L2 teaching)

Based on the present state of Chinese studies in Hungary, research on Chinese phonetics and phonology is becoming increasingly important. Understanding the perception and production of Chinese vowel finals by Hungarian learners presents a unique opportunity to bridge the gap between theoretical phonetic/phonological models and practical second-language acquisition challenges. Chinese and Hungarian have different phonetic and phonological systems, the mid vowel [ɤ] and the apical vowels [ɿ] & [ʅ] which are absent from Hungarian sound system are highly challenging for native Hungarian speakers. These challenges highlight critical areas in linguistic research and pedagogy that have been underexplored. Until now there have been very few production studies on the difficulties of Chinese vowel finals. There have been no studies on the difficulties of Chinese vowel finals from the perspective of Chinese teachers, neither have there been studies on perception from the perspective of Hungarian learners of Chinese.

As Chinese language education continues to expand across diverse educational settings, the presence of both native Chinese teachers and Hungarian teachers of Chinese underscores the importance of exploring how different instructional backgrounds shape learner outcomes. This dissertation considers the role of teacher background in influencing Chinese learning. Rather than implying that only native speakers are qualified to teach pronunciation, this study advocates the complementary strengths of both teacher types. Hungarian instructors often bring deep cultural insight, shared learning experiences, and a strong capacity for empathizing with students' difficulties. They may also be better equipped to explain phonetic features using the learners' first language. Meanwhile, native Chinese teachers provide authentic input and intuitive command of pronunciation patterns. The collaboration between native and non-native instructors can therefore create a richer learning environment. Importantly, the focus on intelligibility and comprehensibility—not native-like pronunciation—means that effective teaching hinges not solely on nativeness, but on pedagogical clarity and learner engagement.

ELTE and ELTE Confucius Institute have published a series of beginner and lower intermediate level textbooks specially written for Hungarian learners (back in the early 2010s)[4]. However, the availability of adequate and specialized teaching materials tailored to Hungarian learners remains a significant challenge. Moreover, the growing number of Chinese degree programs and language courses at universities across Hungary underscores the importance of developing comprehensive research that informs curriculum design, teacher training, and instructional materials.

From a theoretical standpoint, this dissertation contributes to the broader understanding of how second-language learners navigate new phonetic categories that are absent in their first language. By integrating perception and production studies within the frameworks of different perception models and different production models, the present study aims to test each of the models to see which one makes the most correct predictions and offers insights into how Hungarian learners perceive and produce new Chinese vowel sounds. It also provides a valuable perspective on the interaction of markedness and L1 influence in shaping the acquisition process.

From a practical perspective, the increasing importance of Chinese in Hungary – both as a field of academic study and as a tool for global communication – underscores the need to address the specific challenges Hungarian learners of Chinese face. Despite the growing number of Chinese programs and Hungarian learners of Chinese, little research has systematically analyzed the difficulties learners encounter with Chinese vowel finals. This dissertation aims to fill that gap. By integrating both perception and production issues, the present research aims to expand our understanding from a linguistic perspective, particularly in the context of teaching effectiveness.

This dissertation not only explores difficulties of the learners but also incorporates the perspectives of Chinese language instructors, offering a comprehensive view of pedagogical

[4] JIANG Wenyan, HAMAR Imre, YE Qiuyue (szerk.): Kínai nyelvkönyv magyaroknak 1.-4. kötet. 2011, 2012, 2013, 2015. ELTE Konfuciusz Intézet, Budapest

challenges. The findings can inform targeted interventions, such as integrating perception and production training to improve learning outcomes. For instance, understanding why vowels such as [ɤ], [ɿ], and [ʅ] are particularly challenging for Hungarian learners allows educators to design exercises that mitigate these issues.

Finally, by examining both beginner and intermediate learners taught by native Chinese or/and native Hungarian instructors, the study provides insights into how input quality may influence learner outcomes. However, these outcomes are also likely to be shaped by individual differences in aptitude, which play a well-documented role in pronunciation acquisition. The results can contribute to the development of linguistic theories on L2 acquisition, and improve teaching methodologies, and enhance learner success, making this research highly relevant for both academic and pedagogical contexts.

These studies provide crucial insights into the challenges Hungarian learners of Chinese face and the underlying mechanisms shaping their acquisition of Chinese phonetics. Section 7 synthesizes the findings, offering a comprehensive discussion that integrates theoretical considerations, experimental results, and pedagogical implications. This progression ensures a cohesive narrative, linking theory, empirical data, and application in a systematic manner. By integrating findings across all studies, this dissertation provides a comprehensive analysis of the perceptual and productive challenges the Hungarian learners of Chinese may encounter. It emphasizes the importance of addressing these challenges, highlighting the influence of linguistic, context, markedness, and individual factors on second language acquisition.

1.4 Structural introduction

This dissertation explores the use of Chinese vowels among Hungarian learners of Chinese, with a focus on the perception and production of the mid vowel [ɤ] and the high vowels [ɿ] and [ʅ]. The research integrates theoretical, experimental, and pedagogical dimensions, providing a comprehensive investigation of these challenges. The first three sections establish the theoretical foundation and contextual framework for the study.

The introduction outlines the relevance of the study by highlighting the phonetic and phonological differences between Chinese and Hungarian sound systems, particularly in terms of the vowels under investigation. It underscores the importance of this research in addressing gaps in second language acquisition studies, specifically concerning Hungarian learners of

Chinese. This section also connects the theoretical objectives of the study with practical implications for teaching and learning Chinese in Hungary.

Section 2 provides a detailed analysis of the Chinese sound system, with a focus on syllable structure and the phonetic and phonological properties of Chinese vowels. It compares these features with Hungarian phonology to identify potential sources of difficulty for Hungarian learners. Special attention is given to [ɤ], [ɿ] and [ʅ], emphasizing their markedness and absence in the Hungarian vowel system. The section also examines the patterns of orthography on phonological processing in both languages, establishing groundwork for understanding the challenges in perception and production.

The third section reviews relevant second language acquisition (SLA) theories and models, including the Speech Learning Model (SLM), the Perceptual Assimilation Model (PAM), and the Markedness Differential Hypothesis (MDH). These frameworks offer insight into the acquisition of non-native vowels, especially in cases where learners' L1 (Hungarian) lacks equivalent phonetic categories. In addition, the review discusses the potential influence of input quality, orthographic representation, and individual learner differences on L2 sound acquisition. While the current study does not explicitly isolate or control these variables, the results may offer indirect insight into their roles. By situating the findings within this theoretical landscape, the study aims to contribute to a more nuanced understanding of the factors shaping Hungarian learners' perception and production of Chinese vowels.

The subsequent sections build theoretical and contextual groundwork established earlier. Sections 4 and 5 present four empirical studies designed to explore Hungarian learners' perception and production of Chinese vowels.

The first study uses a questionnaire to examine how Hungarian learners of Chinese perceive the difficulty of pronouncing Chinese vowel finals. To investigate this, a questionnaire with a seven-point Likert scale was administered (see Appendix A). The key variables included learners' perceived difficulty ratings of 21 Chinese vowel finals, the frequency with which each vowel final was reported as difficult, and the grouping of these finals based on learner perceptions, as identified through factor analysis. The factor structure was supported by statistical tests, including Bartlett's test of sphericity, the Kaiser-Meyer-Olkin (KMO) measure, and the explained variance.

The second study gathers insights from Chinese teachers through interviews. In the study, several variables were included to investigate teachers' perspectives on Hungarian learners' pronunciation of Chinese vowel finals. Teacher-related variables encompassed their teaching experience in Hungary, proficiency in Hungarian, educational background, methods used to teach Pinyin, teacher beliefs toward error correction as well as their beliefs about the sources of learners' pronunciation problems and the impact of co-teaching with Hungarian or Chinese instructors on students' perception and production of sounds And learner-related variables focused on specific vowel finals that learners struggled with, and other influencing factors were considered, such as the students' academic background and knowledge of other languages.

The third study examines the identification of Chinese vowels [ɤ], [ɿ] and [ʅ] by Hungarian learners of Chinese with different levels of Chinese learning experience and different types of Chinese teacher: native speaker (mother tongue: Chinese), L2-speaker (mother tongue: Hungarian), and a combination of two teachers of both speaker types. Thus, two main variables are: 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 of the results. The aim was to investigate the perceptual patterns of these sounds and, within this scope, the identification of the mid from the high vowels at different learning stages. Meanwhile it is also aimed to explore the potential sources of the observed patterns. A total of 31 participants were divided into two main groups: (i) 21 beginners who have learned Chinese for one semester and (ii) 10 intermediate learners who have learned Chinese for five semesters. The beginner group was further divided into three sub-groups: (i/a) one of the groups is taught exclusively by native Chinese speakers, (i/b) another group taught solely by Hungarian L2 speakers of Chinese, and (i/c) a third group is taught by both a native Chinese and a Hungarian L2-speaker of Chinese. Using an X(AB) identification test, we investigate the identification of Chinese vowels [ɤ] and [ɿ]/[ʅ] among Hungarian learners of Chinese.

The fourth study examines the production of Chinese vowels [ɤ], [ɿ] and [ʅ] by Hungarian learners of Chinese. Previous empirical research explored the production of these vowels by professional learners of Chinese, and the present study extends this work by examining non - professional learners of Chinese, focusing on vowels' differentiation of various L2 learner levels and different group types. In a grapheme-based (Pinyin and characters) L2 word-reading task, the production of [ɤ], [ɿ], and [ʅ] by 30 Hungarian learners of Chinese was compared with that of native Chinese speakers. The experiment involved native Mandarin speakers and Hungarian learners at varying proficiency levels, analyzing vowel productions through formant measurements and Linear Discriminant Analysis. The formant values were normalised into Z-scores by speaker. Linear Discriminant Analysis models were trained on the 6 native speakers' data using all three formants.

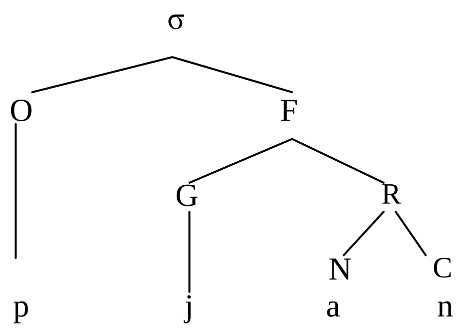
2. The Chinese sound system

Chinese is typically classified as an analytic or isolating language in which each morpheme is usually also a word (Lin, 2007, p.4). In modern Chinese, 95 percent of morphemes are monosyllabic (Chen, 1999, p.138-9). There is almost a one-to-one relationship between morphemes and syllables, and it is sometimes said that conditions for syllable structure and for morpheme structure are hard to distinguish (Battistella, 1987). Syllable as the basic phonological unit is always uncontroversial in traditional Chinese phonological analyses.

2.1 Chinese syllables

When discussing syllables, a set of concepts is always mentioned[5]. Onset (O) is the beginning of a syllable, and Coda (C) is the end of a syllable, and they must be consonantal. The Nucleus (N) of a syllable is at the syllable core as its sonority peak, while nucleus and coda together are called Rhyme. A syllable tree is always used to represent Chinese syllables. In traditional Chinese phonological/phonetic analysis, the concepts of initial and final are used. The initial is the beginning non-glide consonant of a syllable, which is equal to Onset in Figure 1. And the final is the rest of the syllable after the initial consonant. Vowels are associated with nuclear position, and glides are associated with non-onset and non-nuclear positions. The medial glide precedes the nucleus, and a rhyme can be filled by a single nuclear vowel with an optional coda which can be a consonant or a glide.

Figure 1. The analysis of the Chinese syllable
(Cheng, 1966; Lin, 1989, 1990. cf: Van de Weijer & Zhang, 2008)



(σ is syllable, O is onset, F is final, G is glide, R is rhyme, N is Nucleus, C is coda)

[5] In this section we temporarily ignore tones, focusing on the melodic content, and return to the question of tones at the end of the section.

In this research I retain the syllable as a convenient way of referring to the combination of an initial and a final. The reason to use the initial and the final like this is that Chinese textbooks for non-native speakers also follow this model to introduce the Chinese sound system. In Table 2, we can see that only nucleus V is compulsory in a Chinese syllable. Prevocalic glide is limited to [j], [w] and [ɥ] and postvocalic glide is limited to [j] or [w].

Syllable boundaries are clear in Chinese, and the possible inventory of syllable types is small. In Table 2, V refers to vowel, G refers to glide, and C refers to consonant. I have made one minor adjustment in the symbols: Lin used GVV, CVV and CGVV - I have changed it to GVG, CVG and CGVG respectively. The smallest syllable of Chinese is merely a V, and the largest one is CGVG or CGVN. In traditional Chinese phonetic or phonological analysis, the initial and the final are regarded as the main constituents of syllables.

Table 2. The syllable structure inventory[6] of Chinese (disregarding tone)
(Examples in the table are from Lin 2007, p. 107)

V	[ɤ]<e> ‘hungary’	GV	[ja]<ya> ‘duck’	CGV	[two] <duo> ‘many’	CGVG	[xwaj] <huai> ‘bad’
		VG	[aj]<ai> ‘love’	GVG	[jɑw] <yao> ‘medicine’		
		CV	[ma] <ma> ‘horse’	GVC	[jɛn] <yan> ‘salt’		
		VC	[an] <an> ‘peace,safe’	CVG	[laj] <lai> ‘come’	CGVC	[xwan] <huan> ‘to exchange’
				CVC	[lan] <lan> ‘blue’		

The symbols ‘<’ and ‘>’ will be used to denote orthographic forms, / / represents phonemes and [ ] represents allophones. ‘ ’ is used for English translations.

[6] Chinese also has a so-called retroflex vowel *er* [ɚ] which occurs in two cases. One is in words without a suffix like *er*^{51}[ɚ]^{HL} 'two' and *er*^{35}[ɚ]^{LH} 'son' (Duanmu, 2007-40). In this case, [ɚ] is not allowed to spell with any initial consonants and the number of words using *er*[ɚ] is very limited in number. In fact, it is still unclear if *er* consists of a single vowel, a syllabic consonant, a diphthong, or a vowel plus a consonant (Lin, 2007-80). The other case is that it can be combined with a syllable before it, forming a retroflex syllable (*erhua* syllables), for example [paɚ] "handle" is combined by [pa] and [ɚ]. When we pronounce [ɚ], we retract our tongue body backward (Lee & Zee 2003). The suffix [ɚ] merge with the finals it attaches to, and it does not expand the syllable structure.

The orthographic transcription in present research is from Chinese language textbook (Jiang, 2013).

2.1.1 Initials

To look at the Chinese initial consonants, in the Chinese language textbooks for non-native Chinese language learners, it is mentioned that there are altogether 21 consonant sounds at the initial position. Chinese obstruents (stops and affricates) exhibit a two-way contrast between voiceless unaspirated and voiceless aspirated segments. In Table 3, those voiceless unaspirated (e.g. [p]) are on the left and those voiceless aspirated (e.g. [p^h]) are on the right side. In Chinese, there are three pairs of unaspirated and aspirated stops [p, p^h] <b, p>, [t, t^h] <d, t>, and [k, k^h] <g, k>, and three pairs of unaspirated and aspirated affricates [ts, ts^h] <z, c>, [tɕ, tɕ^h] <j, q>, and [tʂ, tʂ^h] <zh, ch>. Fricatives [f, s, ɕ, ʂ, x] <f, s, x, sh, h> do not have aspirated/unaspirated pairs. [tɕ], [tɕ^h], and [ɕ] are in complementary distribution with the dental affricates/fricatives [ts], [ts^h], and [s], retroflex (post-alveolar) affricates/fricatives [tʂ], [tʂ^h], and [ʂ^h], and velars [k^h], and [x] (Duanmu, 2007; Lin, 2007; Zhu, 2010). In Chinese, the voiced nasals [m, n, ŋ] and the liquids [l, ɻ] are represented by graphemes <m, n, ng> and <l, r> respectively. [ɻ] is transcribed as central approximant [ɹ] by Lin (2007, p. 41) and voiced fricative [ʐ] by Duanmu (2007, p. 24).

Table 3. Consonants of Chinese (Zhu, 2010: 332)

vl=voiceless; vd =voiced

		labials		dental		alveolar-palatal		retroflex		velar	
		vl	vd	vl	vd	vl	vd	vl	vd	vl	vd
obstruent	stop	p p^h		t t^h				tʂ tʂ^h		k k^h	
	affricate			ts ts^h		tɕ tɕ^h					
	fricative	f		s		ɕ		ʂ	*ʐ	x	
sonorant	nasal		m		n						ŋ
	liquid				l				ɻ		

The present research follows Zhu's (2010) analysis. Zhu (2010: 332) states that phonetically [ɻ] is not a voiced fricative [ʐ] in the first place, in fact, even if it is realized as a voiced fricative [ʐ] in some cases or on some occasions, from a “systemic” point of view, treating it as a phonetical approximation of [ɻ], which is the versatility of phonemic processing.

2.1.2 Finals

The focus of the present research will be on segmental rather than suprasegmental units, particularly on vowels. Since the time of the "classic example of classical phonemics" of Hartman (1944), extensive studies have been carried out to the phonetic and phonology of Mandarin (Hashimoto, 1970), and the vowel inventory of Mandarin vowel phonemes seemed to be generally accepted as /a, ə, i, u, y/ (e.g. Duanmu, 2007, p. 35; Lin, 2007, p. 82), at least as the basis for further discussions on controversial details.

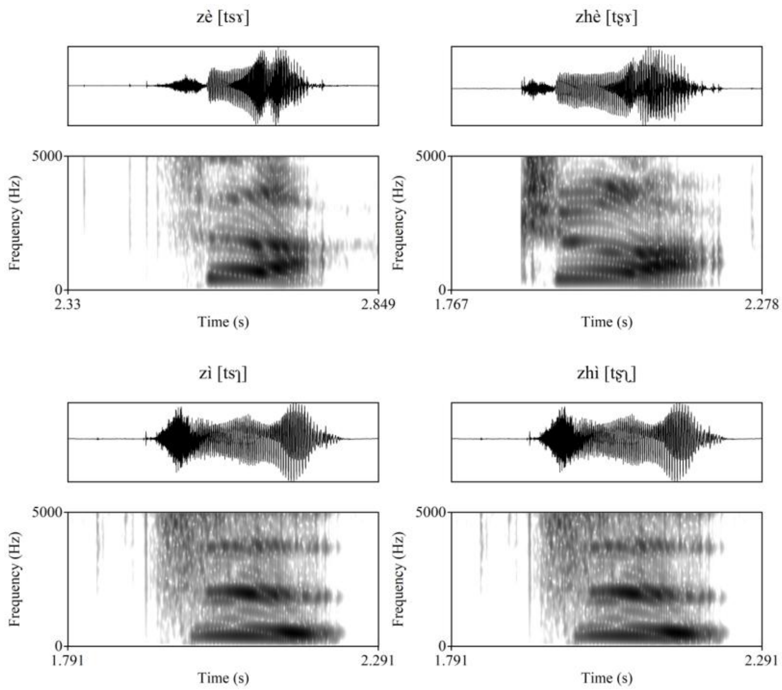
The most difficult problem in any appropriate description of the Chinese vowel system lies in the treatment of allophones of these phonemes. There is a jumbled system of phonetic realizations of Chinese vowel phonemes. The allophonic variations of each phoneme will be introduced in the following order, beginning with high vowels and progressing to low vowels.

As for high vowels, three distinct phonemes /i/, /u/ and /y/ have three corresponding semi-vowels (glides) respectively. There is an alternative way to represent three glides, [j], [w] and [ɥ] can also be written as [i], [u] and [y], respectively. There are debates about variants of /i/, which has been argued to have three contextual allophones [ɿ], [ʅ] and [i] by some scholars.

Since the perception and production experiments in the present research focus on the two high vowels [ɿ] and [ʅ] speech sounds in Chinese, we describe these in more detail. We refer to the /i/-allophones [ɿ], [ʅ] in questions as vowels in accordance with a part of the phonetic and phonological studies (Karlgren, 1915–1926; Chao, 1928; R. Cheng, 1966; C. Cheng, 1973; Howie, 1976; Zhou & Wu, 1963; Zhu, 2010; Hu, 2024). [ɿ] & [ʅ] vowels are denoted by two different IPA-symbols and thus be considered as two segments having different place of articulation. Based on X-ray images taken by Zhou & Wu (1963), the tongue tip/blade gesture of [ɿ] and [ʅ] inherited from the preceding dental and retroflex consonants remains nearly unchanged for the following voiced period. The articulation of these segments is homorganic with the preceding sibilants. The two apical vowels are still in debate. Based on frication noise and formant values, some scholars regard them as consonants, e.g. syllabic fricatives [z] and [ʐ] (Chao, 1934, 1968; Hartman, 1944; Pulleyblank, 1984; Lin, 1989; Wiese, 1997; Duanmu, 2000, 2007. cf: Lee-Kim, 2014), syllabic approximants [ɹ̩] and [ɻ̩] (Lee & Zee, 2001) and approximants [ɹ] and [ɻ] (Lee-Kim, 2014). Lee-Kim (2014) states that the syllabic fricative approach is not quite right because of the absence or at least the tenuous presence of frication noise during the voiced segments. However, some scholars treat them as vowels. R. Cheng (1966) assumed that they are derived from the underlying high front vowel /i/. Since [ɿ] and [ʅ] are in complementary distribution with the high front vowel [i]: [ɿ] only occurs after dentals [ts, tsh, s], [ʅ] only occurs after retroflexes [tʂ, tʂh, ʂ, ɻ], and [i] occurs in other environments. Cheng (1966) assumes that the two apical segments are derived from the underlying high front vowel [i] whereas C. Cheng (1973), who noticed the retracted tongue body posture, posits an abstract underlying segment [ɨ], independent from [i]. Hu (2024) argues that if apical vowels are regarded as syllabic consonants, then it implies that both consonants and vowels can function as the nucleus of a Chinese syllable. However, in the few Chinese dialects where syllabic consonants exist, these are nasals, laterals, and rhotics. Also, these typically appear in weak forms and are generally considered marginal cases. In contrast, apical vowels are not marginal in Chinese dialects by any means. Therefore, both the vowel-based and consonant-based analyses are supported by evidence. Even though the categorization of these two sounds is still questionable and debatable, as a convenient way, from the point of view of the present research, let us regard them as allophones of /i/ that are not apparent in Hungarian. [ɿ], [ʅ] and [i] are treated as contextual allophones in clear-cut complementary distribution. In Chinese, it is indeed important to produce these allophones [i], [ɿ] and [ʅ] correctly because of the similarity of the preceding consonants, e.g. the potential confusion of ‹si› [sɿ] with ‹xi› [ɕi], or ‹zhi› [tʂʅ] with ‹ji›[tɕi], if /i/ is pronounced as [i] might lead to misperception (Bartos Huba, p.c.).

The mid vowel and the high vowels in question are shown in Figure 2.

Figure 2. Oscillograms and spectrograms of [ɤ], [ɿ] and [ʅ]



For comparison of contextual effects, the mid vowel is included in both contexts (non-retroflex and retroflex), even though the influence of the preceding consonant's retroflex feature is considered weaker in the mid vowel. This influence does not result in distinct speech sounds, unlike in the high vowels, where the vowel retains the retroflex feature's value throughout its duration.

The possible variants of the mid vowel /ə/ have been proposed to be the following (Table 4): [o], [ə], [ɤ], [e], [ɛ̝] (Xu, 1980, p. 33).

Table 4. variants of the mid vowel /ə/ in Chinese

Variants	Examples	Environment
[ɤ]	[kɤ] <ge>'song'	In open syllables, not after a labial or a palatal
[e]	[kej] <gei>' give'	Before [j]
[ə]	[mən] <men>'door'	Before [w], [n] or [ŋ]
[o]	[wo] <wo> 'I'	In open syllables, after labials
[ɛ̝]	[jɛ̝] <ye>'grandpa'	In open syllables, after palatals

However, most scholars do not distinguish [ɛ̝] and [e] (Chao, 1968; Cheng, 1973; Lee & Zee, 2003; Zhu, 2010, p. 307), and they write [e] for both cases. Second, the allophone [ə] before [w] is generally transcribed as [o] (e.g. Duanmu, 2007, p. 55; Lin, 2007, p. 71). Since perception and production in the present research focus on the mid vowel [ɤ] and the high vowel [ʅ] and [ʅ̩] speech sounds, we describe [ɤ] in more detail. According to Wang (2018), it is a diphthong [ɤ$^{\Lambda}$] based on formant movement analysis, but most of the scholars state it is a monophthong (e.g. Duanmu, 2007, p. 37; Lin, 2007, p. 73; Chen et.al., 2019). Juhász (2020) also mentioned this in her production research. However, since phonemically there is no reason to treat it as a diphthong, it is treated as a monophthong in present research. [ɤ] is a mid, back, unrounded vowel and is regarded as a contextual allophone of mid central vowel /ə/ in final position of (C)V syllables. [ɤ] occurs not only after [ts, ts^h, s] and [tʂ, $tʂ^h$, ʂ, ɻ̩], but also some other consonants like [k, k^h, x, t, t^h, n, l]. Duanmu (2007) argues that in open syllables after labial consonants [p, p^h, m, f], [ɤ] and [wo] are free variants, which is not included in the present thesis.

The possible variants of the low vowel have also been proposed (Table 5) to be [a], [ɑ], [ʌ], [ɐ], [æ] (Xu, 1980, p. 33).

Table 5. variants of the low vowel /a/ in Chinese

Variants	gloss	Environment

[a]	[k^{h}aj] <kuai> ‘open’	In closed syllables, before [n] or [j], and not after a palatal
[ɑ]	[lɑw] <lao> ‘old’	In closed syllables, before [w] and [ŋ]
[ʌ]	[pʌ] <ba> ‘eight’	In open syllables
[æ]	[jæn] <yan> ‘salt’	In closed syllables, before [n] and after [j]
[ɐ]	[ɥɐn] <yuan> ‘round’	In closed syllables, before [n] and after [ɥ]

[ʌ] and [ɐ] are not shared by some studies (e.g. Lee & Zee, 2003; Zhu, 2010, p. 307; Lin, 2007, p. 78), these two variants are written as [a]. Following Lin (2007, p. 78), we may assume that there are three variants “[ɑ], [ɛ] and [a]” for the low vowel phoneme. Lin states that [æ] and [ɛ] are free variants for some speakers in case of the final <-ian> [jɛn], [a] and [ɛ] are free variants for some speakers in case of the final <-üan> [ɥɛn].

In addition to the vowels mentioned above, in Chinese there is a so-called retroflex vowel final [ɚ], which has been variably transcribed as [əɹ], [ɹ], or [ɚ] (Lin, 2007, p. 80) and [əɚ] (Lee & Zee, 2001). Lee & Zee (2001) state that it has been proved that the formant trajectories for this so-called retroflex vowel in the V syllables is a sequence of [ə] and [ɚ]. [əɚ] is not allowed to spell with any consonant initials, such as [əɚ] ‘son’. But it can be combined with a syllable before it, forming a retroflex syllable, for example [paɚ] ‘handle’ is combined by [pa] and [əɚ]. [əɚ] is not included in vowel finals analyzed in the present study.

In addition to retroflex vowel, 21 Chinese vowel finals V, GV, VG and GVG are shown in Table 6.

Table 6. Chinese vowel finals

[a]<a>	[ja]<l**ia**>	[wɑ] <h**ua**>	
[aj] <ai>		[waj] <h**uai**>	
[ɑw] <ao>	[jɑw] <l**iao**>		
[ɤ] <e>	[je] <l**ie**>	[wo] <l**uo**><b**o**>	[ɥe] <x**ue**>
[ej] <ei>		[wej] <h**ui**>	
[ow] <ou>	[jow] <l**i(o)u**>		
[ɿ] <i> ,[ʅ] <i>	[i]<l**i**>	[u] <l**u**>	[y] <l**ü**>

Duanmu (2007, p. 55) argues that there is one more vowel final [jaj] ‘cliff’ in Chinese, which is a literary word and it is sometimes pronounced as [ja] or [aj]. But this final is not recognized by other scholars (e.g. Lin 2007, p. 288; Lee & Zee, 2003). The present study does not include [jaj]. Furthermore, when the Chinese phonetic alphabet (Pinyin) was formulated and published, [o] after labial consonants was treated as the abbreviation of [uo] (Lin & Wang, 2013, p. 233). Thus, the actual pronunciation of <bo> is [buo].

In the case of diphthongs in Table 6, tongue height features in diphthong finals always change, and sometimes labial roundness also changes, for example diphthongs in Table 7. From Table 7, we can see that Chinese diphthongs are agreed to [back], but tongue height and labial roundness can be changed.

Table 7. Chinese diphthongs

	[VG] $_\sigma$			[GV] $_\sigma$			
height	[-h] [+h]			[+h] [-h]			
backness	[-b] [-b]	[+b] [+b]		[-b] [-b]	[+b] [+b]	[-b] [-b]	[+b] [+b]
roundness	[-r] [-r]	[-r] [+r]	[+r][+r]	[-r] [-r]	[+r] [-r]		[+r] [+r]
diphthongs	[aj]$_\sigma$ [ej]$_\sigma$	[ɑw]$_\sigma$	[ow]$_\sigma$	[ja]$_\sigma$ [je]$_\sigma$	[wɑ]$_\sigma$	[ɥa]$_\sigma$	[wo]$_\sigma$

Notes: r=round; b=back; h=high

In case of triphthongs, not only tongue height and labial roundness always need to change, tongue backness as well, as shown in Table 8.

Table 8. Chinese triphthongs

	[GVG] $_\sigma$		
height	[+h] [-h] [+h]		
backness	[+b] [-b] [-b]	[-b] [+b] [+b]	
roundness	[+r] [-r] [-r]	[-r] [-r] [+r]	[-r] [+r] [+r]
triphthongs	[waj]$_\sigma$[wej]$_\sigma$	[jɑw]$_\sigma$	[jow]$_\sigma$

Notes: r=round; b=back; h=high

We can see that nucleus and coda agree in the feature [back]. Medial glide position and coda position are limited to [+high], but they are not allowed to contain the same feature [back] and [round]. Thus, in the Chinese vowel system, there is a regular nonoccurrence of certain types of finals like /iai, uao, jəi, uəu, yai, yau, yəi, yəu/, they are all interpretable as the result of dissimilation of the medial glide and the ending vowels. OCP (Obligatory Contour Principle) was applied to account for the absence of these triphthongs by Wiese (1997). There can be no adjacent identical feature specifications for at least the so-called melodic features, which is proposed by Leben (1973). Based on this assumption, it turns out that a violation of the OCP occurs not only for those triphthongs with two identical vowels, but also triphthongs including /y/, since /y/ is marked underlyingly both for labiality and for front position. Then if one excludes high vowels in Table 7, we can see the almost complete complementary distribution of sounds with respect to finals, the clear parallelism of occurrence between finals with the low vowel phoneme /a/ and those with the mid vowel phoneme /ə/. The only exception is <ya>, and its nonoccurrence is due to historical change (Hashimoto, 1970).

To look at the Chinese finals ending in a nasal, in the Chinese language textbooks (Jiang, 2013) for non-native Chinese language learners, it is mentioned that there are altogether 16 finals ending in a nasal at the final position (Table 9).

Table 9. Chinese finals with nasal codas

		Finals without intials	Finals with intials
1	VN	<an>[an] ‘to press’	<lan>[lan] ‘blue’
2		<en>[ən] ‘eh’	<ben>[phən] ‘to spray’
3		<ang>[ɑŋ] ‘to hold head high’	<lang>[lɑŋ] ‘wolf’
4		<eng>[əŋ] ‘reins’	<leng>[ləŋ] ‘cold’
5	(C)GVN	<yan> [jɛn] ‘salt’	<lian>[ljɛn] ‘face’
6		<yang>[jɑŋ] ‘sheep’	<niang>[ljɑŋ] ‘to measure’
7		<yin>[in] ‘cloudy’	<lin>[lin] ‘forest’
8		<ying> [jəŋ] ‘hard’	<ling>[ljəŋ] ‘spirit’
9		<wan>[wan] ‘ten thousand’	<luan> [lwan] ‘chaos’
10		<wang>[wɑŋ] ‘king’	<guang>[kwɑŋ] ‘light’
11		<wen> [wən] ‘to ask’	<lun>[lwən] ‘wheel’
12		<weng>[wəŋ] ‘old man’	
13			<zhong>[tʂuŋ] ‘middle’
14		<yuan>[ɥɛn] ‘far’	<quan> [tɕhɥɛn] ‘all’
15		<yun>[ɥyn] ‘cloud’	<jun> [tɕhɥyn] ‘group’
16		<yong> [juŋ] ‘to use’	<qiong>[tɕhjuŋ] ‘poor’

C is consonant, G is glide, V is vowel, N is nasal

The IPA transcription in Table 9 followed Lin (2007: 283-292). We can see that [ən, əŋ, an, ɑŋ] are beginning with non-high vowels [ə], [a] and [ɑ], where [ə] is a central vowel, [a] is a front vowel and [ɑ] is a back vowel. [in], [iəŋ], [jɛn], [jɑŋ] and [juŋ] are beginning with high front vowel [i] or glide [j], where [i] and [j] are in complementary distributions. [wən], [wəŋ], [uŋ], [wan], [wɑŋ] are beginning with high back vowel [u] or glide [w], where [u] or glide [w] are in complementary distributions. [ɥɛn] and [ɥyn] are beginning with high front glide [ɥ], where glide [ɥ] and [y] are also in complementary distributions. [j, w, ɥ] only occur in non-nuclear position, but [i, u, y] only occur in nuclear position.

Mandarin Chinese finals with nasal codas are limited by some phonological constraints. First medial glide position is limited to glides [j, w, ɥ], coda position nasals are limited to [n, ŋ]. Second, if the nuclear position is filled by a low vowel, then [ɥɑŋ] is ill-formed, but [ɥan] is well-formed. The nonoccurrence of [ɥɑŋ] is due to historical reasons (Hashimoto, 1970). Third [uŋ] and [wəŋ] are in complementary distribution in Chinese, [uŋ] could not be used alone, in other words, [uŋ] must spell with consonant initials, but [wəŋ] is not allowed to spell with any consonant initials. Fourth, nasal finals beginning with high front vowel glide [ɥ] only occur after consonant initials [tɕ], [tɕʰ], [ɕ].

2.1.3 Chinese tones

Chinese is a typical lexical tone language: the pattern of pitch change over the span of a syllable can distinguish lexical meanings. Chinese makes the best use of pitch patterns together with the segmental features (vowels and consonants) to build words (Wang, 1973). The perception and production of Mandarin vowels by L2 learners is different from the perception and production of non-tonal vowels by L2 learners. L2 learners must learn how to match pitch with meaning at the syllabic level. In Table 10, we can see that Chinese has four lexical tones including a high-level tone (tone 1), a mid-rising tone (tone 2), a low-falling-rising tone (tone 3) and a falling tone (tone 4) (cf: Zhang et.al, 2016).

Table 10. Chinese lexical tones

Tone number	Pitch pattern	Pitch value	Example
1	High level	55	[ma]55 mā妈 ‘mother’
2	High rising	35	[ma]35 má麻 ‘hemp’
3	Low falling-rising	214	[ma]214 mǎ马 ‘horse’
4	falling	51	[ma]51 mà骂 ‘to scold’

These four words, which share the same segmental parts: ‘妈ma^{55}’ (mother), ‘ma^{35}’ (麻hemp), ‘ma^{214}’ (马horse), ‘ma^{51}’ (骂scold) have different lexical meanings (Wang, 1973; Peng, 2006). It is well established that the main difference among the four Mandarin tones lies in pitch contour i.e., values of fundamental frequency (F0). The pitch value is used for the transcription of tone, which represents relative pitch level. 1 indicates the lowest pitch and 5 indicates the highest pitch on a scale of 1 to 5 (Lin, 2007, p. 90).

2.2 Comparison between the Chinese and Hungarian sound systems

As this study focuses exclusively on the acquisition of Chinese, a comprehensive introduction to Standard Hungarian (hereafter: Hungarian) sound system is beyond the scope of this research. Instead, the Hungarian sound system will be presented in comparison to the Chinese sound system, with an emphasis placed solely on the aspects directly relevant to the current investigation.

2.2.1 Comparison between Chinese and Hungarian syllables

Compared to Chinese, the possible inventory of syllable types in Hungarian (Table 11) is relatively bigger.

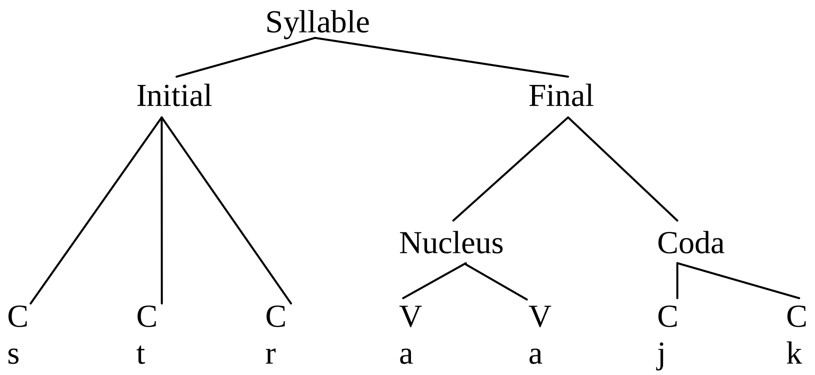
Table 11. Syllable inventory of Hungarian
(The bold graphemes indicate a syllable, and a dot marks the syllable boundary.)

V te.**a** 'tea'	CV ma 'today'	CVC nap 'day'	CVVC kár 'damage'	CVVCC bánd 'repent-imp.def_obj.2sg'	CCCVCC spricc 'squirt'
	VC **em**.ber 'man'	CVV ló 'horse'	VVCC ónt 'tin-acc'	CCVVC prém 'fur'	CCCVVCC sztrájk 'strike'
	VV rá.di.**ó** 'radio'	VVC ágy 'bed'	CVCC rakj 'put-imp.2sg'	CCVCC tromf 'trump'	
		CCV **pla**.kát 'poster'	CCVV **tré**.fa 'joke'	CVCCC karszt 'karst'	
		VCC ott 'there'	CCVC sznob 'snob'	CCCVV **strá**.zsa 'őr'	
			CCCV **stre**.ssz 'stress'	CCCVC **spric**.cel 'spray'	

There is also a sound [j] in Hungarian, which Vago (1980) argues is a glide semivowel, while Siptár & Törkenczy (2000, p.16-17) argue that *j* is a consonant liquid. First, *jV*-initial words select the preconsonantal allomorph of the definite article *a,* not its prevocalic allomorph *az,* e.g. *a játék* 'the toy', **az játék*. Second, the initial consonant of the suffix -*val* 'with' fully assimilates to stem final consonants but appears as [v] after vowel-final stems (e.g. *lábbal* 'with foot', **lábval*, vs. *szóval* 'with a word', **szóal*). Therefore, the fact that e.g. 'with butter' is *vajjal* rather than **vajval* suggests that *j* is a consonant. And finally, the example *vajjal* also shows that intervocalic *j* can be geminated (long): this in itself is enough to render any kind of diphthong interpretation impossible. Based on these facts, the present research follows Siptár & Törkenczy (2000), and it assumes that Hungarian has no glides.

Using traditional Chinese analysis, *sztrájk* 'strike' in Hungarian can be represented by this syllable tree (Figure 3). There is no medial glide in a Hungarian syllable, but this is present in Chinese, and there are consonant clusters in syllable initial position and syllable final position in Hungarian but not in Chinese. Third, length is a distinctive in Hungarian, but this is not the case in Chinese. Fourth, there are diphthongs and triphthongs in Chinese, but they are not present in Hungarian.

Figure 3. Hungarian syllable structure



There are other proposals on the word initial and word medial consonant clusters CCC in Table 11. Siptár & Törkenczy (2000, p.103) argue that there is no complex onset in Hungarian, these consonants are edge clusters. Since Chinese is the focus of the present study, the syllable structure is analyzed based on traditional Chinese analysis.

Chinese imposes certain restrictions on the way segments are distributed in a syllable. For example, /y/ only occurs after /l, n/, and (the alveolo-palatal allophones of underlying) velars, and /i/ does not occur after /f/ (although this may possibly be just an accidental gap, since /i/ does occur with bilabial onsets). While these do not seem to be overarching, they do constitute some justification for syllable as a genuine constituent (Bartos Huba, p.c.).

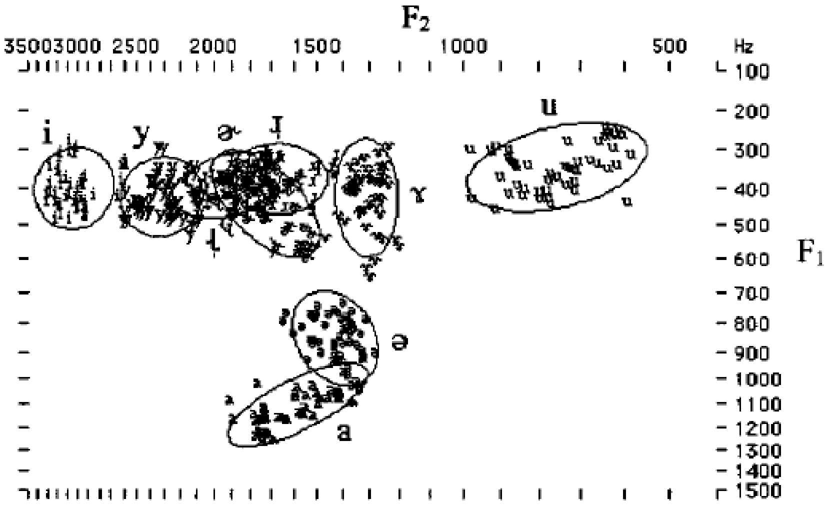
In contrast, phonologically there are no special restrictions between an initial consonant and the following vowel in Hungarian, generally a single initial consonant is not sensitive to the nature of the vowel that occurs in the following position, an initial position almost displays a maximal inventory of contrasts. Thus, syllables like [ki, tsi, sy, fy, ty] are all well-formed in Hungarian. Siptár & Törkenczy (2000, p. 9) state that the 'Principle of Free Cooccurrence' (Kaye, 1995) appears to be true in Hungarian phonology, according to which no restrictions are expected to apply between onset and rhyme, between the nucleus and the coda or within each (sub)constituent.

2.2.2 Comparison between Chinese and Hungarian vowels

According to Ladefoged & Disner (2012), a lower first formant (F1) indicates a higher tongue position, while a lower second formant (F2) reflects a more retracted (back) tongue position—though much of the F2 lowering also results from increased lip rounding. Additionally, r-colored sounds are characterized by a very low third formant (F3), which is strongly influenced by lip position. However, these relations do not mean one can draw broad conclusions on the articulation from the formant values itself (Steven 1968). For example, in many cases, native speakers have strong and mistaken intuition about the relative importance of the two distinctions, e.g., speakers of English believe that "thin" versus "sin" is a place of articulation difference, even though the loudness difference is more perceptible. Stevens et al. proposed that such redundant features evolve as an enhancement of an otherwise weak acoustic distinction, in order to improve the robustness of the language's phonological system (Stevens & Keyser, 1989; Stevens et al., 1986).
Figure 4 shows the Chinese vowel inventories based on previous research, merely in an acoustic sense. We can see that [ɿ] and [ʅ] are represented as [ɹ̝] and [ɻ̍] by Lee & Zee (2001) respectively. Lee & Zee (2001) analysed 9 vowels [a, ɤ, ə, ɚ, i, u, y, ɿ, ʅ], which were made by 20 native Chinese speakers (10 males and 10 females), reading a randomized list of 30 meaningful test monosyllables, and the syllables containing the three vowels [ɤ, ɿ, ʅ] are [xɤ],

[sɿ] and [ʂʅ] respectively with tone 1. The speakers were all college students, aged between 18 and 22.

Figure 4. F1 plotted against F2 for Chinese vowels.



Lee & Zee, 2001: Figure for 10 females

Based on the figure above, [back] plays an important role in distinguishing these three vowels [ɤ], [ɿ] and [ʅ].

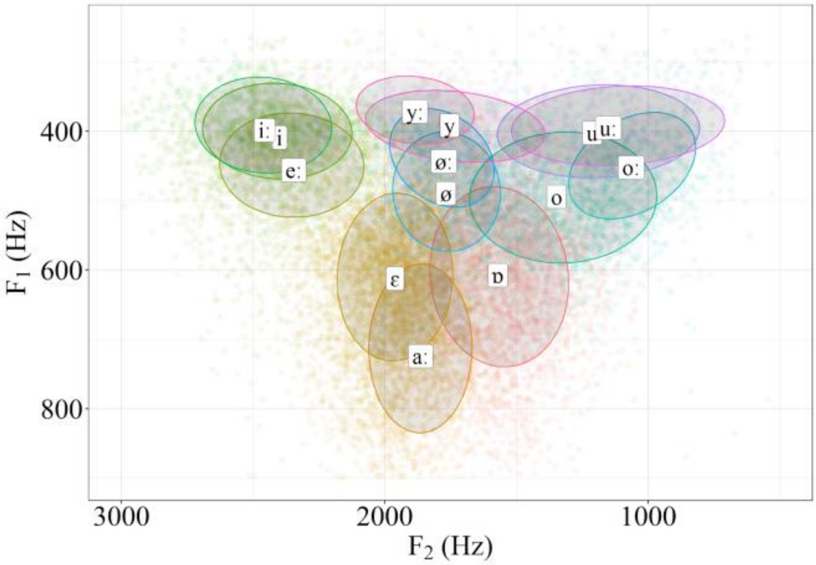
There are 14 vowels in Hungarian, which are /i, iː, u, uː, y, yː, ø, øː, ɑ, aː, o, oː, ɛ, eː/ (Markó, 2017). [ɑ] and [aː] are often written as [ɒ] and [ɑː] respectively.

i [i]	*e* [ɛ]	*ü* [y]	*ö* [ø]	*u* [u]	*a* [ɒ]	*o* [o]
í [iː]	*é* [eː]	*ű* [yː]	*ő* [øː]	*ú* [uː]	*á* [ɑː]	*ó* [oː]

Hungarian not only distinguishes vowel height, backness, and roundness, but also length (Siptár & Törkenczy, 2000). Hungarian vowels are distinguished by length which is different from Chinese. With respect to the specific Chinese segments in question, none of [ɤ, ɿ, ʅ] are part of the vowel system either in a phonemic or phonetic sense of Hungarian.

Figure 5 shows the Hungarian vowel inventories based on previous research, merely in an acoustic sense (Deme 2025). The vowel space is drawn from the first two formants in the spontaneous and read speech of 10 native, female speakers of Hungarian from a speech database called BEA (Neuberger et al., 2014).

Figure 5. F1 plotted against F2 for Hungarian vowels (Deme, 2025).



Although the different speech types in the two figures, the relation among the vowels in the vowels spaces of the two languages are clearly shown. We can see the distribution of Hungarian vowels and Chinese vowels in Figure 4 and Figure 5 shows both the vowel formants and an approximation of the tongue positions, the formants are plotted in a way that approximately follows the articulations. However, we also have to note that the connection between the articulation and acoustics is not linear and these formant values vary along the body and vocal tract size (Stevens, 1989).

According to Figure 4 and Figure 5, the Chinese target vowels of the present study appear in the mid high, mid central region of the acoustic vowel space, where the [y, y:] and [ø, ø :] vowels appear in Hungarian. Hungarian has no mid vowel [ɤ] and apical vowels [ɿ], [ʅ], and formant 2 values for these three vowels are different, but formant 1 values are very close. We have to emphasize, that this comparison is just an estimation, as the formant measurement methods are not compared in detail. Therefore, we can only assume that possibly, these Hungarian vowels may interact in the perception and production of the Chinese vowels in question in the Hungarian language learners.

Second, there are no diphthongs and triphthongs in Hungarian, but Hungarian is rich in hiatus. Hiatus occurs only when the two vowels belong to different syllables, each forming the nucleus of its own syllable. In contrast, Chinese diphthongs are contained within a single syllable. Based on the analysis of Siptár & Törkenczy (2000, p. 123-124), we can find vowel sequences in Table 12:

Table 12. Hungarian hiatus (Siptár & Törkenczy, 2000, p. 123-124)

	i	iː	y	yː	u	uː	ɔ	o	oː	ø	øː	eː	ɑː	ɛ
i					*iu*	*iú*	*ia*	*io*	*ió*	*iö*	*iő*	*ié*	*iá*	*ie*
y							*üa*					*üé*		*üe*
u	*ui*	*uí*			*uu*		*ua*	*uo*	*uó*				*uá*	*ue*
ø					*öu*									
o	*oi*						*oa*	*oo*				*oé*	*oá*	*oe*
ɔ	*ai*	*aí*			*au*				*aó*					*ae*
ɛ	*ei*				*eu*		*ea*	*eo*	*eó*				*eá*	
ɑː	*ái*							*áo*						

Hungarian vowel sequences are not constrained by [+high][-high] or [-high][+high], because [+high][+high] like *ui*, *iu* and [-low][low] like *eá* are all well-formed. Furthermore, Hungarian is rich in round vowel sequences like [yɔ], [oɔ], [uo], [uɔ] etc. (examples are from Siptár & Törkenczy, 2000, p. 123-124). But this kind of round vowel cluster *ou* in Chinese is absent in Hungarian.

Ye & Bartos (2017) did a phonetic comparison between Chinese and Hungarian vowel sequences, and they distinguish two kinds of vowels: continuous vowel and compound vowel (Table 13).

Table 13. Vowel sequences of Hungarian and Chinese (Ye & Bartos, 2017)
Lower case "v" is used to represent glides.

Name	Tongue-position feature	Type	Loudness/Strength
continuous vowel (Hungarian)	fixed	VV	equal
compound vowel (Chinese)	slide	Vv	not equal
		vV	
		vVv	

Table 13 mainly shows the phonetic differences between Chinese and Hungarian vowel sequences. Ye (2013) states that a compound vowel is a part of a syllable and cannot be divided, and a continuous vowel occurs when two vowels appear side by side in a sequence with their own characteristics. The duration of the two sides is the same, they can be divided into two syllables. In other words, a continuous vowel means a hiatus, a compound vowel means a diphthong or a triphthong in one single syllable.

Third, both Chinese and Hungarian have Backness Harmony and round Harmony, but the application domain is completely different. For Chinese rhymes, Duanmu (2007, p. 60) proposed a constraint.

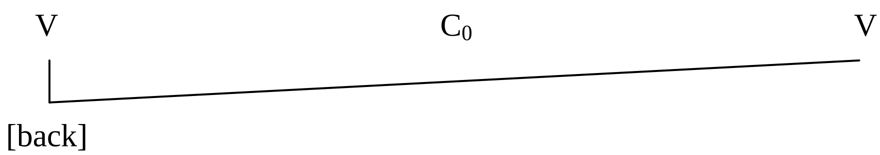
Rhyme Harmony: Nucleus and Coda cannot have opposite values in [round] or [back].
*[+back][-back], *[-back][+back],
*[+round][-round], *[-round][+round]

Backness Harmony and Rounding Harmony are two related but non-identical harmony constraints, applying to different vowels. [front] and [back] are harmonically active in Chinese, when [front] or [back] appear in the right-hand-side position of a rhyme in Chinese, they can trigger place assimilation and extend to the left position, for example [aj], [ɑw], [ej] and [ow]. [round] can also trigger place assimilation and extend to the left position if the nucleus is filled by a mid-vowel, for example [ow].

In Chinese, Backness Harmony and Round Harmony also apply to GV ([ja], [je], [wɑ], [wo], [ɥe]) in Chinese. GV must agree in [back].

Backness Harmony and Round Harmony are also attested in Hungarian, but the application domain is completely different. These two rules are constraints applying to non-adjacent segments belonging to different syllables. A vowel can assimilate to a vowel in another syllable, Marlett (2001) proposed two rules:

(1) Backness Harmony: A vowel assimilates in backness to the vowel in the immediately preceding syllable.



(2) Round Harmony: Mid vowels assimilate in rounding to a front vowel in the immediately preceding syllable.

For example (Siptár & Törkenczy, 2000, p. 72):

tűz-höz ‘fire’ *sofőr-höz ‘driver’*

víz-hez *'water'*   *kódex*-hez *'codex'*

ház-hoz *'house '*   *nüansz*-hoz *'nuance'*

Vowel harmony in Hungarian is from left to right, harmonizing suffixes are controlled by stem.

2.2.3 Chinese and Hungarian orthographies

L2 orthography-induced pronunciations may be part of the acoustic input for instructed learners, and L2 orthographic input interacts with the acoustic input, influencing L2 learners' mental representations of L2 phonology (Bassetti, 2008). Learners' non-target like phonological representations in turn result in non-target like realizations of phonemes, syllables and words. Such orthography-induced pronunciations do not exist in the native speakers' speech L2 learners are exposed to and cannot be attributed to the influence of learners' L1 phonology or to universals of phonological acquisition (Bassetti, 2008).

For instructed learners, orthography-induced pronunciations may even become part of the acoustic input (Bassetti, 2008). Bassetti reported the omitted pronunciation of specific Chinese diphthongs and triphthongs by Italian final-year (third-year) university students of Chinese. Bassetti compared L2 learners' pronunciations of the same diphthongs and triphthongs in syllables spelled with all vowels and syllables spelled omitting one vowel. Results showed that learners often omitted from the pronunciation the vowel that is omitted from the standard Romanized orthographic representation of Mandarin (pinyin). For instance, learners pronounced [iou] correctly in the syllable /iou/ (spelled as <you>), but pronounced it as *[iu] in /liou/ (spelled as <liu>). No omissions took place in diphthongs and triphthongs that are always spelled consistently, such as /iɑu/ (which is always spelled with three letters).

Hungarian and Chinese orthographies differ extensively in the association between the written form, sound, and meaning for some sounds, which may affect how native speakers encode phonological information of a word in memory. The Hungarian writing system uses the Roman alphabet to represent phonemes, so it can be classified as a phonemic orthography. In addition, there are many regularities for vowels in the grapheme-to-phoneme correspondences in Hungarian. The Hungarian orthography is not entirely transparent, but the grapheme-segment relationship is exclusive in most cases. As a reviewer points out, the the graphemes in most cases correspond to exclusively one segment (or phoneme), but there are some exceptions, e.g. regressive voicng assimilation in fűZFa, when the /z/ becomes voiceless due to the subsequent voiceless Consonant, haNg, where the /n/ becomes velar instead of the standard alveolar realization. As a result, Hungarian is considered as a transparent orthography in general. Therefore, Hungarian speakers may rely more heavily on orthographic information in word learning, identification, and reading.

Chinese has a distinct written form compared to Hungarian in that Chinese uses characters whereas Hungarian uses letters to form words. In addition, Chinese also employs the Romanization system (*Hanyu pinyin* - Chinese sound spelling or simply) to indicate the pronunciation of Chinese characters. The Pinyin system adopts graphemes of the Roman alphabet, which is the standard transcription system of Chinese words. Both characters and *Hanyu pinyin* have been taught in education inside and outside China. For second language learners of Chinese, when they start to learn Chinese, Pinyin system is dominating.

Compared to Pinyin, the Hungarian writing system is orthographically highly transparent. For example, in Chinese the grapheme <i> or <e> can denote different segments, but in Hungarian <i> and <e> only represent [i] and [ɛ] respectively (Figure 6 and Figure 7)

Figure 6. *Comparison between Chinese and Hungarian <i>*

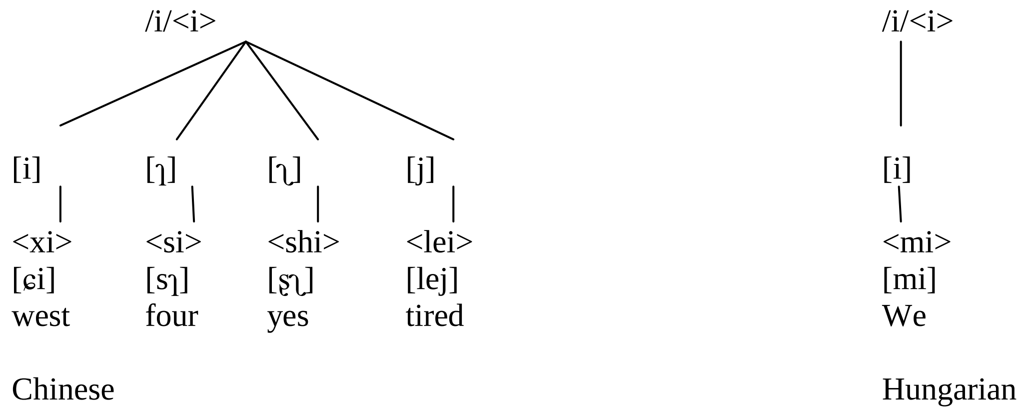
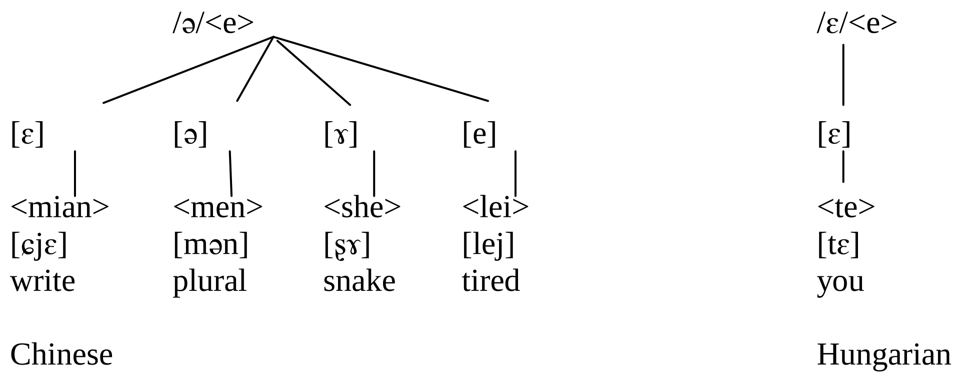


Figure 7. *Comparison between Chinese and Hungarian <e>*



Therefore, compared to Hungarian, one grapheme can represent several sounds, even one grapheme only represents only one phoneme in Figure 6 and Figure 7. Hungarian contains more transparent orthographies in which phoneme to grapheme correspondences are more consistent, and one grapheme only represents one sound. And in the present perception and production studies in Chinese the grapheme <i> denotes two different segments [ɿ] and [ʅ], in terms of orthography these apical vowels are differentiated from the mid back vowel [ɤ] because it is denoted by a different grapheme <e>.

Above all in section 2, we can see that the Chinese sound system is characterized by its analytic and monosyllabic structure, where most morphemes correspond to single syllables composed of an initial consonant and a final. Chinese syllables follow a strict CGVG or CGVN structure, with 21 initial consonants featuring contrasts between voiceless unaspirated and aspirated stops/affricates (e.g. [p] vs. [p^h]), while fricatives lack voicing distinctions. The vowel system centers on /a, ə, i, u, y/, with notable allophonic variations, including the controversial apical vowels [ɿ] and [ʅ] (sometimes analyzed as syllabic consonants) and the mid vowel /ə/, realized as [ɤ], [e], or [o] depending on context. Backness and rounding harmony are applied within syllables. Additionally, four lexical tones distinguish meaning. Compared to Hungarian, Chinese has a more restricted syllable inventory and no consonant clusters. Hungarian, features long/short vowel distinctions, hiatus (vowel sequences across syllables), and left-to-right vowel harmony, unlike Chinese's syllable-internal harmony. Orthographically, Chinese Pinyin is less transparent than Hungarian, with graphemes like <i> and <e> representing multiple sounds. These structural differences highlight key challenges for Hungarian learners of Chinese, vowel allophony, and syllable constraints.

Given these linguistic differences between Chinese and Hungarian, understanding how learners acquire a second language (L2) is crucial for explaining the challenges they face. In this dissertation, the focus is on Hungarian learners acquiring Chinese in a foreign language context (Hungary), where the target language is not widely used. The following section explores key seconde language acquisation theories that help explain how Hungarian learners of Chinese process and internalize the unique phonological and orthographic features of Chinese.

3. Theories on second language acquisition

Second language acquisition (SLA) refers to the learning of a nonnative language after the learning of one’s native or primary language. And the term ’second language (L2)’ is used broadly to refer to any language acquired after the first language (L1), regardless of whether it is the second, third, fourth, or even fifth language learned (Gass et al., 2013). Although many different languages are learned as an L2 by Hungarian learners of Chinese, the focus is on the acquisition of Chinese in the present study. Furthermore, the distinction between acquiring an L2 in natural communicative settings and learning it in more structured environments generally emerges when comparing second language acquisition in regions where the target language is dominant with foreign language acquisition in regions where the target language is not commonly used. In this study, I use the general term SLA to encompass learning in both second language and foreign language contexts. The term ’native language’ (NL) refers to the first language acquired by a child, often called the primary language, mother tongue, or first language (L1). Meanwhile, ’target language’ (TL) denotes the language being learned. As for Hungarian learners of Chinese in the present study, their L1 or NL is Hungarian, and their L2 or TL is Chinese. They are studying Chinese in Hungary where the target language is not widely used.

In the following sections, let us consider the main models of L2 learning, with a particular emphasis on perception and pronunciation, to examine their potential effects on both speech perception and production.

3.1 The Contrastive Analysis and the Markedness Differential Hypothesis

The native language (L1) significantly influences how second language learners perceive the speech sounds of the target language, particularly for individuals who begin acquiring the L2 later in life (Barrientos, 2023). Adults consistently encounter difficulties in perceiving and producing various foreign phonetic segments (Strange, 1995). One of the most prominent and well-defined assertions about the influence of an L2 learner’s native language on pronunciation errors emerged in the mid - 20th century. This claim first was developed within the framework of the Contrastive Analysis Hypothesis (CAH) proposed by Lado (1957). The Contrastive Analysis Hypothesis posited that differences between the native language (L1) and the target language (L2), along with the influence of L1 transfer, played a central role in shaping L2 production.

Transfer, as a broad concept, refers to the impact of similarities and differences between the target language and any previously acquired language (Cook, 2016). When elements from the first language align with those in the second language, this results in positive transfer, often occurring unnoticed. Conversely, when discrepancies arise, leading to visible challenges in the second language, it is referred to as negative transfer (Selinker, 1983). Thus, interference from the L1 was considered the main phonological cause of a foreign accent.

Within the framework of the Contrastive Analysis Hypothesis, Lado explicitly asserted that individuals are inclined to transfer the forms, meanings, and distributions of forms and meanings from their native language and culture to the foreign language and culture. Elements that align closely with the native language are expected to be easier to learn, while those that differ significantly will present greater difficulty (Lado, 1957, p. 2). It is assumed that learners transfer the structural patterns of their native language to the foreign language they are learning, which serves as the primary factor influencing the ease or difficulty of acquiring the foreign language's structure. According to Lado (1957), structures that closely resemble those of the native language are easier to learn, as the transfer process allows them to function effectively in the foreign language. Conversely, structures that differ significantly from the native language are more challenging, as their transfer often results in errors, requiring learners to make adjustments for successful use in the foreign language (Lado, 1957, p. 59). Stockwell and Bowen (1965) further developed and enhanced the predictive framework of the Contrastive Analysis Hypothesis. They introduced an eight-level hierarchy of difficulty, assigning the greatest phonological challenge to situations where a learner must acquire a target language allophone that does not exist in their native language. Several studies have investigated the production of L2 vowels by adults who began learning their second language during childhood, late adolescence, or early adulthood. These studies have consistently demonstrated that the vowel system of the first language often exerts a noticeable influence on the production of L2 vowels, particularly during the initial stages of learning (Major, 1987; Munro, 1993).

However, the Contrastive Analysis Hypothesis has faced criticism from many researchers for its inability to accurately predict all errors (Gass et al., 2013). The CAH does not account for L2 errors that are not the result of interference or L1 transfer, which led to the development of a new approach called error analysis. Error analysis produced several studies in which patterns of language acquisition could be clearly linked to knowledge of both the target language and the native language (Schumann, 1979). It is important to note, however, that understanding error rates requires considering more than just native and target languages.

The Markedness Differential Hypothesis (MDH) makes further efforts to take into account markedness differentials, which are used to predict and explain the order or rate of L2 acquisition (Eckman, 1977). MDH sought to revise the Contrastive Analysis Hypothesis by introducing additional principles. In the Markedness Hypothesis Framework, the emphasis is not on denying the importance of transfer, but rather on identifying the principles that underline its use. MDH aims to resolve the problems of the Contrastive Analysis Hypothesis by suggesting that although differences between the native language and the target language are crucial for explaining challenges in second language acquisition, they alone do not offer a comprehensive explanation. Under the framework of MDH, an understanding of how second language learners learn a new phonological system must consider linguistic differences between the NL and the TL systems, as well as universal facts of phonology. MDH was grounded in a phonological theory of markedness, advancing from the least marked (easiest) language types to the most marked (most difficult).

Eckman (1977) argues that the areas of difficulty a second language learner will encounter can be predicted by comparing the native language (L1) with the target language. Aspects of the TL that differ from the native language and are relatively more marked will be more challenging. The level of difficulty associated with these aspects of the TL corresponds to the degree of markedness they possess compared to the NL. Thus, the MDH assumes that L2 learners can apply or transfer those structures that have been already acquired in the native language to the target language, and the level of difficulty will correspond to the relative degree of markedness of that structure. An unmarked phonological form is more frequent and considered easier to acquire than a marked one. Markedness follows a progression from the least marked (easiest) to the most marked (most difficult). This systematic approach underscores how typological markedness can serve as a framework for predicting and understanding challenges in second language acquisition. Thus, a native speaker of Hungarian, which has no marked sounds [ɤ], [ɿ] and [ʅ] in any position, may have more difficulty learning them than other sounds.

[ɤ] is a high-mid unrounded back vowel. Most back vowels are rounded, and back unrounded vowels are relatively uncommon in the world's languages (Ladefoged & Maddieson, 1996). So [ɤ] is regarded as a marked sound (typologically unusual), in relation to its rounded counterpart of unmarked [o]. In the present study, only the marked member is tested – since [o] is reported to not cause problems by both Hungarian learners of Chinese and their teachers based on the first and the second study in present research, and Hungarian has the [o] sound as well, which is phonetically very similar to Chinese [o].

[ɿ] is frequently equated with the unrounded high central vowel, [ɨ], its unmarked counterpart is [ʉ]. Central vowels are normally rounded (Ladefoged & Maddieson, 1996), so the unrounded version counts as marked. There is no regular IPA vowel symbol corresponding to [ʅ]. In standard IPA, [ʅ] corresponds approximately to [ɨ] with a retroflex articulation. And the sound can be represented by [ɨ˞] which is a valid IPA transcription using diacritics to indicate that the vowel [ɨ] is retroflexed. The unmarked counterparts of [ɿ] and [ʅ] are not present in Mandarin Chinese. Following the long-standing tradition of studies on the Mandarin sound system, especially from an educational perspective, we will use the non-standard phonetic symbols, [ɿ] and [ʅ], for the two apical vowels under scrutiny (Bartos Huba, p.c.).

3.2 L2 Perception and production models

In addition to the hypotheses previously mentioned, in the study of second language acquisition, many models have been developed to predict and explain the processes of speech perception and production. This diversity underscores the intricate nature of L2 speech processing, and the complexity involved in both perception and production. Notably, there is a greater emphasis on models addressing speech perception compared to those focused on production. Although the models discussed here are not exhaustive, they represent some of the most frequently referenced frameworks in the field of L2 sound acquisition, offering valuable insights into these processes.

3.2.1 L2-Speech perception models

The study of second language (L2) speech perception has been informed by many theoretical models. These frameworks provide valuable insights into how learners perceive speech sounds, accounting for the complex interplay of linguistic experience, cognitive processes, and phonological systems. A central theme of these models is the role of linguistic experience, which shapes their proposals and predictions. Specifically, these frameworks focus on how L1 speech sounds influence and compare to those in the target language, providing insights into the mechanisms underlying L2 sound perception.

In the field of phonetics, several influential models have been developed to explain L2 sound perception. Notable among these are Kuhl's Native Language Magnet (NLM) model (Kuhl, 2000), the Speech Learning Model (SLM; Flege, 1995, 2003), and its revised version (SLM-r, Flege & Bohn, 2021). The Native Language Magnet model primarily seeks to explain the transition from general auditory processing to language-specific perceptual processing. Based on data gathered through functional MRI techniques, the NLM suggests that infants' mapping of the ambient language distorts the acoustic dimensions underlying speech, creating a complex network or filter through which language is perceived. This language-specific filter alters the acoustic dimensions we attend to, stretching or shrinking them to emphasize the differences between language categories. Once established, these filters make it significantly harder to learn a second language, as the perceptual mapping suited to one's native language is often incompatible with the requirements of other languages. Therefore, certain phonetic distinctions, both in perception and production, can be particularly challenging for second language learners (Kuhl, 2000). The NLM model suggests that constraints on L2 speech sound perception stem from prior experience, rather than from a loss of neural plasticity due to normal maturation (Kuhl, 2000). The influence of prior experience on the perception of L2 speech sounds is also supported by the Natural Referent Vowel framework (Polka & Bohn, 2011). According to this framework, language experience builds upon the default patterns established by initial vowel perception biases. As learners are exposed to a specific language, they expand their vowel category repertoire and reorganize their vowel perception to optimize their native language perception. As for the other two models, the Speech Learning Model (SLM; Flege, 1995, 2003) and its revised version (Flege & Bohn, 2021) are widely recognized in research on speech production. Given their significant role in production area, these models will be discussed in detail in Section 3.2.2, which focuses on speech production models.

In comparison to phonetic field, phonological perception models generally focus on the challenges learners face in distinguishing between sound contrasts. For instance, the phonological approach to L2 contrasts is exemplified by Brown's Phonological Interference Model (Brown, 1998, 2000), Polka and Bohn's Natural Referent Vowel framework (NRV;

Polka & Bohn, 2011), and Escudero's Second Language Linguistic Perception Model (L2LP; Escudero, 2005; Escudero et al., 2009). These models contribute significantly to our understanding of how L2 learners perceive and process phonological contrasts in their target language. The Phonological Interference Model seeks to explain the source of L1 phonology's influence on the acquisition of L2 segments by proposing that an L2 learner transfers the feature geometry of their L1 into their interlanguage (cf: Escudero, 2005). Selinker (1972) defined interlanguage as a distinct linguistic system exhibited by adult second-language learners when they attempt to convey meaning in a language they are actively learning. Within the non-linear phonological framework of Phonological Interference Model, the representation of a particular sound segment in a given language is viewed as a subset of the universal feature geometry (Brown, 1998).

The conceptual framework of NRV is focused on understanding the development of vowel quality perception. And the NRV framework integrates principles of linguistic typology with its unique approach to speech perception, connecting with other models in the field. At its core, the NRV hypothesis suggests that peripheral vowels are perceptually prioritized by young infants, representing an innate bias independent of language-specific phonemic status. Cross-language studies with adults demonstrate that this natural vowel bias is shaped by language experience, evolving as listeners adapt to their linguistic environment. Adult data indicate that while this bias persists in mature listeners, it can be overridden when necessary to optimize the perception of functional vowel contrasts unique to their language. Furthermore, research using natural speech within the NRV framework reveals that directional asymmetries in vowel perception emerge when differences in vowel quality are substantial enough to cross a phonetic boundary (Polka & Bohn, 2011).

The Second Language Linguistic Perception model seeks to describe and explain how sound perception and its acquisition occur by mapping the speech signal onto phonological representations (Escudero, 2005). According to the L2LP model, L1 transfer plays a central role in this process, leading to the cognitive formation of an L2 perception system that mirrors or duplicates the learner's L1 perception system. Additionally, the L2LP model aims to outline the principles governing the development of L2 phonological systems, providing a comprehensive framework for understanding the interplay between perception and phonological learning in a second language (Escudero, 2005).

Best's Perceptual Assimilation Model (PAM; Best, 1995) and its extension for second language learning (PAM-L2; Best & Tyler, 2007) integrate principles from both phonetics and phonology. The two models incorporate insights from a direct realist approach and focus on the interplay between language-specific phonetic patterns and underlying phonological systems. They also explore how developmental changes shape the perception of non-native speech sounds and contrasts, providing a framework for understanding cross-language speech perception (Best & Tyler, 2007).

The core principle of PAM in cross-language speech perception is that non-native speech segments are interpreted based on their resemblance to and divergence from native segmental patterns most similar to them within the listener's native phonological space. This similarity is evaluated through the spatial configuration of the vocal tract and the dynamic properties of articulatory gestures, which serve as the dimensions for assessing likeness (Best, 1995). The relative ease or difficulty in perceiving a specific contrast depends on the listener's native language. These variations arise from the perceived similarities or differences between the phonetic features of non-native sounds and the native phonological system. PAM categorizes L2 contrasts into three types: i) Two-Category: Clear differentiation between the two sounds as belonging to distinct categories. ii) Category Goodness: Both sounds are assigned to the same category but differ in how closely they align with the prototypical features of that category. iii) Single Category: Both sounds are perceived as belonging to the same category, leading to poor discrimination because they are equally distinct from, or similar to, a single native category (Best, 1995). PAM-L2 predicts that a learner's speech perception undergoes continuous refinement with extended exposure to both native (L1) and non-native (L2) languages. This model introduces the concept of a phonological space, where native categories are positioned based on similarities in their articulatory and phonetic structures. According to PAM-L2 (Best & Tyler, 2007), listeners may assimilate L1 and L2 sounds as functionally equivalent at the phonological level. However, this phonological assimilation does not necessarily mean that the corresponding phones are perceived as identical at the phonetic level. Differences in L2 phonetic category formation may arise due to contrasts at the phonological level. For instance, consider an English speaker learning French as an L2 (Best & Tyler, 2007). This learner may develop distinct phonetic categories for /p/ and /b/ in both languages, as these sounds contrast phonologically in English and French. However, there is a mismatch in how stop voicing contrasts are realized phonetically in the two languages. In French, the short-lag unaspirated [p] (associated with /p/) overlaps phonetically with the English phoneme /b/, which is often realized as a short-lag unaspirated [p] in word-initial positions. This overlap may lead to dissimilation between French [p] and English [p], as well as between the interlanguage phonemes /p/ and /b/. Such mismatches could result in a global reorganization of related L1 and L2 phonetic categories, possibly extending to dimensions beyond primary voice onset time differences, such as glottal phasing (Best & Tyler, 2007). This example illustrates how PAM-L2 accounts for the complexity of L2 phonetic category differentiation, emphasizing the dynamic interaction between phonological and phonetic properties during second language acquisition.

All in all, these models collectively provide a comprehensive understanding of L2 speech perception. They emphasize the interaction of linguistic experience, innate biases, and the adaptation of phonological and phonetic systems, offering valuable insights into the challenges of second language acquisition.

3.2.2 L2-speech production models

When examining speech production models, phonetic approaches are more prominent than phonological ones. The Speech Learning Model (SLM; Flege, 1995, 2003) and its revised version (SLM-r; Flege & Bohn, 2021) are among the most frequently referenced models for speech production. The Speech Learning Model (SLM) focuses on the impact of linguistic experience on the acquisition and refinement of L2 phonetic systems, making them central to discussions about L2 speech production. The SLM primarily investigates the phonetic similarities and dissimilarities between a speaker's first language (L1) and second language (L2), particularly focusing on how these factors influence the rate and order of language acquisition in both production and perception. One of the key concepts in SLM is 'equivalence classification.' According to Flege (1995, 2003), sounds in the L2 that are perceived as 'equivalent' or 'similar' to those in the L1 are more challenging for learners to acquire. This is because speakers tend to classify or perceive these sounds as equivalent to those in their L1. In contrast, sounds that are 'new' or dissimilar to those in the L1 are easier to acquire, as the learner is more likely to notice the differences. Flege (1995) proposed that the phonetic category an L2 learner forms for an L2 sound may differ from the category formed by monolingual native speakers of the target L2. This occurs when the L2 sound includes features not exploited in the learner's L1, or when the perceptual cues defining the L2 sound are weighted differently than those of the closest L1 sound (pp. 239-243). However, the original SLM does not provide a theory-based explanation for how L2 sounds are equated with L1 sounds. Recently, the SLM was revised into the SLM-r (Flege & Bohn, 2021). While the SLM-r continues to focus on the production and perception of position-sensitive allophones of L2 vowels and consonants in sequential bilinguals, it no longer emphasizes the differences between early and late learners, or between individuals with varying levels of experience in the L2. The SLM-r replaces the ’age’ hypothesis from the original SLM with the ’category precision’ hypothesis. According to the SLM-r, the formation or non-formation of a new phonetic category for an L2 sound is primarily influenced by several factors: the precision with which the closest L1 category is specified at the start of L2 learning, the quantity and quality of L2 input received through meaningful conversation, and the degree of perceived phonetic dissimilarity between the L2 sound and the closest L1 sound. The ability to discern cross-language phonetic differences can be influenced by individual differences in auditory acuity, early-stage (pre-categorical) auditory processing, and working auditory memory. Additionally, the SLM-r proposes that late learners can access features used to define L2 categories that are not exploited in their L1, thus challenging the previous assumption that age limits the ability to acquire certain L2 phonetic features (Flege & Bohn, 2021). SLM and SLM-r do not deal with the role of universal or markedness, thus there have been no explicit claims regarding the relative roles of markedness.

3.3 Influencing factors of non-native perception and production

The present thesis aims to investigate how each of the hypotheses and the models can offer further insight into the possible learning outcomes for Hungarian learners of Chinese. The empirical and the experimental research have been designed to provide evidence either supporting or challenging these theoretical perspectives. Beyond evaluating the hypotheses and models, this study explores the linguistic and pedagogical implications of the findings. The insights gained from the analysis aim to inform language teaching practices and contribute to a deeper understanding of second language acquisition processes, particularly for learners navigating typologically distinct languages such as Hungarian and Chinese.

3.3.1 Different levels of representation

L2 learning models describe the process of second language acquisition in terms of different levels of segmental representation. These representations can range from more concrete to more abstract, with various theoretical models proposing different approaches. However, research in second language phonetics and phonology has not yet reached a clear consensus on the exact type of representation that late learners create during their learning process (e.g. McCarthy, 2007; Pierrehumbert, 1990). As a result, the nature of segmental representation in late L2 learners remains an area of ongoing debate.

Being based on phonological distinctions in the learner's native language (L1), the aim of the contrastive analysis (CA) approach was to identify learning problems that would need to be addressed through instruction in the foreign-language classroom. Its general prediction was that L2 phonemes that do not have a counterpart in the L1 would be difficult to learn whereas those having an L1 counterpart would be relatively easy to learn. CA approach was that allophonic distributions of the "same" phonemes found in two languages often differ cross-linguistically, making point-by-point comparisons of phonemes difficult or meaningless (Kohler, 1981) (cf: Flege & Bohn, 2021: 4).

The Contrastive Analysis Hypothesis and the Interference Model represent two different approaches to understanding second language (L2) phonology. The CAH is based solely on phonological distinctions between the first language and the second language, using phonemic representation (Lado, 1957). It involves making direct, point-by-point comparisons of phonemes between L1 and L2. In this framework, learners may respond by substituting the closest L1 phoneme for an unfamiliar L2 phoneme (Lehiste, 1988; cf: Flege, 1995). However, allophonic distributions of the 'same' phonemes often differ cross-linguistically (Kohler, 1981). The Interference Model, on the other hand, uses feature representation (Brown, 2000) and focuses on how specific features of L1 phonemes might interfere with the learning of L2 sounds. Different languages prioritize different phonological contrasts, which makes a fixed universal feature hierarchy problematic. For example: In Chinese, obstruent contrasts involve voiceless unaspirated and voiceless aspirated pairs, whereas in Hungarian, the contrast is between voiced and voiceless unaspirated segments. Universal feature geometry assumes a fixed structure that may not reflect language-specific priorities. Studies have shown that learning one position-sensitive allophone of an L2 phoneme does not guarantee successful production or perception of other allophones of the same phoneme (e.g. Mitterer, Reinisch, & McQueen, 2018; Strange, 1992; cf: Flege, 2021). In fact, most empirical studies on L2 phonology have focused on the production and perception of position-dependent allophones (Strange, 1995).

Flege (1995) argues that the phonetic level of analysis is more concrete than the phonemic level. According to the hypothesis of the Speech Learning Model, learners perceptually map positionally defined allophones in the L2 to the closest corresponding allophone (or ’sound’) in their L1. The revised version of the model, SLM-r (Flege & Bohn, 2021), continues to emphasize phonetic details. The term ’common phonetic space’ is used to describe the interactions between sounds in the L1 and L2 phonetic systems. The Perceptual Assimilation Model considers both phonetic details and phonological categories or contrasts. The PAM-L2 framework expands on this by explaining how L1 influence on L2 learning occurs through perceptual learning. This learning involves both the abstract, contrastive level of phonological categories and, importantly, the non-contrastive, gradient phonetic details that shape speech patterns (Best & Tyler, 2007).

L1 effects in the L2 at the phonetics/phonology level have been acknowledged by phonological theory. And many language teachers are aware of the effects of L2 orthography on L2 pronunciation. However, it is important to highlight that there are a small number of studies on the effect of orthographic input on speech perception and speech production. L1 effects in the L2 at the orthographic level have not been widely discussed. And graphemes are generally not stated by different perception and production models. In this regard, the present research considered the effects of the orthographic representation of the second language on learners’ L2 phonology as well. I have to mention that the terminology used by researchers in perception and production includes more levels, but for the sake of simplicity the present manuscript only focuses on these three levels, an orthographic level, a phonemic level and a phonetic level.

3.3.2 Relation between Perception and Production

The relationship between speech perception and production in second language learning is a key focus of models such as the Perceptual Assimilation Model and its extension, PAM-L2. These models propose that the pronunciation encountered by L2 learners is shaped by perceptual limitations (Best, 1995; Best et al., 2001). Perceptual skill levels are positively correlated with accuracy in producing L2 vowels and with the ratio of L2/L1 usage (Best & Tyler, 2007). In PAM, speech perception and production are closely linked, with perception constraining production – a principle also supported by the Speech Learning Model. Flege (1995) argued that accuracy in L2 production is limited by perceptual accuracy. He hypothesized that the production of an L2 phonetic segment will generally not surpass its perceptual representation and may be less native-like, particularly in the early stages of learning. This view is consistent with studies by Major and Kim (1999) and Leather (1999), who reviewed evidence suggesting that L2 perception typically develops before production and that accurate perception is a prerequisite for accurate production. Neufeld (1988) described this phenomenon as a 'phonological asymmetry', noting that learners often exhibit greater skill in perceptually detecting sound errors than in avoiding them during production. Barry (1989) found that learners with well-established perceptual categories also demonstrated accurate production, suggesting that perceptual tests could be useful for identifying difficulties in producing L2 vowels and consonants. Additional support for the hypothesis that perception precedes and facilitates production is provided by studies such as Flege (1993) and Rochet (1995).

The relationship between perception and production in second language (L2) learning is a subject of debate. Strange (1995) argues that modifications in phonetic perceptual patterns do not necessarily correspond to changes in production patterns. In some cases, L2 learners may produce non-native contrasts more accurately than they perceive them, either in their own speech or in the speech of others. This suggests that assessing pronunciation alone cannot reliably indicate perceptual mastery, highlighting a practical implication for foreign language teachers. The revised Speech Learning Model (SLM-r) posits that L2 segmental production and perception develop concurrently, with neither taking precedence over the other (Flege & Bohn, 2021). This perspective emphasizes the dynamic and interdependent nature of perception and production in L2 acquisition.

3.3.3 Native language effect

It is widely accepted that speech perception and production in a second language are affected by the learner's native language (L1). When we hear words from an unfamiliar language spoken by a native of that language, we often have difficulty perceiving the phonetic differences among contrasting consonant or vowel sounds that are not distinct phonemes in our own language. It is apparent, then, that the phonology of the native language comes to exert substantial influence on L2 speech perception and production (Best, 1994). Later learning may be constrained by the initial mapping that has taken place. The specialization mechanism results in reduced perceptual sensitivity within the L1 phoneme inventory, and this may become an obstacle when learning non-native sounds in adulthood (Kuhl et al., 1992). The SLM proposed that LI and L2 sounds are perceptually linked to one another through a cognitive process called ’interlingual identification’, which operates automatically and subconsciously. When first exposed to the L2, learners interpret the ’full range’ of L2 sounds they encounter on the phonetic surface of the L2 as being instances, some better than others, of existing LI phonetic categories (Flege, 1995, p. 241). The Markedness Differential Hypothesis further highlights the role of language experience by offering predictions about the relative difficulty of acquiring specific L2 structures. MDH posits that learners face greater challenges with marked structures – those that are less common or more complex cross-linguistically. However, once the most marked structure in a set is learned, learners can generalize this knowledge to less marked structures. This generalization explains why advanced learners outperform beginners; they have already acquired the more marked structures, enabling better overall performance. Adults have persistent perceptual difficulties and production difficulties with many foreign phonetic segments (Strange1995), and a number of studies have examined the production of L2 vowels by L2 late learners. These studies have shown that the influence of the first language (Ll) vowel system is often readily apparent in late learners’ production of L2 vowels, especially in early stages of learning (Flege & MacKay, 2004)

Different researchers have proposed different means to explore how L2 sounds are mapped onto L1 sounds. Firstly, PAM and PAM-2 propose that the accuracy with which L2 speech sounds are distinguished will depend on how, or if, they are perceptually assimilated by L1 speech sounds, including notions of phonetic goodness of fit. The SLM-r maintains the earlier SLM hypotheses that learners subconsciously and automatically relate L2 sounds to L1 phonetic categories is, and that the greater the perceived phonetic dissimilarity of realizations of an L2 phonetic category from the realizations defining an L1 category, the more likely a new phonetic category will be formed for the L2 sound (Flege & Bohn, 2021). Flege and Bohn cited the words that cross-language dissimilarity must be assessed perceptually rather than acoustically because acoustic measures sometimes diverge from what listeners perceive (e.g. Levy and Strange, 2008, p. 153) (cf: Flege & Bohn, 2021). Although perceived similarity or dissimilarity seems convincing in different perception and production models, how to design a ‘perceptual similarity task’ is not always clear. How does one define a wide range of variants of L1 and L2 sounds in a specific phonetic context? A standard measurement procedure has not emerged (for discussions, see Bohn, 2002; Strange, 2007). In other words, an obstacle to testing perceived similarity or dissimilarity is the lack of an objective means. And the conditions under which perceptual assimilation patterns might change were not specified. Secondly, Best (1995) suggested that L1 versus L2 distances can be gauged by the 'spatial proximity of constriction locations and active articulators' and by 'similarities in constriction degree and gestural phasing'. And the Interference Model compares native versus non-native phonemic similarities on the basis of pure phonological features (Brown, 2000). Although these methods are appealing, the metric is also difficult to apply. Finally when predicting and accounting for the ease or difficulty in perceiving second language sounds, L2LP model native versus non-native phonemic similarities on the basis of pure acoustic features (Escudero, 2005). Flege (l995) also states that mismatches between L1 and L2 sounds may be described in terms of differences in the acoustic cues.

3.3.4 Other relevant factors

While similarities and differences between L1 and L2 segments are key factors, other variables also play a crucial role in predicting and explaining the challenges in mastering L2 vowels or vowel contrasts. Although definitive answers are not yet available, several factors have been highlighted across various perception and production models.

First, previous psycholinguistic research has shown that orthography influences word recognition (Perre & Ziegler, 2008; Taft, 2001) and plays a significant role in phonemic awareness (Cheung et al., 2001; Tyler & Burnham, 2006). Furthermore, studies suggest that visual information is more heavily utilized when attending to nonnative speech (e.g. Fuster-Duran, 1996; Sekiyama & Tohkura, 1993). Orthographic input interacts with acoustic input to shape learners' mental representations of L2 phonology. For instructed learners, orthography-induced pronunciations may even become part of the acoustic input (Bassetti, 2008). Spoken and written languages differ in structures and vocabulary (Halliday, 1990), and for many second language learners, a significant portion of L2 input is written rather than spoken. Cross-language differences may also be influenced by visual properties associated with L1 and L2 sounds. Research has indicated a possible age-sensitive period for learning to use visual information for L2 speech perception, with earlier exposure leading to more accurate interpretation of visual cues (Weikum et al., 2013). Additionally, sensitivity to visual cues may diminish if not used for perceiving L1 contrasts, which can make their application in L2 contexts more challenging (Hazan et al., 2005). Unlike children acquiring their L1, L2 learners are often exposed to written input early in the learning process, and written input frequently constitutes a substantial part of their overall L2 exposure. Erdener and Burnham (2005) found that while all L2 learners were better at repeating L2 words when graphemes were provided, the strength of this effect depended on the phonological transparency of the orthographies in both L1 and L2. Learners with experience in transparent orthographies tended to focus more on orthography, whereas those with experience in opaque orthographies appeared to rely more on auditory-visual signals. Notably, native users of transparent L1 writing systems were more negatively affected by less transparent L2 orthographies. A critical aspect of written information is orthographic depth, which varies across alphabetic writing systems. Orthographic depth refers to how closely a writing system aligns with a simple one-to-one grapheme-to-phoneme correspondence (Van den Bosch et al., 1994). Writing systems are conceptualized along a transparent-to-opaque continuum. Transparent orthographies feature unambiguous, straightforward phoneme-to-grapheme correspondences, ideally aligning each phoneme with a singular grapheme (Erdener & Burnham, 2005). However, no writing system fully achieves the one-to-one correspondence seen in phonetic transcriptions (Cook & Bassetti, 2005). In the orthographic conditions, participants were instructed to look at the screen and silently read the orthographic input. This methodology highlights the significant role of orthographic depth and its impact on L2 phonological development and speech perception. Pinyin in Chinese is characterized by its deviation from relatively consistent phoneme-to-grapheme correspondences, but in Chinese one grapheme or one phoneme can represents several sounds. In this case, Chinese sounds-to-grapheme correspondences are inconsistent compared with Hungarian, Hungarian has more transparent orthographies in which phoneme to grapheme correspondences are more consistent, and one phoneme generally only represents one sound (section 2.2.3).

Second, Flege and Bohn (2021) emphasize the importance of considering the context in which input is assessed, as it can significantly impact how well the input is consolidated and, consequently, influence speech learning. Cross-language mapping patterns are not static but can vary depending on phonetic contexts, which may alter the realization of an L2 sound in a language-specific manner (Levy and Strange, 2008, p. 153). For instance, Levy (2009, p. 2680) found that native English speakers perceived the French vowel /y/ as ‘most similar’ to the American English vowel /u/ more frequently when the French vowel occurred in an alveolar context compared to a bilabial context. This highlights how consonantal environments can affect the perception of vowels. Studies examining the effect of consonant contexts on vowels further illustrate this phenomenon. In Mandarin Chinese, strident fricatives are often followed by fricative vowels, where the tongue body assumes a position similar to that used in the corresponding fricative (Ladefoged & Maddieson, 1996). The tongue body backing gesture in Mandarin alveolar and palatoalveolar fricatives is consistent with a lowered F2 frequency in the following apical and retroflex vowels, and vowels following alveopalatal sibilants in Mandarin often exhibit raised F2 values, typically attributed to coarticulatory vowel fronting (Stevens et al., 2004). Hauser (2019) noted that raised F2 values after alveopalatal sibilants frequently persist throughout the entire duration of the subsequent vowel in Mandarin. These findings suggest that phonetic context plays a crucial role in shaping the acoustic and perceptual characteristics of L2 sounds. Recognizing the impact of such contexts is vital for accurately analyzing cross-language sound mappings and for understanding the intricacies of second language speech learning. Accordingly, the present study also takes phonetic context into account.

Third, it is increasingly recognized that there is considerable variability among L2 learners, even within a given group. A wide range of individual differences that may influence the ability to learn speech have been identified in literature. These include factors such as musical ability (e.g. Slevc & Miyake, 2006), selective attention (e.g. Mora & Mora-Plaza, 2019), and phonemic coding ability (Saito & Tierney, 2019). Such differences play a critical role in shaping how learners perceive, process, and produce second language (L2) sounds. The recently revised Speech Learning Model offers a framework for understanding how individuals' phonetic systems reorganize in response to phonetic input. According to the SLM-r, individual differences can significantly influence the process of phonetic category formation for L2 sounds. These differences affect how much L2 phonetic input is needed for learners to progress through various stages of category development. The model also emphasizes the importance of category precision, which it regards as an endogenous factor tied to several individual traits, such as auditory acuity, which affects the ability to detect subtle differences in sounds, early-stage (pre-categorical) auditory processing, which governs how auditory input is initially processed before forming stable categories, auditory working memory, which supports the temporary storage and manipulation of sound information during learning. By considering these factors, the SLM-r provides a nuanced perspective on how individual variability shapes the trajectory of phonetic learning and contributes to differences in L2 acquisition outcomes (Flege & Bohn, 2021).

Fourth, the quality and quantity of L2 input play a pivotal role in the formation of new L2 phonetic categories and the development of composite L1-L2 phonetic categories (Flege & Bohn, 2021). According to Bohn and Flege (1992), success or failure in learning L2 sounds can be explained by the sound correspondences between L1 and L2 as well as the learner's L2 experience. Initially, L2 sounds may be identified in terms of a positionally defined allophone of the L1. However, as learners gain experience, they begin to discern phonetic differences between L2 sounds and their closest L1 equivalents. Flege (1995) emphasized the importance of conversational experience, defining it as the cumulative speech input received during face-to-face verbal interactions in the L2. This experience allows learners to refine their perceptual skills, a process positively correlated with the ratio of L2-to-L1 usage (Best & Tyler, 2007). Learners with minimal L2 experience are likely to strongly assimilate non-native phones to native phonemes. However, with increased exposure to L2 input, learners become better at recognizing the discrepancies between L1 and L2 phonemes. This recognition fosters a gradual reduction in the assimilation of L2 phones to L1 categories, leading to the emergence of distinct L2 phoneme categories (Best & Strange, 1992). Best and Strange (1992) proposed that L2 experience reorganizes perceptual assimilation patterns, thereby enhancing the discriminability of L2 sounds. This progression underscores the importance of extended, meaningful interaction with L2 input in reshaping phonological perception. Such longitudinal insights emphasize the dynamic nature of language acquisition as learners progress through varying levels of proficiency.

In the context of L2 learning, SLM-r stresses the importance of input quality, particularly in meaningful conversational contexts, for forming new L2 phonetic categories (Flege & Bohn, 2021). High-quality input can significantly reduce the disparity between L2 learners and native speakers. However, in environments where the target language is not widely spoken, the role of the teacher becomes central. Teachers are often the primary source of L2 input in such settings (Gass et al., 2013). Loewen (2009) highlights the teacher's capacity to increase the frequency and perceptual salience of targeted linguistic features. While cognitive abilities may not be easily influenced, teachers can optimize instruction by focusing on high-frequency and perceptually salient input. The SLM suggests that L2 speech learning is a gradual process requiring substantial exposure to native-speaker input (Flege, 2003). For SLM-r, both the quantity and quality of input are crucial for developing distinct L2 phonetic categories and integrating them into learners' phonological systems (Flege & Bohn, 2021).

This study investigates the perception and production of Chinese sounds by Hungarian learners across different levels of Chinese language experience. It also examines the qualitative input provided by different groups of Chinese teachers, highlighting the influence of instructional practices on L2 phonetic category formation. By focusing on the interplay between input and experience, the research contributes to a deeper understanding of the mechanisms underpinning L2 speech learning.

3.4 Previous studies on Chinese vowel finals with a special attention to [ɤ], [ʅ] and [ɿ] among Hungarian learners of Chinese

Hungarian learners of Chinese have difficulty in pronouncing Chinese vowel finals. There are several studies on this matter. Zhang (2015) conducted a study on the pronunciation of Hungarian learners of Chinese in which 18 (5 male, 13 female) Hungarian university students from Károli Gáspár University of the Reformed Church in Hungary participated, who learnt Chinese as a regular L2-learner. All participants can speaker other languages before they start to learn Chinese. Up to the time of the experiment, they had learnt Chinese for 96 classes. The learners were asked to read aloud a text which contained 57 monosyllabic words and 67 disyllabic words. The pronunciation was evaluated by Zhang who listened to the recordings and considered the pronunciation correct or not and if not, the segment was noted. Then the segment was analysed by Praat. According to this study, Hungarian learners of Chinese made the most mistakes with the [ɤ] sound, with an error rate of 71.67%, followed by [o] at 54.17% and [y] at 26.73%. [ɤ] was pronounced as [ə]/[e]/[ʅ]/[ɿ]. The two high vowels had the error rates, with [ʅ] at 25% and [ɿ] at 20.83%, respectively. [ʅ] was pronounced as [ə]/[i]/[ʮ]/[ɿ] and [ɿ] was pronounced as [ə]/[i]/[ʮ]/[ʅ].

Jia (2017) also studied Hungarian learners' Chinese pronunciation. 42 Chinese learners from ELTE Confucius Institute. In this study, the subjects had learnt Chinese for 32 classes before the experiment. The pronunciation was evaluated by 3 native Chinese who listened to the recordings and considered the pronunciation correct or not and if not, the segment was noted. According to the results, [ɤ] sound showed the highest error rate (70.26%). As Jia only reported error rates above 40%, and he did not mention the two high vowels, therefore we assume that these were pronounced correctly in more than 60%.

Gao (2018) also did research about the pronunciation of Hungarian Chinese learners. In the study, participants were 14 secondary school students. After 144 classes, learners were asked to read aloud the basic sounds from the Chinese sound chart and some words (non-natural contexts), and to answer their teacher's questions (natural contexts). The pronunciation was evaluated by Gao who listened to the recordings and considered the pronunciation correct or not and if not, the segment was noted. Then the segment was analysed by Praat. Considering the non-natural contexts' [ɤ] showed the highest error rate again (83.23%), and the two high vowels were less frequently pronounced in an incorrect way ([ɿ]: 24.93% and [ʅ]: 17.86%). In the "natural context', the highest error rate was observed for the [o] (93.34%), followed by [ɤ] at 92.85%, and the two high vowels showed relatively low error rates ([ʅ] at 23.79% and [ɿ] 20.04%) again.

Based on formant patterns in the production of 5 native Chinese and 10 Hungarian speakers of Chinese who were BA students of Chinese, Juhász (2020) showed that Hungarian learners produced the velar [ɤ] significantly different from native Chinese speakers. She suggested that in terms of formant values, the Hungarian vowels [y]/[yː] are closest to the Chinese apical vowels which were represented as approximates, while the Hungarian vowel [ø]/[øː] is closest to the Chinese [ɤ], what also can be assumed based on Figure 4 and Figure 5. She suggested, that although Hungarian learners of Chinese did not pronounce [ɤ] similar to [ø]/[øː], but produced the [ɤ] with higher F1, F2 and F3 than native Chinese speakers. [ɿ] and [ʅ] were close to the native pronunciation, in other words there is no significant difference between native Chinese speakers and native Hungarian speakers of Chinese. The 10 L2 speakers were divided into beginner and advanced groups (1st and 3rd years), however, their results did not show significant difference.

To summarize the previous findings, Hungarian learners of Chinese encountered difficulty in the pronunciation of [ɤ]: All studies showed the highest error rate in this vowel, which was between 70% and 93%. In terms of Chinese vowels, it appeared either as [ə]/[e]/[ʅ]/[ɿ], and the formant ratios showed significant difference from the native speakers' data. Opposed to the mid vowel, the high vowels showed much lower error rates, but still above 20%, and their formant ratios did not show any significant difference from the native speakers' pronunciation. [ʅ] was pronounced as [ə]/[i]/[ɥ]/[ɿ] and [ɿ] was pronounced as [ə]/[i]/[ɥ]/[ʅ]. The acoustic production studies (Juhász, 2020) did not find any difference between beginner and advanced learners.

The present study also investigates the production of [ɤ], [ɿ] and [ʅ] by Hungarian learners of Chinese and native Chinese speakers. The present dissertation aims to expand Juhász's production results in three specific directions. One is that in the present research, the speaker groups are different: while Juhász's subjects were undergraduate students majoring in Chinese, the speakers in this study are studying Chinese in an extra-curricular language course at their university. This also means a lower number of language classes per week. The second is to give an insight not only on studying experience of learners but also on their teachers' background. The third is to attest the vowel differentiation in production, i.e., attesting the vowels classification, not (only) the group differences.

4 Research questions

Building on the growing interest in Chinese language education in Hungary and the phonological contrasts between Chinese and Hungarian, this dissertation investigates the specific perceptual and articulatory challenges Hungarian learners face when acquiring Chinese vowel finals. The study pays particular attention to the mid vowel [ɤ] and the apical vowels [ɿ] and [ʅ], which are absent in the Hungarian sound system. Through a multi-method approach involving questionnaires, teacher interviews, and experimental perception and production studies, the research aims to uncover how learners perceive and produce these sounds, the extent to which these vowels are confused, and the influence of general factors such as learning experience and instructional background. The overall goal is to better understand L2 phonological acquisition in a typologically distant language pair and to inform teaching strategies that address these specific phonetic challenges. The following research questions guide this inquiry.

4.1 Research questions of the questionnaire study

RQ1. What are the Mandarin Chinese vowel finals Hungarian learners of Chinese regard most difficult to pronounce?
RQ2. What can be the general factors behind the difficulty of specific Mandarin vowel finals for Hungarian learners of Chinese?

4.2 Research questions of the interview study

RQ3. Which Chinese vowel finals do Chinese teachers consider the most difficult for Hungarian learners of Chinese to pronounce?

RQ4. What discrepancies exist between Hungarian learners' self-assessments and their teachers' evaluations regarding the pronunciation of Chinese vowel finals?

RQ5. What phonetic, phonological, or pedagogical factors contribute to the difficulty of specific Mandarin Chinese vowel finals?

4.3 Research questions of the perception research

Based on theoretical considerations, the production study by Juhász (2020), and the questionnaire and the interview results of section 6.1 and 6.2, the following questions were formulated:

RQ6. Do the Chinese vowels [ɤ], [ɿ], and [ʅ] cause perceptual difficulties for Hungarian learners of Chinese?

RQ7. Do advanced Hungarian learners of Chinese perform better than beginners in the perception of [ɤ], [ɿ], and [ʅ]?

RQ8. Does the L1 of the teacher affect the perception of Hungarian learners of Chinese?

RQ9. Does the consonant context affect the perception of Chinese vowels?

4.4 Research questions of the production research

Based on theoretical considerations, the production study by Juhász (2020), the questionnaire, the interview and the perception results of section 6.1, 6.2 and 6.3, the following questions were formulated:

RQ10. Whether Hungarian learners of Chinese are capable to diffrenciate [ɤ], [ɿ], and [ʅ] in their production?

RQ11. How can the between speaker variance be interpreted in the production of [ɤ], [ɿ], and [ʅ] among L2 learners, as suggested by previous research and the effects of orthography and perception performance?

RQ12. Does the L1 of the teacher affect the production of Hungarian learners of Chinese?

RQ13. Do advanced Hungarian learners of Chinese show better pronunciation of [ɤ], [ɿ], and [ʅ] compared to beginner learners?

RQ14. What phonetic, phonological, or pedagogical factors contribute to the production difficulty of specific Mandarin Chinese vowel finals?

5. Methodology and research design

5.1 Survey of Hungarian students of Chinese

In order to know the views of Hungarian learners of Chinese, an adapted version of the questionnaire from Fan & Myintzu (2022) was designed based on the 21 vowel finals (See the questionnaire in Appendix A). The research questionnaire includes a mix of closed-ended and open-ended questions to gather both quantitative and qualitative data. And in the questionnaire (Appendix A), the term “learning” was used broadly to capture participants' perceived difficulty with specific Chinese vowels. The questionnaire aimed to explore learners’ subjective experiences of difficulty, which may include both perception and production.

5.1.1 Participants

107 respondents completed the questionnaire from 8 universities and 4 Confucius Institutes in Hungary providing Chinese courses. The respondents’ Chinese language proficiency level ranges from beginner (~A1) to -intermediate (~B1) to -advanced (~C2). Intermediate-level students were around HSK 3 (~B1), however specific proficiency has not been measured in a Chinese language exam, thus Chinese language proficiency level is based on their self-reported answers. The data set was constructed for the responses received from the Hungarian learners of Chinese.

5.1.2 Materials and methods

After designing the questionnaire (Appendix A), it was distributed online or within physical classrooms to the Hungarian learners of Chinese. After collecting learners’ responses to the questionnaire, factor analysis was used to categorize the difficulty groups. Factor analysis of perceived difficulty of Chinese sounds was supported by both Bartlett’s test and the KMO measure of sampling adequacy. The sample size was the 107 respondents who filled in the questionnaire.

Median score was used to measure the perceived difficulty level. In the questionnaire, a seven-point Likert-scale has been used, in which 1 = “no difficulty at all”, 2 = “very easy”, 3 = “somewhat easy”, 4 = “neutral”, 5 = “somewhat difficult”, 6 = “very difficult”, 7 =

"extremely difficult", was used to investigate to what extent each vowel final was considered to be difficult by learners to pronounce.

5.2 Interviews with Chinese teachers

In order to know the views of teachers, I have conducted interviews with Chinese and Hungarian teachers who are teaching Chinese from all the major education institutes of Hungary where students learn Chinese, covering six Confucius Institutes in Hungary, three universities with Chinese majors, a Sino-Hungarian bilingual school, and some other educational institutions offering Chinese courses. This sample provides an understanding of similarities in the Chinese language learning process. Qualitative interviews can help to understand which type of problems may occur and especially what kind of pronunciation problems teachers face more often. Interviews may also illustrate which kinds of problems are more general, and what types of interference that teaching professionals feel significant between Hungarian and Chinese languages, if they feel any. The qualitative interview research does not allow generalizations to be made about the experimental data, but it can provide a more detailed picture about the research topic.

There was an assumption in this study that the views of Hungarian learners of Chinese on their ability to pronounce Chinese vowel finals is different from their teachers.

5.2.1 Participants

Table 14. Participants of the interview

		Chinese	Hungarians
Sex	Female	9	4
	Male	1	6
Education background	Ph.D	2	2
	M.A	8	6
	B.A	0	2
Teaching experience in Hungary	Range in years	3~12	5~20
	Chinese major	3	4
	Confucius Institute	5	0
	Training School	2	3

The participants of the study were 20 Chinese foreign language teachers (10 native Chinese teachers and 10 native Hungarian L2-speaker of Chinese) with B.A, M.A, and Ph.D. degrees (Table 14), 13 being male and 7 being female. All the interviews were conducted in Chinese. Among 10 native Chinese teachers, only one could speak Hungarian. The ten native Hungarian teachers had studied Chinese in China before. The Hungarian teachers did not have educational background in second language teaching, but they had many years of Chinese teaching experience. All these 20 teachers had at least 4 years of Chinese language teaching experience to non-native Chinese learners and 3 years of teaching Chinese language experience to Hungarian learners of Chinese.

7 Chinese teachers (3 native Chinese teachers and 4 native Hungarian L2-speaker of Chinese) are currently teaching Chinese majors. Chinese courses at the three universities in Hungary that offer Chinese majors are taught by native Chinese teachers and Hungarian Chinese teachers. 5 native Chinese teachers are from five Confucius Institutes in Hungary and have never worked with Hungarian Chinese teachers. 5 Chinese teachers (2 native Chinese teachers and 3 Hungarian Chinese teachers) are currently teaching in training institutions, and one of them has also had teaching experience in Chinese majors and Confucius Institutes. The three native Hungarian Chinese teachers who taught in other institutions were from two universities offering Chinese elective courses and a bilingual school (which built a Confucius institute in 2024 with Capital Normal University).

The 10 native Hungarian L2-speaker of Chinese had never taught in any Confucius Institutions, and the three teachers in training institutions and the three teachers in other institutions had never worked with native Chinese teachers. Nine native Hungarian L2-speaker of Chinese taught adult Chinese in universities and training schools, and one taught teenagers Chinese in primary and secondary schools. He also taught adults in training schools. The 10 native Chinese teachers also mainly taught adults.

5.2.2 Materials and methods

In order to investigate Chinese teachers' beliefs to the perception and pronunciation of Hungarian learners of Chinese in case of Chinese vowel finals, this thesis conducted in-depth interviews with Chinese teachers using a semi-structured interview method. The interviews were mainly one-on-one face-to-face interviews, and online interviews were also used. For offline interviews, the interview location was determined according to the actual situation, and each in-depth interview lasted no less than half an hour.

The interviews (see Appendix B) included questions regarding the teachers' educational background, their knowledge of Hungarian, and their approach to teach Pinyin (e.g. syllable-first vs. segmental progression). Teachers were asked to identify which vowel finals or graphemes they found most problematic for learners and to describe the nature of the difficulties, such as articulation issues or confusion between sounds. Additional questions explored views on pronunciation error correction, factors influencing learners' phonological acquisition (e.g. academic background, multilingualism), and the impact of co-teaching with native Chinese or Hungarian instructors. Finally, teachers were invited to offer suggestions for improving pronunciation teaching. These interviews were designed to provide contextual data that would help us interpret the experimental results on perception and production.

The interview questions were designed based on the Chinese teachers' experience in teaching Chinese to Hungarians and the teachers' educational background. Before the interview began, the purpose of the interview was explained to the interviewees, and the consent of all interviewees was obtained in advance, and the interview content was recorded throughout. After the interview, a small gift was given to the interviewee in return.

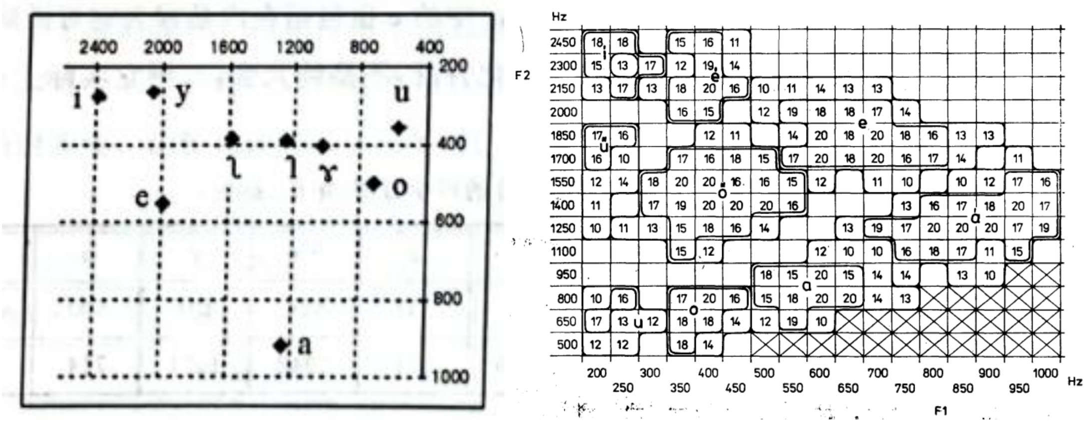
In the process of organizing the interview content, the interviewees were numbered. At the same time, for the writing of the thesis and understanding of the interviewees' narratives, the interviewers' oral narratives were 'standardized' in text conversion based on verbatim transcription, that is, the colloquial language and repetition in the interview narratives were deleted. The Chinese teachers in the interview believed that the vowel finals that Hungarian learners of Chinese were most likely to make mistakes in came from the interview notes.

5.3 Perception experiment

The third study examines the identification of Chinese vowels [ɤ], [ɿ] and [ʅ], and a comparison between acoustic values of Chinese [ɤ], [ɿ], [ʅ] and the Hungarian perceptual vowel map/space is also show in Figure 8. We can see the Hungarian perceptual vowel map/space and the Chinese acoustic data. Comparing two graphs, we can see that the acoustic measurements of [ɤ], [ɿ] and [ʅ] are mapped into the same Hungarian perceptual category of

<ö/ő> [ø]. It can also be seen that the three analyzed vowels [ɤ], [ɿ] and [ʅ] are quite close to each other in Figure 7 as well. Hungarian learners of Chinese produced the [ɤ] in a more acoustically palatalized way (Juhász, 2020). These facts might result in interference in the perception/identification, which is also the motivation to design a two - alternative forced - choice.

Figure 8. Comparison between Formants of Chinese vowels and Hungarian perceptual vowel map/space



(Source: Mandarin vowels (left) (Zhu, 2010: 268); and Plot of Hungarian vowels (right) (Kiss, 1985)

5.3.1 Participants

31 native Hungarian speakers participated in a discrimination task (Table 15). All participants were born and raised in Hungary. They were divided into four groups: The intermediate group included 10 subjects who had been learning Chinese for five semesters. The first two semesters they were studying online with a native Chinese and a native Hungarian L2-speaker of Chinese. From the third semester, they started to have physical classes, but they still studied with two teachers, a native Chinese and a native Hungarian L2-speaker of Chinese. During the fifth semester in which the perceptual test was carried out. The rest of the participants were divided into three groups. All these participants had been learning Chinese for one semester at the time the experiment was administered. 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 intermediate group had spent half a year in Taiwan as an exchange student studying economics, but she did not learn the language. 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 these sounds [ɤ], [ɿ] and [ʅ] are part of 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 to 22 years.

Table 15. Participants of perception research

	Teachers	Number of students	Group names	Number of semesters
Beginner groups	Native Chinese speaker	7	BegChi	1
	Native Hungarian speaker	6	BegHun	1
	Both teacher	8	BegMix	1
Intermediate group	Both teacher	10	IntMix	5

5.3.2 Materials and methods

[ɿ] & [ʅ] vowels are denoted by two different IPA-symbols and thus are considered as two segments having different places of articulation. Since [ɿ] and [ʅ] are in complementary distribution with the high front vowel [i]: [ɿ] only occurs after dentals [ts, tsh, s], [ʅ] only occurs after retroflexes [tʂ, tʂh, ʂ, ɻ], and [i] occurs in other environments. Compared with the two apical vowels, [ɤ] enjoys a more expansive phonotactic context, not only after [ts, ts^h, s] and [tʂ, tʂ^h, ʂ, ɻ], but also some other consonants like [k, k^h, x, t, t^h, n, l]. On this basis, in this dissertation the apical vowels are going to be treated as one category within the identification task, and the test only contrasts the following two pairs: [ɿ] with [ɤ] and [ʅ] with [ɤ]. More precisely, the present research only focuses on the syllables starting with consonants [tʂ, $tʂ^h$, ʂ, ɻ, ts, ts^h, s] and ending with vowels [ɤ, ɿ, ʅ] in Table 16, to see whether Hungarian learners of Chinese can correctly identify these syllables. In addition, in the perception test, the participants perceived the whole syllable. [ɤ] and [ɿ]/[ʅ] contrast in the minimal pairs in Table 16, but [ɿ] and [ʅ] do not.

All stimuli consisted of CV-syllables ending in [ɤ], [ɿ] or [ʅ] (Table 16). 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 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 analyzed in this study. The items were recorded isolated, along with fillers (additional items for future analysis). The 14 stimuli of the present study were part of 128 items (test items and distractors) played three times each (N = 384) in a 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 (Speech Recorder: Christoph & Klaus, 2004). After recording the sounds, a professional native Chinese voice actor evaluated how close the pronunciation is to ideal Mandarin Chinese variety (Appendix C). The voice actor had passed Mandarin Chinese proficiency test with I-B, which means that he is qualified to be a broadcaster[7], he was born in Hebei and graduated from university in Beijing. The actor had also dubbed Chinese textbooks. Therefore, he can be considered as a highly ideal-close speaker of Mandarin. He listened to the speech samples from the six native speakers of Chinese and scored them. Among the 6 native speakers, a male speaker was selected, because he was one of the best speakers. The male speaker is from Mainland China, and he has been studying in Hungary for 3 years. During these three years, he was mostly in a Chinese-speaking environment, except for attending classes at the university. ‘man’

[7] The Putonghua Shuiping Ceshi (PSC), or Putonghua Proficiency Test, is an official standardized test of spoken fluency in Standard Mandarin (Putonghua) in China. It's developed by the Ministry of Education (MOE) and the Institute of Applied Linguistics, and is intended for native Chinese speakers. The test assesses a candidate's ability to use and understand Standard Mandarin effectively. The test is required for many jobs in broadcasting, education, and government in China. The PSC typically consists of three levels (I, II, and III), each with sub-levels (A and B). Successful completion of the PSC results in a "Certificate of Proficiency in Standard Chinese"(https://www.cltt.org/#/webHome).

Table 16. Stimuli of the perception research (consonant context as a variable)

Non-retroflex [ts, tsh, s]		Retroflex [tʂ, tʂh, ʂ, ɻ]	
Pinyin	IPA	Pinyin	IPA
<zè> - <zì> ‘tone’ - ‘letter’	[tsɤ] - [tsɿ]	<zhè> - <zhì> ‘this’ - ‘sign’	[tʂɤ] - [tʂʅ]
<cè> - <cì> ‘side’ - ‘time’	[tshɤ] - [tshɿ]	<chè> - <chì> ‘remove’ - ‘red’	[tʂ^{h} ɤ] - [tʂ^{h}ʅ]
<sè> - <sì> ‘color’ - ‘four’	[sɤ] - [sɿ]	<shè> - <shì> ‘club’ - ‘be’	[ʂɤ] - [ʂʅ]
		<rè> - <rì> ‘hot’ - ‘day’	[ɻɤ] - [ɻʅ]

In Table 16, the retroflex feature indexes the consonant context, since [tʂ, tʂ^{h}, ʂ, ɻ] are retroflex sounds.

The present research examined identification of these vowels in a two-alternative forced-choice test (2AFC) by adult native Hungarian speakers. The perception test was administered in a quiet room through headphones by Praat Experiment MFC (Boersma & Weenink 2022).

The subjects were instructed to select the word they heard as soon as possible after listening to the stimulus (Table 16). The subject listens to a single stimulus and must make a choice between two possible 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, the participants did not have to decide between any two high vowels, but between the mid vowel and one of the high vowels. The response and the reaction time were recorded.

5.3.3 Statistics

Among different perception models, error rate or correct rate is used to describe perception data. Difficulty in this view was equated with errors. If a learner made an error, or errors, this was a signal that the learner was having difficulty with a particular structure or sound. Thus, the present study uses error rate to analyze the perception difficulty of Hungarian learners of Chinese.

The correct answers were analyzed in R (R Core Team, 2022). Generalized mixed models were run: Binomial Generalized Linear Mixed Models (BGLMM) for the correctness of the answers, and Mixed Effects Logistic Regression for the reaction times (lme4: Bates et.al., 2015; lmerTest packages: Kuznetsova et.al., 2017). Tukey post hoc test was run 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 largest possible largest, converging model.

The correctness of the answer was set as dependent variable. The factors were: phoneme category (i.e., mid or high vowel), learner group (intermediate, 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 analyze the possible effect of the retroflex context and own feature of the vowel on the correct identification, the results for the mid and high vowels were tested separated by two further BGLMMs. The correctness of the answer was set as 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 by ggplot2 (Wickham, 2016).

5.4 Production research experiment

Lado emphasized that the list of problems derived from comparing the foreign language with the native language should be regarded as hypothetical until it is validated through analysis of the actual speech of students (Lado, 1957, p. 72). Ladefoged & Disner (2012) states that the most useful representation of the vowels of a language is a plot showing the average values of Formant one (F1) and Formant two (F2) for each vowel spoken by a group of speakers. The present research examines the production of Mandarin Chinese vowels [ɤ], [ɿ], and [ʅ] among Hungarian learners of Chinese based on an acoustic analysis of F1, F2, and F3. However, we also have to note that the connection between the articulation and acoustics is not linear (see

Stevens' quantal theory (Stevens, 1989), therefore we are not going to draw conclusions on the possible articulatory background of the acoustic results.

5.4.1 Participants

The acoustic experiment consisted of several subject groups. Six native speakers of Chinese (CNS, 3 women and 3 men), and the four groups of native speakers of Hungarian (HNS, total n = 30). There are one native Chinese group and four native Hugnarian groups. The four groups of Hungarian learners came from language classes (learning Chinese as an optional course).

The participants of the CNS group were Mainland Chinese who had lived in a Hungary for a minimum of half year (maximum 1 year). 3 males had passed the required exams to begin studies at the University of Hungary. These students were not engaging with the local community. 2 female participants are Chinese teachers from the Confucius Institute and the third female participant is a housewife who has just moved to Hungary from Beijing. None of the six native speakers can speak Hungarian. They were all born and raised in Mainland China until they moved to Hungary. They spoke standard Mandarin. They reported no history of speech or hearing disorders. As mentioned in section 5.3.2, the male voice actor who had passed Mandarin Chinese proficiency test heard the recordings of the 6 CNS. He listened to the speech samples from six CNS, and evaluated how close the pronunciation was to ideal Mandarin Chinese variety, using a five-point scale, ranging from 非常不好 (Very bad) to 非常好 (Very good). Based on his opinions (appendix C), two were rated as very good, three were received as good, one was rated as 不好也不坏 (not good, not bad)', suggesting moderate pronunciation skills.

The participants of the HNS group are all university students whose age is from 18 to 22 years old, and they were born and raised in Hungary, so Hungarian was used in their daily lives. They all can speak English, some of them can speak other languages too. But none of these languages include our target vowels [ɤ], [ʅ] and [ɿ]. In the present study we treat all these languages as L2 languages. Hungarian learners of Chinese in the present study are studying in Hungary, and they all had similar formal education. None of them had previous experience conversing in Chinese. All the Hungarian participants started Chinese language training at the beginners' level at the Hungarian universities. They have L2 exposure primarily through formal instruction in a restricted setting, with little or unsystematic conversational experience with native speakers.

The participants of the HNS group were divided into four groups (the intermediate group and the three beginner groups) based on their different levels of Chinese language learning experiences. The participants of the intermediate group (intMix) are the same as the participants in section 5.3.1, only the student who had spent half a year in Taiwan as an exchange student did not participate in the production research. The intMix group included 9 members (5 females and 4 males). The beginner group arrived from three sub-groups based on their Chinese teachers; they never had online Chinese courses. In the first group (7 participants: 3 males and 4 females) the teacher was a native Chinese (begChi), the second group (6 participants: 4 males and 2 females) the teacher was a native Hungarian (begHun), and the third one (8 participants: 1 male and 7 females) was taught by these two teachers together (begMix) (Table 17). After the experiment the only male speaker of the begMix group reported dyslexia, therefore his data were eliminated from further analysis.

Table 17. Participants of the production research

	Teachers	Number of students	Group names	Number of semesters
Beginner groups	Native Chinese	7 (3m, 3f)	BegChi	1
	Hungarian L2 speaker	6 (4m, 3f)	BegHun	1
	Both teachers	7 (0m, 7f)	BegMix	1
Intermediate group	Both teachers	9 (5m, 4f)	IntMix	5

Therefore, most of the participants in production research and the participants in perception research are identical. Only one student from the intermediate group did not participate in the production research.

5.4.2 Materials and methods

The primary focus of the present study is on differences among vowels, Table 18 shows the number of tokens analyzed for this study.

Table 18. Stimuli of the production research

(an empty cell means there is no meaningful words with that tone in Chinese.)

Phonetic context	Sounds	Tone 1	Tone 2	Tone 3	Tone 4
Retroflex [tʂ, tʂ^h, ʂ, ɻ]	[tʂɤ] - [tʂʅ]	<zhē> <zhī> 蛰 之 ‘sting’‘of’	<zhé> <zhí> 折 值 ‘fold’‘cost’	<zhě> <zhǐ> 者 纸 ‘ -er’‘paper’	<zhè> <zhì> 这 志 ‘this’‘sign’
	[tʂ^{h}ɤ] - [tʂ^{h}ʅ]	<chē> <chī> 车 吃 ‘car’‘eat’		<chě> <chǐ> 扯 尺 ‘pull’‘ruler’	<chè> <chì> 撤 赤 ‘remove’‘red’
	[ʂɤ] - [ʂʅ]	<shē> <shī> 奢 师 ‘luxury’‘master’	<shé> <shí> 蛇 十 ‘snake’‘ten’	<shě> <shǐ> 舍 使 ‘shed’‘make’	<shè> <shì> 社 是 ‘club’‘be’
	[ɻɤ] - [ɻʅ]				<rè> <rì> 热 日 ‘hot’‘day’
Non-retroflex [ts, ts^h, s]	[tsɤ] - [tsɿ]		<zé> <zí> 则 蓛 ‘then’‘bud’		<zè> <zì> 仄 字 ‘tone’‘letter’
	[ts^{h}ɤ] - [ts^{h}ɿ]				<cè> <cì> 侧 次 ‘side’‘time’
	[sɤ] - [sɿ]	<sē> <sī> 閪 斯 ‘lost’‘this’			<sè> <sì> 色 四 ‘color’‘four’

All stimuli consisted of CV-syllables ending in [ɤ], [ʅ] or [ɿ] (Table 18). All consonantal environment that trigger /ə/ to appear as [ɤ] and /i/ as either [ʅ] or [ɿ] were used in all tones of Chinese, and the consonant-tone combination in question for the CV - structure represents an existing lexical item in Table 18.

Altogether 34 words (17 minimal pairs) were selected. The wordlist included 94 distractors as well. Each item was repeated 3 times (total Number: 384, target Number: 34*3 = 102) in random order.

The present study examines formant values of vowels in Chinese for native Chinese speakers and Hungarian learners of Chinese.

The experiment was carried out in the phonetic laboratory at the HUN-REN Hungarian Research Center for Linguistics. The material was recorded in a sound-treated room using a head-mounted Beyerdynamic omnidirectional condenser microphone and Speech Recorder (Christoph & Klaus, 2004). The words/material were recorded at a sampling frequency of 44,100 Hz.

Before the experiment, participants were asked to fill out a language background questionnaire first. After filling out the form, participants were asked to perform a training task, producing 3 trial items. Stimuli were presented isolated one-by-one in random order using Speech Recorder. One speaker was eliminated due to dyslexia. Therefore, the begMix group consists only of female speakers. Therefore, all other data could be used for acoustic analysis. If the repetition was noisy (e.g. cough), the speaker was asked to repeat the task.

5.4.3 Formants measurement

Segmentation was carried out in Praat (Boersma & Weenink, 2022). The start and cessation of F2 were used as V boundaries. In the case of [ɻɤ] and [ɻʅ], the spectrogram had to be set to the range of 0-8 kHz 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 as the start of the vowel. The first three formants of the stimuli were measured in Praat automatically by a script. The formant range was set to 5000 Hz for male speakers and 5500 Hz for female speakers, except in the case of the four [ʅ] vowels, where the F2 and F3 fall close to each other leading to mismeasurements. 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, window: 0.025 s, time step: 6.25 ms (automatic), pre-emphasis from 50 Hz). The median formant values were taken at the mid 40 ms of the vowels. The data were manually checked and manually corrected for the outliers by speaker and by vowel.

5.4.4 Statistics

Data analysis and statistics were carried out in R (R Core Team, 2024; RStudio: Posit Team, 2025). The formant values were normalised into Z-scores by speaker. Linear Discriminant Analysis (LDA) models were trained on the 6 native speakers' data using all three formants. The testing was carried out on the L2 learners' data. The LDA were built in MASS package (Venables–Ripley, 2002). Figures were prepared in ggplot2 (Wickham, 2016) and ggord (Beck, 2024).

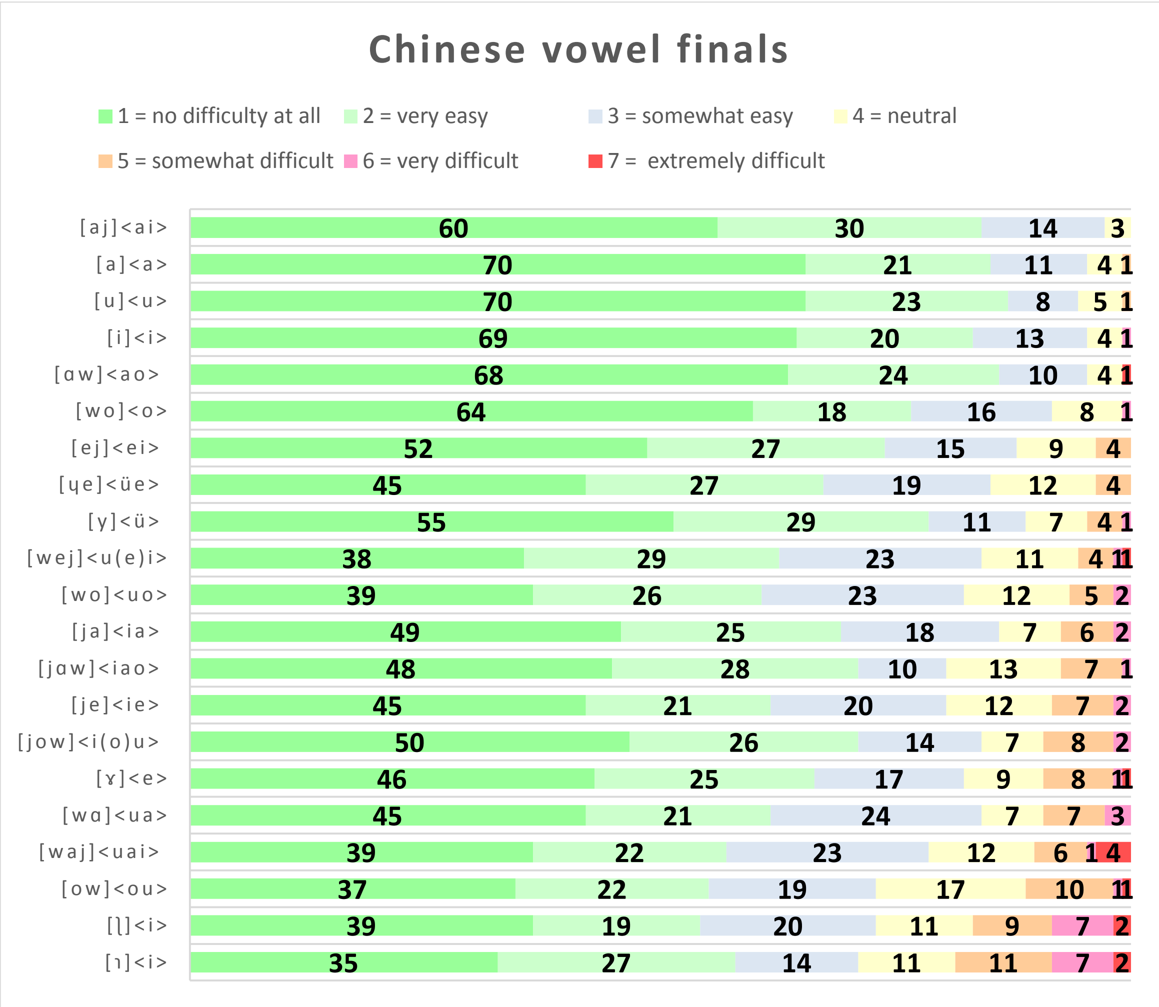
6. Empirical Findings and Discussions

6.1 The beliefs of Hungarian students of Chinese

6.1.1 Results

Hungarian learners' responses about their perceived difficulties in learning Chinese vowel finals can be seen in Table 19. <o> [uo] after labial consonant is transcribed as [o] by most scholars, e.g. Zhu (2010, p. 268) and Lin (2007). but when the Chinese phonetic alphabet (Pinyin) was formulated and published, it was clearly explained that the actural pronunciation of [o] after labial consonants is [uo] (Lin & Wang, 2013, p. 233). Wang (2001) also states that <o> after labial consonant is [uo] based on formant cues. In the present research, the graphemes used for teaching students are considered. Specifically, <o> represents [uo] after labials, while <uo> represents [uo] after non-labials. Therefore, <o> [uo] is listed independently.

Table 19. Perceived difficulty of Hungarian learners of Chinese on a 7-point Likert scale



Based on the results of the questionnaire, among Chinese vowel finals 21 out of 107 (20%) learners in the study thought that [ou] was difficult to learn. 20 out of 107 (19%) learners thought that [ɿ] and [wai], 19 out of 107 (18%) learners thought that [ɤ], and 18 out of 107 (17%) learners thought that [ʅ] was difficult to learn. In addition, despite the fact that <o> after labial initial and <wo> after non-labial initial are both pronounced as [wo] in Mandarin Chinese, Hungarian learners tend to treat these two items as phonemically distinct. The factor analysis of the questionnaire showed that Hungarian learners of Chinese divided Chinese vowel finals into four difficulty groups with all difficulty groups being significantly different from one and other: (i) [ɤ], [ɿ] and [ʅ]; (ii) four triphthongs and diphthong [ye]; (iii) monophthongs [a, i, u, y, o] and [ai], [ɑu]; (iv.) the remaining diphthongs [ou], [wo], [ja], [wɑ], [ei], [je]. The analysis was supported by Bartlett's test which showed significance, and the analysis was also supported by the 0,887 KMO value, which can be considered very good. The 4 factors (groups) explained 72% of the total variance, and it is not worth adding more factors to explain a higher level of total variance, as any additional factors' Eigen values would have been less than 1.

All of the sounds had a median score of less than 3, indicating generally low ratings across the dataset. Specifically, out of the 21 sounds evaluated, 8 received a median score of 1, reflecting the lowest possible score, while the remaining 13 sounds had a slightly higher but still low median score of 2.

6.1.2 Discussion

To answer the first research question, the analysis of learner self-assessments shows that all 21 Chinese vowel finals received median difficulty scores of 1 or 2, indicating that Hungarian learners of Chinese, overall, did not perceive any specific vowel final as highly difficult to learn. However, patterns within the data suggest that certain vowel finals were consistently perceived as similarly difficult by learners.

Notably, the vowel finals [ɤ], [ɿ], and [ʅ] are treated as the same difficulty group. It is reasonable that students put them into one difficulty group, since Hungarian doesn't have these sounds in its vowel system, and these three sounds are marked vowels, which require articulatory configurations unfamiliar to Hungarian speakers. Moreover, acoustic data (Figures 3 and Figure 8) show more overlapping F1 values and close F2 values for these vowels may also make the students treat them as similar sounds. These findings suggest that learners' difficulty ratings are strongly influenced by the absence of corresponding sounds in their native language and by the acoustic proximity of the unfamiliar Chinese vowels. A similar pattern was observed in learners' treatment of Chinese triphthongs. Since Hungarian does not have triphthongs, it is unsurprising that learners grouped them into a shared difficulty category. The complex articulatory transitions are likely to pose additional challenges for learners whose L1 lacks such structures.

Another interesting observation involves the diphthongs [ai] and [ɑu], which Hungarian learners grouped together with the monophthongs [a], [i], [u], and [y]. This classification may be due in part to the phonetic tendencies observed in Chinese, where diphthongs like [ai] and [ɑu] often undergo monophthongization in natural speech (Wang, 2008). Additionally, as a reviewer of the dissertation points out, although Hungarian does not possess diphthongs as phonemes, similar sound sequences (e.g. [au] in *autó*, [ai]/[aj] in *háj*) exist in the language, albeit within different syllabic structures. This partial familiarity may have made these vowel combinations more accessible to learners.

Furthermore, Hungarian learners appeared to differentiate between the syllables <o> and <wo>, despite both being pronounced as [wo] in Chinese. This indicates that some learners treat <o> and <wo> as separate phonological units, suggesting a disconnect between the actual phonetic forms and the mental representations formed during learning. Such a discrepancy may be attributed to orthographic influence.

Thus, to address the second research question, the general factors such as L1 phonological structure, acoustic characteristics of L2 sounds, orthographic input, and individual cognitive factors likely play an influential role influencing learners' beliefs of difficulty with Chinese vowel finals. These findings underscore the need for pronunciation instruction that explicitly addresses marked sounds, includes perceptual training, and clarifies mismatches between orthography and phonetics. Further research is needed to examine how these factors interact over time.

6.2 The beliefs of Chinese teachers

To better contextualize the findings of the perception and production experiments, semi-structured interviews were conducted with Chinese language instructors teaching in Hungary. The interviews focused on identifying which vowel finals were the most difficult for Hungarian learners of Chinese, teachers provided insight into common learner errors, such as mixing up vowels, incorrect articulation, and influence from Hungarian phonological patterns. Responses also shed light on diverse instructional strategies for teaching Pinyin, as well as the benefits and challenges of co-teaching with native or non-native instructors. Notably, the attitutes of correcting pronunciation errors, learners' prior language knowledge and learners' academic background are also investigated. These qualitative findings help frame and interpret the observed patterns in learner performance and provide pedagogical context to the experimental data.

6.2.1 Results

We did not find any special differences between males and females. But there are some interesting findings.

First, from the perspective of Chinese teachers, 15 out of 20 participants mentioned that there were different types of problems with [ɤ]. Two of the teachers mentioned that some learners would pronounce it as Hungarian <e>[ɛ]/[e:] when they first started learning Chinese. For example, when students saw <de> 德, they would pronounce [tɤ] as [tɛ] or [te:] possibly as a result of the effect of orthography, but this problem did not exist after the elementary stage. Another five teachers mentioned that students always confused [ɤ] with [ɿ]/[ʅ].

As for [ɿ] and [ʅ], in addition to the confusion between [ɤ] and [ɿ]/[ʅ] mentioned above, three other teachers mentioned that learners would pronounce them as Hungarian [i] possibly as a result of the effect of orthography.

In addition to [ɤ], [ɿ] and [ʅ], fourteen out of twenty teacher participants mentioned [ou] and [uo] have mixed problems, for example, when the students would like to say [kou] 'dog' and [xəntuo] 'many', they pronounce them like [kuo] and [xəntou] respectively. Four teacher participants also reported [y] has some problems. Three of them said they mispronounce [y], because they did not grasp the spelling rule of Pinyin and one of them said the students mispronounce [y] because the mouth shape is not good.

Second, a wide range of instructional strategies are used by Chinese language teachers in Hungary when teaching Pinyin. The most commonly reported method involved systematic practice of initials, finals, and tones, often supported by the use of Pinyin tables or example characters. This structured approach was typically implemented at the beginning of Chinese instruction and, in some cases, continued through the first semester. Several teachers reported they follow the sequence outlined in standard textbooks to introduce initials and finals. In contrast, a small group of teachers preferred a holistic approach, beginning with whole syllables and only breaking them down into their components if difficulties arose. A number of instructors emphasized imitative learning, focusing on repetition and listening while avoiding explicit phonetic explanations. Additionally, a few teachers employed unique approaches. For example, one teacher introduced Pinyin through an explanation of phonemes, while another kept pronunciation instruction and character learning separate.

Instructional practices also varied depending on the teacher's background and institutional setting. Generally, native Chinese teachers— particularly those working in degree programs —focused on pronunciation through imitation, often using new vocabulary and reading texts for practice. In contrast, their colleague native Hungarian teachers tended to provide explicit explanations of the phonetic system, including rules and structure. Teachers working outside Chinese degree programs, such as those in Confucius Institutes or language schools, reported that time constraints and language barrier often limited the scope of phonetics instruction. In such settings, teachers tended to prioritize speaking practice and vocabulary recognition over explicit pronunciation training. Pronunciation practice was typically embedded in vocabulary reading or dialogues. At intermediate and advanced levels, phonetics instruction was often de-emphasized in favor of improving students' overall communicative abilities.

Conversations with two Hungarian native teachers holding doctoral degrees revealed a more theory-driven and reflective approach to phonetics instruction. One teacher incorporated contrastive phonetic analysis between Hungarian and Chinese, while the other used the International Phonetic Alphabet (IPA) to explain Chinese phonology. Both tailored their instruction to Hungarian learners and brought a strong linguistic awareness to their teaching. By contrast, the two interviewed Chinese teachers with doctoral degrees had previously taught university students in China rather than L2 learners abroad. One, with an English-teaching background, used English as a bridge language for explaining Chinese pronunciation, including comparisons between English and Chinese. The other, specializing in Chinese philosophy, emphasized imitation over analysis, expressing the belief that teaching individual vowels and finals was unnecessary.

Third, Chinese teachers who teach Chinese major think that it is not necessary to offer listening classes at the elementary level, but it is necessary at the intermediate and advanced levels, because non-beginners have a rich vocabulary, furthermore, Chinese voice input should not come from only one native Chinese teacher. Most Chinese teachers who are not teaching Chinese major think that offering listening classes is not a question of necessity but the conditions. For example, in Confucius Institutes, training schools and other higher education institutions that offer Chinese elective courses, there are few Chinese classes (every week two or four class hours). In the limited time, teachers all think that speaking is more important, and students can practice listening on their own after class. One of the native Hungarian Chinese teachers also mentioned: "Listening classes are not suitable for Westerners. Listening classes are boring. I still remember that [when I was a student myself] I was daydreaming after five minutes. Our brains are different, and we don't have that habit. A few minutes is fine, but not the whole class." At present, in Hungary, except for Chinese majors, other institutions do not offer listening classes specifically for students.

Fourth, the survey found that Chinese teachers have pragmatic beliefs towards students' pronunciation errors. Eight out of the twenty respondents (five native Chinese teachers and three native Hungarian teachers) said that they were more tolerant when correcting pronunciation errors, for example, they would only correct pronunciation when there were common problems or when they did not understand. A native Chinese teacher said:

"I will ask students to repeat and answer questions. I will not interrupt students' answers because of pronunciation, nor will I tell them that they have pronunciation problems. I generally do not correct students individually but will summarize their pronunciation problems and practice with the students together. I think adults know where their problems are, so there is no need to tell them deliberately. Pronunciation correction should not be repeated in the same place or on the same person. However, I will also consider the student's personality. If one particularly wants to know which sound he is not speaking well, I will tell him. I believe there is no perfect pronunciation, so if it is not necessary to correct it and not influence communication, I will not do anything."

A native Hungarian teacher said: "I don't know. If I don't understand what the students are saying, I will help them correct it. If I can understand most of it, I will consider the students' needs, sense of achievement or pressure." Three other teachers mentioned that if students have pronunciation problems, they must correct their pronunciation, but after correction, if students still make the same mistakes, their feedback will still use positive words such as 'much better', and they will consider the students' self-esteem and self-confidence. Another four teachers said that they would correct, but if it still doesn’t work after several corrections, they may not be so strict. For example, one teacher said: “I will correct if there are pronunciation problems, but if it still doesn’t work after several corrections, I will give up. But there are a few sounds that I will never give up, such as 老师(<laoshi> [lawʂʅ] ‘teacher’). I usually give up on the difference between aspirated and unaspirated. I know it’s not that good.”

All in all, most Chinese teachers are positive and affirmative about students’ pronunciation, and they pay more attention to communication skills. Only five teachers said that they are very strict in correcting pronunciation, and they will correct students if their pronunciation is not standard.

Fifth, regarding the influence of Chinese teachers on pronunciation, a common agreement among the teachers is that it is beneficial for native Chinese teachers to lead the oral language classes, while native Hungarian teachers should be responsible for explanations and clarifications. Six native Hungarian teachers of Chinese who had experience working alongside native Chinese teachers all expressed the view that having a native Chinese teacher for pronunciation instruction was highly advantageous for students. These teachers highlighted that native Chinese teachers could provide authentic pronunciation models, which is essential for students learning the correct tones and sounds. However, two of these teachers noted that the effectiveness of this collaboration could be compromised if the native Chinese teacher was either unwilling to cooperate or lacked an understanding of local Hungarian cultural customs. They warned that such issues could have a negative impact on the students’ learning experience. One teacher mentioned: “If the Chinese teacher cannot speak Hungarian and fails to understand the local culture, they might unintentionally dampen students’ enthusiasm for learning.” Another teacher shared her experience, stating that her Chinese colleague never engaged in discussions about the students’ learning progress or teaching content, which led to a disconnect in their teaching approach. These insights reflect the importance of not only linguistic proficiency but also cultural understanding and collaboration between teachers for fostering a positive learning environment.

Sixth, regarding whether students’ knowledge of more languages will affect their Chinese phonetics learning, 5 native Chinese teachers and 6 Hungarian Chinese teachers all believed that the languages learners knew before would have a positive impact on their Chinese learning, such as making them more receptive to a new way of thinking, having a sense of language, knowing more language learning strategies, and being familiar with more sounds. However, 7 teachers emphasized that students' knowledge of more languages would not help them learn Chinese phonetics. In addition, two teachers believed that the languages learners knew before would have a negative impact on their Chinese phonetics learning. A native Chinese teacher said: "Learning more languages may make them more skilled, but the pronunciations will be confused with each other." Another Hungarian teacher said: "It will have an impact, but it's not good. When they first started learning Chinese, they tried their previous learning experience, but they don't want to leave their previous knowledge and are unwilling to make a new start."

Seventh, when talking about suggestions for Chinese phonetics teaching and learning, 9 native Chinese teachers and 3 Hungarian teachers all mentioned that the basic skills in the initial stage of phonetics teaching are very important. If students do not learn phonetics well in the initial stage, improving vocabulary in the later stage will not help well. In addition, 7 native Chinese teachers and 7 Hungarian teachers emphasized that individual differences of students have a greater impact on Chinese phonetics learning, including personal language talent and degree of effort.

Finally, it is worth noting that two teachers said they were not satisfied with the existing teaching materials. For example, one teacher said: "In terms of teaching materials, ... I hope there will be Hungarian-specific phonetics teaching materials, so that teachers do not need to accumulate many years of experience to understand the problems of Hungarian students." Another teacher mentioned that although she had received training from the Language Integration Center before coming to Hungary, it did not help her with Chinese phonetics teaching in Hungary. She believed that it was necessary to provide Chinese native-speaking teachers with comparative training on Chinese-Hungarian phonetics.

6.2.2 Discussions

Table 20 presents findings relevant to Research Questions 3 and 4. It was found that the teachers' views on the Hungarian learners' ability to pronounce Chinese vowel finals was different to some extent from the learners' beliefs on their own difficulties to pronounce Chinese vowel finals. Based on the results of the questionnaire, all vowel finals had a median score which was less than 3. Thus, generally Hungarian learners of Chinese do not think

Chinese vowel finals are difficult to pronounce. However, their teachers mentioned that [ɤ] is easy to be mixed with [ʅ] or [ʯ] by their students, round vowels [uo] and [ou] are always mixed by the learners, and the students generally pronounce [ou] as [uo] etc. (Table 20).

Table 20. Beliefs of difficulty between Chinese teachers and Hungarian students

	Chinese teachers	Hungarian students
[ɤ]	15/20 = 75%	19/107 = 18%
[ʅ]	8/20 = 40%	20/107 = 19%
[ʯ]	8/20 = 40%	18/107 = 17%
[ou]	14/20 = 70%	21/107 = 20%
[y]	4/20 = 20%	5/107 = 5%

The comparison in Table 20 is just for visualizing the results, and not for direct mathematical comparison, because the number of the participants is extremely different. The data in Table 20 reveal a disparity between Chinese teachers' perceptions and Hungarian learners' experiences regarding the difficulty of specific Chinese vowel finals. This discrepancy suggests important implications for pronunciation instruction and learner awareness in second language acquisition.

Most strikingly, 75% of the Chinese teachers identified the vowel [ɤ] as problematic for Hungarian learners, whereas only 18% of learners themselves reported difficulty with this sound. The views of [ɤ] from Chinese teachers are in line with the previous literature review. Hungarian learners of Chinese do not think Chinese vowel finals are difficult, but they treat [ɤ], [ʅ] and [ʯ] as the same difficulty group. The retroflex vowels [ʅ] and [ʯ] were identified by 40% of teachers as problematic, while 19% and 17% of learners, respectively, considered them difficult. This suggests slightly closer alignment between teacher observations and learner perception, though a gap remains. Some teachers mentioned that [ʅ] and [ʯ] are reported to be mixed with [ɤ], so this result strengthened the results of Juhász (2020). She states that Hungarians produced the [ɤ] in a more acoustically palatalized manner than Chinese native speakers, in other words, they may mix the mid vowel [ɤ] and the high vowels [ʅ]/[ʯ]. In case of round vowels, 70% of teachers viewed the diphthong [ou] as a challenge for Hungarian students, yet just 20% of the learners perceived it as such.

Overall, the findings reveal a noticeable mismatch between learners' metalinguistic awareness and their actual pronunciation difficulties. These discrepancies highlight a gap between subjective learner beliefs and the more systematic errors observed by teachers, indicating the need for more explicit instruction.

In case of Research Questions 5, several factors may contribute to the difficulty of specific Mandarin Chinese vowel finals.

First, Hungarian learners of Chinese consistently struggle with Mandarin vowel finals that lack equivalents in their L1 phonological inventory, such as [ɤ], [ɿ], [ʅ] and diphthong [ou]. According to CAH, learners transfer the phonological forms from their L1 to the L2, and errors occur when the target language features are dissimilar from those of the native language. The absence of [ɤ], [ɿ], and [ʅ] in Hungarian explains the frequent substitution with more familiar vowels like [e:], [ɛ], or [i], which were reported by teachers. And the frequent confusion between [ɿ]/[ʅ] and Hungarian [i] is a clear case of negative transfer (Selinker, 1983), where learners map unfamiliar L2 phonemes to the nearest L1 category. These examples confirm the Contrastive Analysis Hypothesis prediction that phonological divergence between L1 and L2 increases acquisition difficulty (Lado, 1957), especially at the initial stages (Major, 1987; Munro, 1993).

While CAH highlights L1-L2 differences, the Markedness Differential Hypothesis (Eckman, 1977) deepens this view by explaining which differences are likely to be harder to acquire. The MDH posits that more marked (less common, more complex) phonological structures are more difficult for learners to acquire, especially when they do not exist in the L1. Mandarin [ɤ], a mid-back unrounded vowel, is considered typologically marked because most back vowels are rounded (Ladefoged & Maddieson, 1996). Similarly, [ɿ] and [ʅ] are also marked due to their apical articulations, and they lack direct counterparts in either IPA or most world languages — including Hungarian. Thus the difficulty with [ɤ], [ɿ], and [ʅ] may be not only due to L1 absence, but also to their intrinsic phonological markedness, as predicted by MDH.

In addition to phonological difficulty, orthography played a key role. According to Bassetti (2008) and Erdener & Burnham (2005), orthographic depth and transparency affect how learners form phonological representations. Hungarian has a transparent orthography, with consistent grapheme-to-phoneme correspondence. Pinyin, by contrast, is relatively opaque: one grapheme <i> may represent [i], [ɿ], or [ʅ], and <u> may represent [y] or [u]. These mismatches cause orthographic interference, learners expect one-to-one grapheme-sound mappings, leading them to misinterpret <i> as [i] and <u> as [u] across the board, and some rely heavily on orthography for pronunciation cues, especially in written-focused classroom environments. This is supported by Perre & Ziegler (2008) and Cheung et al. (2001), who found that orthographic input can shape or even distort L2 phonemic awareness, especially when the learner's L1 uses a transparent system.

Third, the interviews revealed that instructional approach plays a crucial role in either mitigating or reinforcing these difficulties. Teachers who explicitly taught phoneme-grapheme-phonetic correspondences (including IPA or contrastive analysis) enabled learners to develop more accurate mental representations. Instructors relying solely on imitative practice or delaying correction contributed to fossilization of errors, especially where orthographic cues were misleading. The lack of Hungarian-specific phonetics teaching materials and insufficient training in contrastive phonology between Chinese and Hungarian emerged as pedagogical gaps. In addition, time constraints in non-degree settings (e.g. Confucius Institutes) result in reduced phonetic training. Instructional variation in phonetic focus, feedback frequency, and teacher expectations directly impacts learners' mastery of challenging sounds.

As for learner variables, while there's disagreement among teachers, some believe knowing more languages helps students develop phonological awareness and metalinguistic strategies. However, others argue it can lead to interference or resistance to new phonetic systems. Additionally, individual differences such as language aptitude and learning effort were repeatedly cited as strong predictors of success. This underscores that learner-internal factors significantly mediate the difficulty of vowel finals.

Thus, the difficulty of specific Mandarin vowel finals for Hungarian learners stems from an interplay of phonological transfer, orthographic confusion, instructional practices, learner characteristics, and institutional factors. Addressing these challenges requires a multifaceted pedagogical approach, including explicit instruction, more focused correction strategies, improved collaboration between native and non-native instructors, development of teaching materials designed for Hungarian learners and greater awareness of individual learner differences and prior linguistic experience. The teachers' attitudes towards students' pronunciation errors reflect a shift from the Nativeness Principle—which emphasizes native-like pronunciation as the goal—to the Intelligibility Principle, which prioritizes effective communication (Levis, 2005). The current shift away from native-like models is beneficial in reducing unrealistic expectations, but it should not come at the expense of overlooking persistent, intelligibility-reducing errors. Pronunciation instruction should not only prioritize intelligibility and communicative effectiveness but also ensure that learners develop accurate phonological representations. Integrating both teacher assessments and learner self-perceptions into pronunciation pedagogy is therefore critical. Teachers may need to offer more explicit corrective feedback and raise learner awareness of subtle but systematic pronunciation issues. These insights not only help interpret learner performance in perception and production tasks but also inform more effective pronunciation instruction in L2 Chinese classrooms.

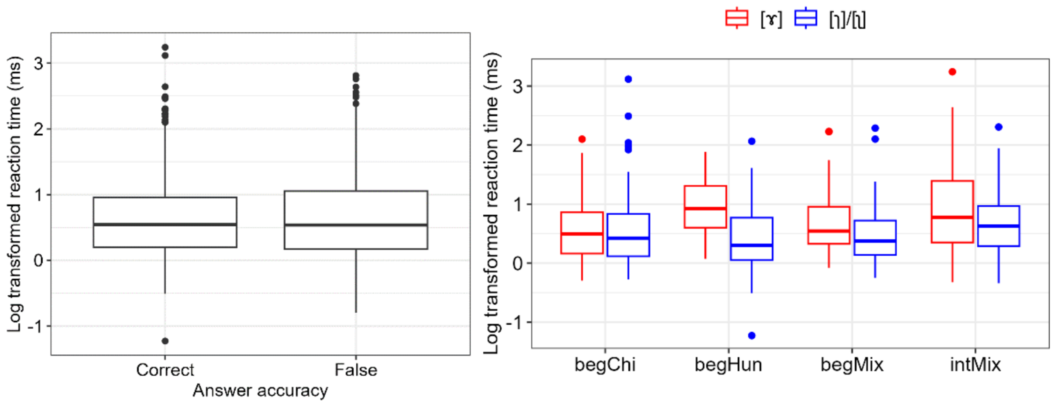
6.3 Empirical results about perception

6.3.1 Results

First, the reaction times for the correct answers were attested by generalized mixed models with Generalized Mixed Effects Logistic Regression. The best fitting model was the one including the interaction of the vowel and the group and random slope on the vowels by learners and a random intercept by the stimuli. The results did not indicate significant differences by any of the factors or their interaction, however, there was a tendency that the reaction times for the correct answers for <i> [ɿ]/[ʅ] were shorter than for <e> [ɤ] (Figure 9. right). Accordingly, the best fitting Generalised Mixed Effects Logistic Regression included the interaction of the vowel and the group and 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, the reaction time is not consulted in the later results.

Figure 9. The reaction times (log-transformed, ms)

Left: Reaction times for the correct vs. false answers, right: Reaction times for the correct answers for [ɤ] vs [ɿ]/[ʅ] among the learner groups.

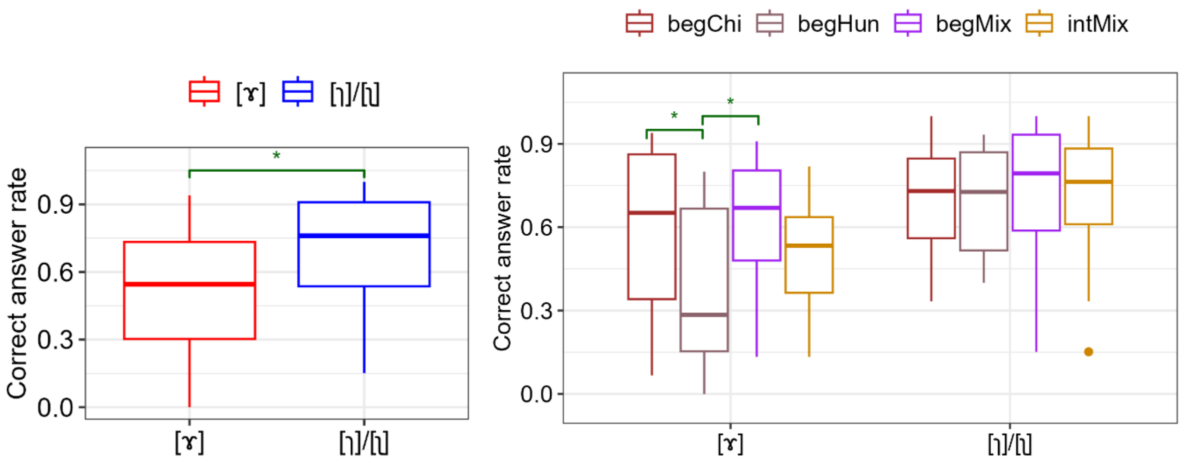


Then, the accuracy rate was calculated speaker-by-speaker for the mid and the high vowels separately. The ratio of correct answers revealed a noticeable difference in performance when identifying the two vowel phonemes (Figure 10, left). The results were grouped by the vowels and the groups. Mean is not used in present research because the data are not normally distributed, thus median and interquartile range are used for description. The best fitting model was the one that included the interaction of the vowel and the group with 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 factor vowel: $\chi^2(1, 1302) = 64.99$, $p < 0.001$) (Figure 10, left). 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 are lower than that of <i> [ɿ]/[ʅ]. 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 10. The accuracy rate for the mid and high vowels.

Left: The accuracy rate for the mid vs. high vowel regardless the learner group. Right: The accuracy rate for the vowels within the learner groups.



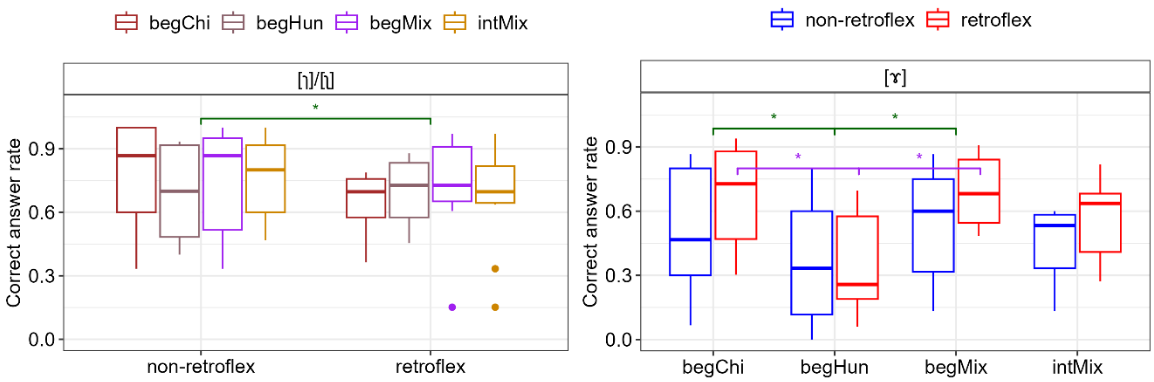
The interaction of the two factors was also significant ($\chi^2(3, 1302) = 13.38$, $p = 0.004$) (Figure 10, right). According to the Tukey post hoc test there is a significant difference (H1) between the begChi and begHun, and also (H2) between the begMix and begHun groups' results in the case of <e> [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) have significantly higher accuracy rates than the begHun group does (with a native Hungarian teacher) for [ɤ] sound. Comparatively, the accuracy rate of <e> [ɤ] in begHun is 0.30, while it is 0.60 and 0.64 in begChi and begMix respectively. As for apical vowels, the accuracy rates of the begChi and begMix (with a native Chinese teacher) also higher than the begHun group (with a native Hungarian teacher). Comparatively, the accuracy rate of <i> [ɿ]/[ʅ] in begHun is 0.67, while it is 0.77 and 0.75 in begChi and begMix respectively.

Figure 10. also shows that the intermediate (intMix) group's results did not show significant differences from any of the beginner groups, which means that their results were not significantly better even intMix has more experience with Chinese language.

Furthermore, the possible influence of tongue tip position (retroflex or not) is shown in Figure 11. 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 all along the vowel duration. Based on the ultrasound study of Lee-Kim (2014), both the tongue tip raising and tongue back retraction of [ɿ] and [ʅ] are maintained throughout the entire syllable. The accuracy rate of the non-retroflex vowel [ɿ] is higher than that of the retroflex vowel [ʅ]. While this seems to be a difference within the groups with a native teacher, the best fitting generalized linear mixed model was the one that included only the factor of retroflexion, but not the group, and included a random intercept by the learner. There was a significant difference between the retroflex and non-retroflex vowel ($\chi^2(1, 651) = 9.536$, $p = 0.002$).

In the case of the mid vowel, 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 begChi and begHun, and between begHun and begMix, where the begHun group has lower accuracy 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 cleared 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 significance level in the non-retroflex context.

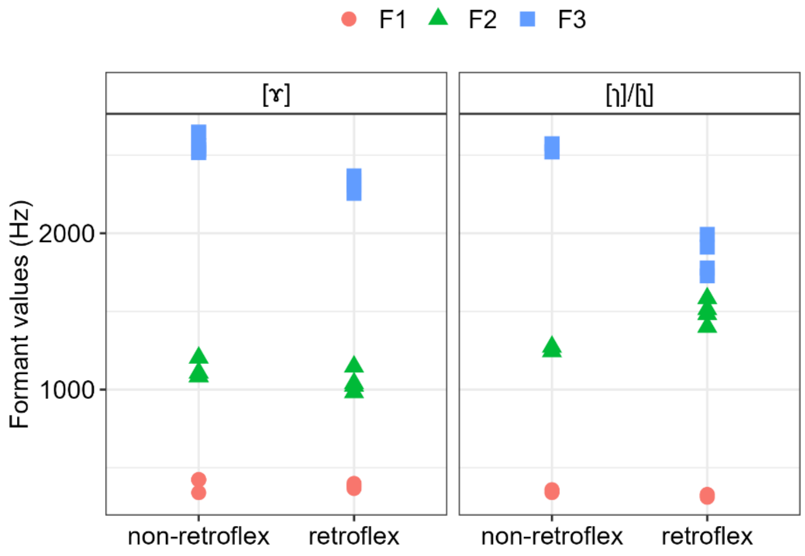
Figure 11. The accuracy rate of [ɿ]/[ʅ] (left) and [ɤ] (right)



Finally, 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.

Figure 11 shows the formant frequencies for the test items (Appendix D shows the mean values). The present data lies in 14 items in Table 16 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 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 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 considerably lay closer to each other in the present stimuli in the non-retroflex scenario than in the retroflex scenario, in correspondence with Lee & Zee's data (2001). We assume that phonetic context may influence vowels perception of Hungarian learners of Chinese. As the results in Figure 11. showed, the correct identification for [ɤ] shows higher differences across the learner groups in retroflex context than in non-retroflex one, however it is the other way for the /i/ allophones. The retroflex context resulted in somewhat higher accuracy 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 the specific question.

Figure 12. The mean formant frequencies (Hz) and their values in the test items



6.3.2 Discussion

First, to answer Research Question 6, the results of the perception test showed that the accuracy rates of the mid vowel were lower in all listener groups than that of the high vowels, which suggests that <e>[ɤ] is misperceived as <i> [ɿ]/[ʅ] more often than <i> [ɿ]/[ʅ] is misperceived as <e>[ɤ] for Hungarian learners of Chinese. These results are in correspondence with Juhász's (2020) results in the production domain. We assume that beside 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 like orthographic interference and phonetic context influencing the acquisition of these segments.

Second, to answer Research Question 7, the intermediate group did not have better results than the beginner groups in our mid-high differentiation task. As discussed in section 3, orthographic input plays an important role in second language acquisition. More experienced, i.e. intermediate learners of Chinese get longer time of more orthographic input, while the amount of focused pronunciation (and thus perception) training decreases. Generally Chinese pronunciation practice is limited to 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 the time goes by everyday communication does not require this efficient distinction between these speech sounds, thus, if the mental discrimination of the categories are not established in the beginning in the L2 learners mind, it is likely that advanced learners face difficulties when trying to tell them apart -- because the lack of "high-quality and well differentiated" input is getting more and more absent to help discriminating these sounds. 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 learners mind persists and fossilizes, 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 opaquer than the Hungarian orthography – as mentioned above.

Third, to answer Research Question 8, the present results showed that the 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 without her, which also implies that Hungarian teachers of Chinese have accent and possibly also have a difficulty in discriminating these sounds in production. However, based on the results of the phonetic context for vowel perception, it is interesting that BegHun show the opposite results compared to the other groups in Figure 11: [ɤ] after non-retroflexes [ts, ts^{h}, s] induce higher identifying rate than [ɤ] after retroflex [tʂ, tʂ^{h}, ʂ, ɻ] in groups of BegHun. In addition, [tsɿ, ts^{h}ɿ, sɿ] also induce higher identifying rate than [tʂʅ, tʂ^{h}ʅ, ʂʅ, ɻʅ] in groups of BegHun. This result suggested that the experience with a native teacher may have an impact on identification.

Fourth, based on Figure 4 and Figure 8, [ɤ] and [ɿ] vowels are situated closer in terms of their formant values compared to the distance between [ɤ] and [ʅ] vowels. We expect that identifying [ɤ] and [ɿ] vowels after non-retroflex [ts, ts^{h}, s] could induce lower accuracy rate, compared to the [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 lower identifying rate than [ɤ] after retroflex [tʂ, tʂ^{h}, ʂ, ɻ] in groups of BegMix, BegChi and intMix, while [tsɿ, ts^{h}ɿ, sɿ] do not induce lower identifying rate than [tʂʅ, tʂ^{h}ʅ, ʂʅ, ɻʅ] in groups of BegMix, BegChi and intMix. This may be caused by the inherent difficulty for retroflex perception (Tabain et. al., 2020). Thus, to answer Research Question 9, the consonant context may affect the perception of Chinese vowels. In syllables [tʂʅ, tʂ^{h}ʅ, ʂʅ, ɻʅ], both the consonant context and the apical vowel are retroflex, which is perhaps the major contributor to their difficulties. Since 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].

6.4 Empirical results about production

6.4.1 Formant spaces

The formant spaces (F2-F1 and F3-F1 in Z-scores, 67% = mean ± ~1sd) of the native speakers are shown in Figure 13. The mid vowel and the high vowels can be separated based on the first two formants, while the two high vowels based on the second and first formants in the raw data. Considering the L2-learner groups (Figure 14), we can expect that the mid and high vowels can be separated in the begChi and begHun groups similarly to the native speakers' vowels, while the begMix and intMix groups show larger overlap between the mid and high vowels. Taking a look at the two high vowels of the L2-speakers, we cannot expect these to be separated based on F2, and to be separated to a lower extent based on the F3, that seems to be more reliable in the begMix and intMix groups (in a contrast to the mid vs. high vowels' results).

Figure 13. The distribution of the three target vowels in case of native Chinese speakers (Z-score values)

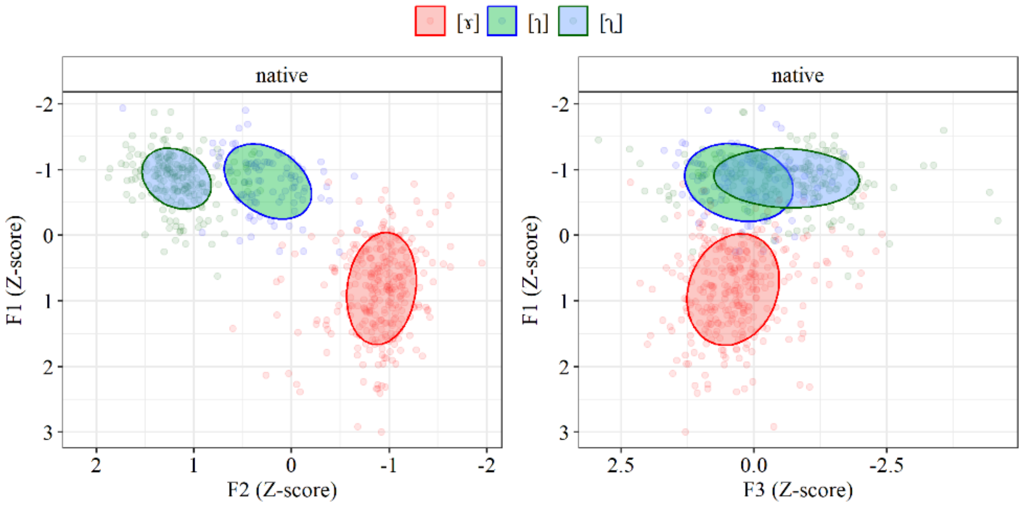
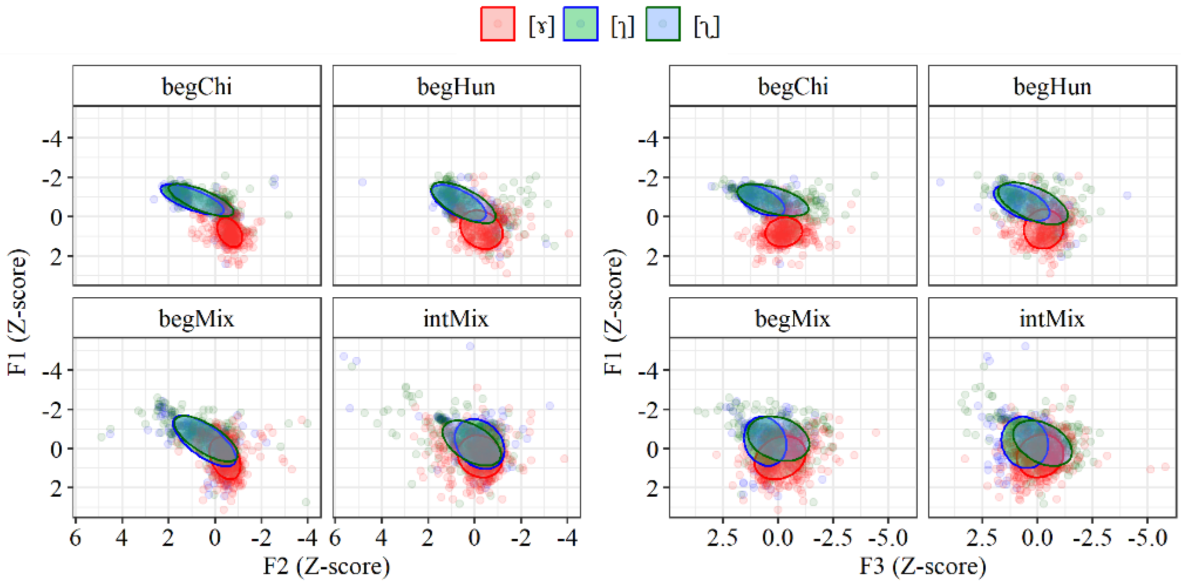


Figure 14. The distribution of the three target vowels in the L2-learner groups (Z-score values)

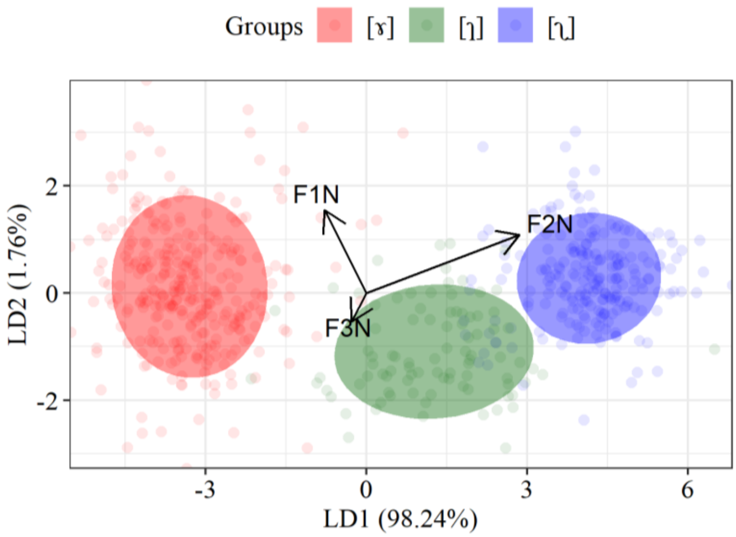


Based on the formant data, one can expect that during the classification based on the algorithm taught on the native speakers data, (i) the mid vowel will have higher true positive and lower false positive ratios in the begChi and begHun groups compared to the begMix and intMix groups, and that (ii) the high vowels will be less successfully classified in the begHun and begChi groups, than in the other two. Also, (iii) the overall classification rate for the mid vowel is expected to be higher than the high vowels.

6.4.2 LDA-results

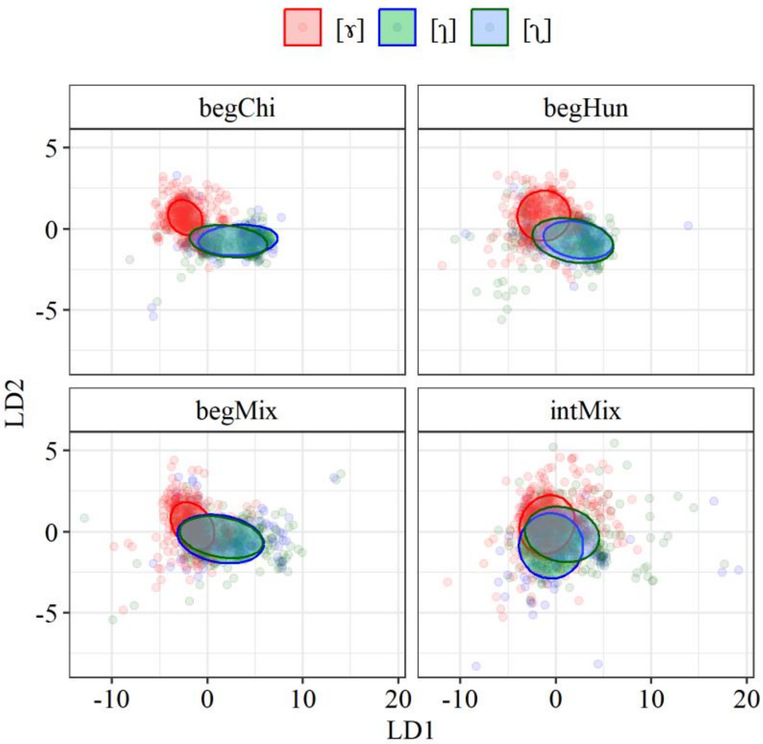
The distribution of the three target vowels were 50% ([ɤ]), 14.7% ([ʅ]), and 35.3% ([ɿ]) in the native speakers' dataset. Two linear discriminants (LDs) were calculated, as we used the three formants. While LD1 mostly relied on F2, less on F1, and only in a small ratio on F3, LD2 relied mostly on F1, and F2, and less on F3. LD1 explained the 98.2% of the training dataset, and LD2 explained the remaining 1.8% (Figure 15). And in case of native Chinese speakers, the discrimination of the three vowels reached a true positive ratio of 95.6% for all the realisations ([ɤ]: 98.4%, [ʅ]: 88.9%, [ɿ]: 94.4%). The false positive ratios were higher between the two high vowels (7.8%, 5.6%), than between any high and the mid vowel (1.6%, 3.3%, 0%) (Figure 16).

Figure 15. The classification of the three vowels in the native group



The prediction for the test dataset (Figure 16)., i.e., for all non-native data regardless of the learner group reached an overall true positive ratio of 54.3%, which is much lower than in the case of the native speakers

Figure 16. The discrimination of the three vowels in the L2-learner groups



The classification of the three vowels reached a true positive ratio of 54.3% for all the realisations ([ɤ]: 71.8%, [ʅ]: 26.9%, [ʅ]: 40.9%, Figure 16). These results mean that while there was a lower ratio of correct classification for the L2-learners' vowels in general, there was a difference regarding the vowel height. The mid vowel was more successfully classified than the high vowels. The 28.2% of the mid vowels were classified into any of the high vowels, and the 32.9%, and 30.5% of the high vowels were classified as mid. The misclassification between the two high categories was also higher: 28.6%, and 40.2%.

Figure 17. The distribution of the three target vowels in non-native groups

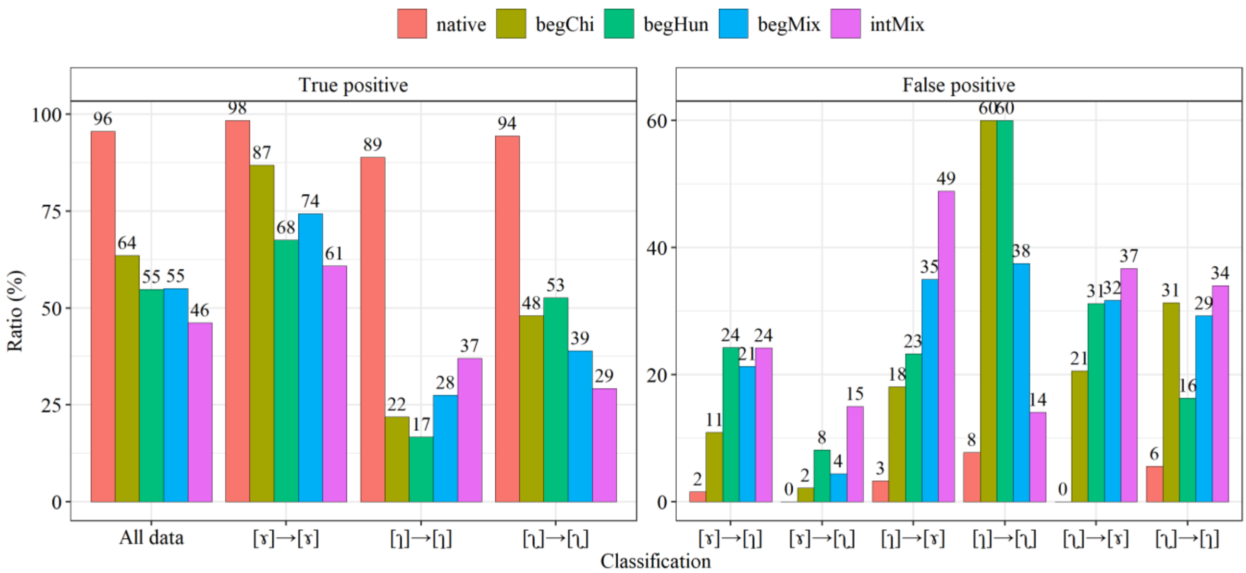
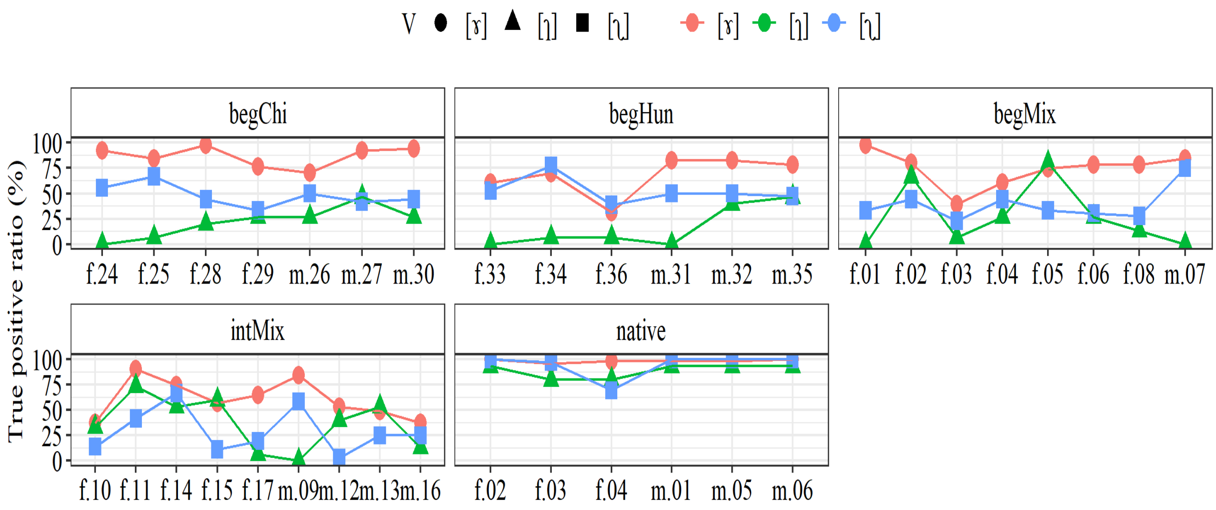


Figure 16 shows the true positive and the false positive ratios separated for the speaker groups in order to see whether there can be found any differences among the groups. A general tendency can be seen that the mid vowel is most correctly classified in all L2-leraner groups, while the non-retroflex high vowel hardly reaches the chance level. The correct classification of the retroflex high vowel shows larger difference across the groups, native speakers have high correct classification (94%), and L2 learners groups show much more variability, ranging from begHun (53%) to intMix (29%). But in terms of performance, the retroflex vowel's classification tends to fall in between the better-recognized mid vowel and the poorly recognized non-retroflex high vowel. The false positive ratios indicate that if the mid vowel is misclassified, it is classified into the non-retroflex high vowel's group in most cases. The non-retroflex high vowel is most often misclassified as the retroflex vowel in the begHun and begChi groups (60%), while as the mid vowel in the intMix group (49%). The begMix group's production of this vowel is classified at a change level into any categories. While the high retroflex vowel is correctly classified above the chance level in most L2-learner groups, their misclassification can be considered mostly at a chance level into any of the other two vowels, except for the begHun group, where it is mainly classified into the mid vowel. Altogether we can see that the classification results not seem to vary based on the expected proficiency level of the learner groups.

6.4.3 Patterns of inter-speaker differences

While the group-level data confirms systematic challenges with specific vowel segments, increasing evidence points to significant individual variability in L2 speech learning—even among learners who share similar linguistic backgrounds. To address this, the present study also examines speaker-level differences in vowel production (Figure 18).

Figure 18. Patterns within individual differences



We can see that native speakers show consistently high a true positive ratios across all vowels, with minimal inter-speaker differences; a slight drop is observed in one speaker (nat.f.04) for the high vowels, though overall performance remains stable. But the speaker-wise true positive ratios show a wider range of variation in all L2 learner groups compared to native speakers. In the beChi group, there is a strong distinction in the mid vowel [ɤ], and among the high vowels, the non-retroflex [ɿ] shows lower or equal true positive values compared to the retroflex [ʅ]. In other groups, although the overall true positive values are high at the group level, individual speakers differ notably in their pronunciation of the mid vowel [ɤ]. For high vowels, the pattern in beChi and begHun is consistent, with [ɿ] showing lower or equal true positive values relative to [ʅ]. However, in the remaining groups, this distinction is more speaker-dependent and less consistent. Overall, in begMix and intMix, there is substantial individual variation across all three vowels—mid vowel [ɤ] and high vowels [ɿ] and [ʅ]. The inter-speaker variability implies that certain learners lack stable phonetic representations, leading to fluctuating or unpredictable pronunciation patterns. Notably, beginner groups—especially begChi and begHun—display less inter-speaker variability compared to the intermediate mixed group (intMix). In other word, L2 learner groups show wider inter-speaker variation than native speakers, with the greatest variability evident in the intMix group.

6.4.4 Discussion

Native Chinese speakers demonstrated high classification accuracy (95.6%), their data were used as teaching set, but the high accuracy rate was expected as they constitute the native speakers group. The Hungarian learners' data served as testing set in order to answer the question if the three vowels are pronounced by them with similar distinction to the native speakers. This means that the results show us if the vowels are distinguished in a similar way in their pronunciation to the native speakers, but if they use a non-native like (mislearnt) distinction, it will not be found in this method. Based on the results, the Hungarian learners averaged only 54.3%, with substantial variation across groups. Hungarian does not have the Chinese vowels [ɤ], [ɿ], or [ʅ], either phonemically or phonetically. This absence may contribute directly to the different rates in producing these vowels between native Chinese speakers and Hungarian learners of Chinese.

Among the three targeted vowels, the mid vowel [ɤ] achieved the highest true positive classification rate (71.8%). This relatively higher accuracy appears to stem from its clearer acoustic separability from the two high vowels. Specifically, despite the absence of [ɤ] in the Hungarian vowel system, its distinct position within the vowel space in Figure 4 and Figure 8, particularly relative to the high apical vowels, may render it more perceptually salient for learners. The higher classification success of [ɤ] seems not in agreement with the previous research. However, this higher classification success does not imply accurate production in absolute terms. The distribution of [ɤ] in the learners' formant space, as visualized in Figure 14, remains consistently misaligned with that of native speakers in Figure 13, most notably along the F2 dimension. This misalignment suggests that learners employ non-native articulatory strategies, likely substituting acoustic targets with the nearest available articulatory configurations from their first language. Among the three targeted vowels, the non-retroflex vowel [ʅ] got the lowest true positive classification rate (26.9%), and confusion between [ʅ] → [ɻ̩] reached 60%, 60% and 38% - in begChi, begHun and begMix respectively. These suggest that [ʅ] is the most difficult to distinguish for Hungarian learners of Chiense among the three vowels. The apical vowels [ʅ] and [ɻ̩] exhibit substantial mutual confusion, evidenced by both low true positive rates and frequent bidirectional misclassification. This result also answers Research Question 10, the sounds [ɤ], [ʅ], and [ɻ̩] present discrimination challenges for Hungarian learners of Chinese, and the most difficult vowel among the three targeted vowels to diffrenciate is the non-retroflex vowel [ʅ].

To answer Research Question 11, The observed inter-speaker differences, however, highlight the substantial variability in L2 learners' pronunciation performance, particularly within the intMix group. While native speakers maintain stable, near-ceiling a true positive ratios, and beginner learners—especially in begChi and begHun—display relatively consistent patterns, the intermediate learners demonstrate the widest range of outcomes. The increased variability in the intMix group may reflect diverging learning trajectories, fossilization, or a reduced emphasis on precise segmental pronunciation as learners progress. These findings emphasize the importance of analyzing inter-speaker differences rather than relying solely on group averages, which can obscure significant individual-level variation. Pedagogically, the relatively stable performance of beginner learners may result from more focused pronunciation instruction during initial stages of learning. As learners advance, however, the lack of sustained attention to phonetic detail might lead to less consistent production unless early distinctions are solidly established. This pattern is consistent with earlier production data from Zhang (2015), which documented difficulties in the production of Mandarin vowels among Hungarian learners of Chinese. Specifically, [ɤ] was often mispronounced as [ə], [e], [ɻ̩], or [ʅ]; [ɻ̩] was realized as [ə], [i], [ɥ], or [ʅ]; and [ʅ] was produced as [ə], [i], [ɥ], or [ɻ̩]. The current data reveal that the difficulties are not uniformly distributed across learners: individual speakers vary substantially in how accurately they produce these sounds. This inter-speaker variability supports the growing recognition in L2 phonetics that individual learner differences play a central role in acquisition outcomes. The Revised Speech Learning Model (SLM-r; Flege & Bohn, 2021) offers a useful framework for interpreting these findings, suggesting that factors such as auditory acuity, phonemic coding ability, and early-stage perceptual sensitivity influence the stability and precision of L2 phonetic categories. As such, the observed variability is likely shaped not only by L1 transfer but also by these learner-specific traits, which govern how phonetic input is processed and internalized during L2 development.

To address Research Question 12, the results reveal a noticeable influence of the teacher's L1 on learners' production accuracy. Among the L2 learner groups, the begChi group—taught by a native Chinese teacher—achieved the highest overall classification accuracy (64%), especially for the mid vowel [ɤ], with a remarkable 87% true positive rate and the lowest false-positive rates. This supports the idea that consistent exposure to native phonetic input enhances learners' ability to distinguish acoustically salient vowels, such as [ɤ], which do not exist in Hungarian. In contrast, the begHun group, taught exclusively by a native Hungarian teacher, showed moderate accuracy (55%), with their best result in the classification of the retroflex high vowel [ʅ] (53%). This may reflect a focus on explicit phonetic instruction, often emphasized by non-native instructors familiar with the learners' L1. The begMix group, taught by a combination of native and non-native instructors, also scored a moderate overall accuracy (55%), suggesting that mixed instructional input may offer both the benefits and drawbacks of each teacher. These findings suggest that the L1 background of instructors plays a measurable role in shaping learners' phonetic accuracy.

Surprisingly, some beginner groups outperformed the intermediate group, challenging the commonly held assumption that greater exposure to an L2 linearly correlates with improved pronunciation accuracy. The intMix group, despite having the longest exposure to Chinese (five semesters), achieved the lowest overall classification accuracy (46%). Even this group exhibited the best performance for [ɿ] (37%) among all learner groups, their overall performance does not align with the expected benefits of prolonged exposure. Thus to answer Research Question 13, the results do not reveal a consistent proficiency effect, which is in agreement with Juhász's (2020) production test results. This finding supports the the Speech Learning Model revised (SLM-r), which emphasizes that successful phonetic category formation depends not merely on cumulative exposure (Flege & Bohn, 2021). And as discussed in section 6.3.2, 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.

To answer Research Question 14, while Contrastive Analysis Hypothesis and Markedness Differential Hypothesis offer general predictions about difficulty in learning unfamiliar and marked L2 sounds like [ɤ], [ɿ], and [ʅ] for Hungarian learners of Chinese. 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 explain production difficulty in this context as well. Orthographic influence emerges as a significant factor contributing to pronunciation difficulties, particularly for the Chinese apical vowels [ɿ] and [ʅ]. Hungarian learners appear to rely heavily on the Pinyin spelling system, which fails to transparently encode the phonetic distinctions necessary for accurate production. Specifically, the Pinyin grapheme <i> simultaneously represents [i], [ɿ], and [ʅ], with the phonetic realization depending entirely on the nature of the preceding consonant. This ambiguous orthographic representation likely leads learners to erroneously map both [ɿ] and [ʅ] onto the Hungarian high front vowel [i], promoting systematic high-front substitutions. As a result, the critical articulatory differences between the canonical vowel [i] and the apical vowels are often ignored in learner production. A parallel, albeit less pronounced, orthographic effect is observed with the vowel [ɤ], represented by <e> in Pinyin. Hungarian learners may assimilate this to their native <e> grapheme, which corresponds to vowels such as [ɛ]. This misinterpretation reinforces inaccurate articulatory settings, leading to productions that are phonetically deviant but still sufficiently distinct to achieve partial acoustic separability from the high vowels. The impact of orthography on vowel production is evidenced in the formant distributions illustrated in Figures 12 and 13, where the vowel ellipses for learners display consistent misalignment with those of native speakers. Notably, the elevated F2 values for the targeted vowel productions suggest an articulatory pattern skewed towards a high front position—an outcome plausibly shaped by reliance on Pinyin orthographic conventions rather than perceptual input alone. These findings are consistent with prior research (e.g. Bassetti, 2008), which demonstrates that orthography can directly influence the formation of phonological representations in second language learners. When a single orthographic symbol encodes multiple phonetic realities, learners are prone to collapse these distinctions, integrating orthography-induced biases into both perception and production.

Furthermore, [ɿ] got the lowest true positive classification rate among the three vowels. This pattern highlights several primary sources of difficulty such as their acoustic proximity, particularly within the F1–F3 dimension, and the articulatory complexity of producing apical vowels, which require fine-grained tongue-tip control that is not utilized for vowel production in Hungarian phonology. Confusion between [ɿ] → [ʅ] reached 60% in both begChi and begHun, suggesting that the perceptual boundary between these two apical vowels remains especially difficult for learners, regardless of the instructor's nativeness.

7. General discussions

This research explored the acquisition of Chinese vowel finals by Hungarian learners, with a focus on the vowels [ɤ], [ɿ], and [ʅ], through a multi-method approach including a learner questionnaire, teacher interviews, a perception test and a production task. Each study offered a distinct perspective, and together, they form a comprehensive picture of the challenges learners face and the factors influencing vowel acquisition.

The questionnaire results indicated that Hungarian learners generally did not regard Chinese vowel finals as highly difficult, with median difficulty scores below 3 for all vowels. However, teachers identified specific vowels such as [ɤ], [ɿ], [ʅ], and [ou] as particularly problematic for learners, highlighting a mismatch between learner self-awareness and actual pronunciation difficulties. This discrepancy suggests that learners may lack metalinguistic awareness of their pronunciation errors, highlighting the importance of teacher insight for effective diagnosis of learner difficulties. The gap between learner and teacher perceptions underscores the need for more explicit pronunciation instruction, including targeted feedback and perceptual training to help learners recognize and correct persistent errors.

Findings from all four studies underscore the influential role of L1-L2 phonological differences, especially in line with the Contrastive Analysis Hypothesis (Lado, 1957) and Markedness Differential Hypothesis (Eckman, 1977). None of the three target vowels [ɤ], [ɿ], [ʅ] appear to be fully acquired by the Hungarian learners in the sample. The problematic vowels are absent from Hungarian and typologically marked, which increases their acquisition difficulty. However, the findings also suggest that native language effect and markedness donot fully explain the observed difficulties. For instance, [ɤ] was more difficult to be identified than [ɿ] and [ʅ] in perception task, but [ɤ] was distinguished more accurately than [ɿ] and [ʅ] in production task, despite all three being absent in Hungarian. This indicates that additional factors play a role in shaping learner performance.

Orthographic interference emerged as a major factor. Across all data sources, orthographic interference consistently contributed to misperception and misproduction: i) questionnaire results showed that learners treated phonologically identical forms (e.g. <o> and <wo>, both realized as [wo]) as different due to spelling differences. ii) interviews revealed that students associated Chinese vowel symbols with Hungarian orthographic conventions, e.g. mispronouncing [ɿ]/[ʅ] as [i] due to the <i>. iii) Perception data demonstrated that learners failed to distinguish between [ɤ] and [ɿ]/[ʅ]. iv) Acoustic data demonstrated that learners failed to distinguish between [ɿ] and [ʅ]—all represented as <i> in Pinyin—leading to production patterns overly reliant on vowel height (F1) rather than backness (F2). These results suggest that orthography may override auditory input, distorting learners' phonological representations and impeding accurate perception and production. The phonetic memory of Hungarian learners is not kept awaken 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. And the incorrect pronunciation may persist and become fossilised even after abandoning Pinyin. Therefore, pronunciation instruction must address the limitations and inconsistencies of Pinyin, emphasizing phonetic awareness and auditory training early and consistently throughout instruction.

The finding in perception results showed that the experience with a native language teacher may have a positive effect on Hungarian listeners' perception. The production study found that native vs. non-native teacher input influenced learner performance differently. Even the data show this tendency, that needs further investigation. Additionally, the decline in intermediate learners' performance highlights the lack of sustained pronunciation training in advanced courses. Since pronunciation instruction is often limited to early stages, learners may develop ingrained errors that persist without correction. Intermediate learners did not outperform beginners in perception or production tasks, suggesting that prolonged exposure without explicit phonetic training may lead to fossilization of mispronunciations. The findings align with the SLM-r's notion that phonetic category formation depends not only on input quantity but also input quality.

Furthermore, the SLM-r offers a useful framework for interpreting the gradual and highly variable acquisition process. While group-level trends point to the highest true positive classification rate (71.8%) of the mid vowel [ɤ] achieved, individual variability remained high—demonstrating that individual differences significantly shape phonetic outcomes. The questionnaire results show that some learners do have good awareness — they know what they struggle with. Others don't — they might misjudge what's hard for them. 14 out of 20 Chinese teachers emphasized that individual differences of students have a greater impact on

Chinese phonetics learning, including personal language talent and degree of effort. When interpreted through the theoretical frameworks provided by Järvelä (2006), Glaser (1972), Jonassen & Grabowski (1993), and Tomlinson et al. (2003), these findings strongly support a personalised instructional approach. Järvelä's emphasis on balancing personal needs with collaborative learning is particularly relevant: although group-level trends provide useful diagnostic insight, they cannot replace instruction tailored to individual learner profiles. Glaser's notion of "adaptive education" is also validated here, especially his call for focusing on modifiable cognitive processes rather than fixed aptitudes like IQ. The study's findings echo Jonassen and Grabowski's view that individual differences in aptitude, cognitive style, and learning profile directly influence instructional effectiveness. Tomlinson et al.'s advocacy for differentiated instruction further reinforces the need for teaching that responds to student readiness and perceptual strengths. The findings underscore that Chinese phonetic learning is shaped by both predictable patterns and unique individual differences. Therefore, I view personalised instruction as essential—not as individual tutoring in isolation, but as a responsive, adaptable framework that integrates diagnostic insight, scaffolds learner-specific challenges, and leverages personal aptitudes to maximise progress. Personalisation must attend not only to what learners struggle with, but how they perceive, process, and regulate their own learning, making it a cornerstone of effective, inclusive phonetic instruction. These results also reinforce the SLM-r claim that the development of new phonetic categories depends not only on input quantity but also on input quality, individual attention to phonetic detail, and the nature of the L1–L2 contrast.

Taken together, the studies reveal that vowel acquisition difficulty arises from a complex interaction of linguistic, cognitive, instructional, and perceptual factors. The acquisition of Chinese vowels by Hungarian learners is influenced by factors such as phonological distance and markedness of target vowels, orthographic opacity of Pinyin, instructional practices such as the balance of imitation vs. explanation, learner variables including language aptitude, motivation, and multilingual background. The findings call for a systematic, phonologically-informed, and personalized approach to pronunciation instruction. Explicit training, especially on marked sounds like [ɤ], [ɿ], and [ʅ], must begin early and continue through intermediate levels to prevent fossilization and to build accurate phonological representations. Orthographic transparency, targeted auditory practice, and high-quality teacher input—both native and non-native—must be central to this endeavor.

Future research should implement the perceptual assimilation test to quantify learners' perceptual mapping between L1 and L2 vowels. And future research should focus on longitudinal studies such as tracking learners over time to examine how perception and pronunciation evolve with different instructional approaches, neurocognitive investigations such as using EEG or fMRI to explore how orthographic and auditory processing interact in L2 phonological acquisition, and comparative studies such as examining how learners from other L1 backgrounds (e.g. English, Japanese) acquire Chinese vowels to disentangle universal vs. L1-specific challenges.

References

Barrientos, F. (2023). On segmental representations in second language phonology: A perceptual account. *Second Language Research, 39*(1), 259-285. Doi: 10.1177/02676583211030637

Barry, W. (1989). Perception and production of English vowels by German learners: instrumental- phonetic support in language teaching. *Phonetica*, 46(4), 155-168. Doi: 10.1159/000261840

Bassetti, B. (2008). Orthographic input and second language phonology. In: Piske, T. and Young Scholten, M. (eds.), *Input Matters in SLA* (pp. 191-206). Clevedon, UK: Multilingual Matters. Doi: 10.21832/9781847691118-013

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. Doi: 10.18637/jss.v067.i01

Battistella, E. (1987). On segment-structure conditions. *Linguistics, 25*(2), 383-394. Doi: 10.1515/ling.1987.25.2.383

Beck, M. W. (2024). ggord: Ordination Plots with ggplot2. R package version 1.1.8. Retrieved from: https://zenodo.org/badge/latestdoi/35334615

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words, 167*(224), 233-277.

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

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of phonetics, 20*(3), 305-330. Doi: 10.1016/S0095-4470(19)30637-0

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Munro, M. J. & Bohn, O.-S. (Eds.): *Language experience in second language speech learning: Inhonor of James Emil*

Flege (pp. 13-34). Amsterdam: John Benjamins Publishing Company. Doi: 10.1075/lllt.17.07bes

Best, C. T., McRoberts, G. W., & 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*(2), 775-794. Doi: 10.1121/1.1332378

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

Bohn, O. S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in second language acquisition*, *14*(2), 131-158. Doi: 10.1017/S0272263100010792

Bohn, O. S. (2002). On phonetic similarity. In Burmeister, P., Piske, T. & Rohde, A. (Eds.), *An integrated view of language development: Papers in honor of Henning Wade* (pp. 191-216). Trier, Germany: Wissenschaftlicher.

Brown, C. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research 14*(2), 136-193. Doi: 10.1191/026765898669508401

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In Archibald, J. (ed.), *Second Language Acquisition and Linguistic Theory*. Oxford: Blackwell.

Chao, Y. R. (1928). *Studies in the modern Wu dialect*s. Monograph 4. Tsinghua University Research Institute.

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

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chen, P. (1999): *Modern Chinese: History and Sociolinguistics*. Cambridge: Cambridge University Press. Doi: 10.1017/CBO9781139164375

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*(2), 371-405.

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

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

Cheung, H., Chen, H. C., Lai, C. Y., Wong, O. C., & Hills, M. (2001). The development of phonological awareness: Effects of spoken language experience and orthography. *Cognition*, *81*(3), 227-241. Doi: 10.1016/s0010-0277(01)00136-6

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).

Cook, V. (2016). Transfer and the relationships between the languages of multi-competence. In: Alonso, R. A. (Ed.): *Crosslinguistic influence in second language acquisition* (pp. 24-37). Bristol, UK. Multilingual Matters

Cook, V., & Bassetti, B. (Eds.). (2005). *Second language writing systems*. Clevedon, UK, Multilingual Matters.

Deme, A. (2025). *A magánhangzók változatossága a magyarban*: Eredmények a koartikulációs ellenállásról és a koartikulációs agresszióról a magánhangzók közötti koartikulációban artikulációs és akusztikai adatok alapján. Nyelvtudományi Értekezések 173. Budapest: Akadémiai Kiadó.

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

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

Eckman, F. R. (1977). Markedness and the Contrastive Analysis Hypothesis. *Language Learning, 27*(2), 315–330. Doi: 10.1111/j.1467-1770.1977.tb00124.x

Erdener, V. D. & Burnham, D. K. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, *55*(2), 191-228.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Utrecht: LOT.

Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. A. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America, 126*(3), 1379-1393. Doi: 10.1121/1.3180321

Fan, L. & Myintzu, W. (2022): A study on Hungarian Chinese learners' pronunciation for Chinese consonant initials. In: Simon S., Brdar-Szabó R. & Litovkina A. (eds)., *14th International Conference of J. Selye University. Language and Literacy Section. Conference Proceedings*. pp.25 -37. Komárno, Slovakia. Doi: 10.36007/4508.2023.25

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of* America 93(3), 1589-1608. Doi: 10.1121/1.406818

Flege, J. E. (1995). Second language learning: Theory, findings, and problems. In Strange, W. (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-276). Timonium, MD: York Press.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In Schiller, N. O. & Meyer, A. S. (eds.): *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319-355.). New York: De Gruyter Mouton. Doi: 10.1515/9783110895094.319

Flege, J. E., & Bohn, O. S. (2021). The revised speech learning model (SLM-r). In Wayland, R. (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3-83.) Cambridge: Cambridge University Press. Doi: 10.1017/9781108886901

Flege, J. E., & MacKay, I. R. (2004). Perceiving vowels in a second language. *Studies in second language acquisition, 26*(1), 1-34. Doi: 10.1017/S0272263104261010

Fox, J. & Weisberg, S. (2019). *An R Companion to Applied Regression,* Third edition. Thousand Oaks: SAGE Publications

Fuster-Duran, A. (1996). Perception of conflicting audio-visual speech: An examination across Spanish and German. In. Stork, D. G. & Hennecke, M. E. (Eds.): *Speechreading by Humans and Machines: Models, Systems, and Applications* (pp. 135-143.). Berlin: Springer Verlag.

Gao, Y. (2018). *匈牙利中学生汉语语音偏误分析(Analysis of Chinese Phonetic Errors for Hungarian Middle School Students – A Case Study of Beginners in Xantu Janos Middle School).* M.A. Dissertation, Beijing Foreign Studies University.

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

Glaser, R. (1972). Individuals and learning: The new aptitudes. *Educational Researcher, 1*(6), 5-13. Doi: 10.3102/0013189X001006

Halliday, M. A. K. (1990) *Spoken and Written Language*. Oxford: Oxford University Press.

Hartman, L. M. (1944). The segmental phonemes of the Peiping dialect. *Language*, *20*(1), 28–42.

Hashimoto, M. (1970). Notes on Mandarin phonology. In. Jakobson, R. & Kawamoto, S. (Eds.): *Studies in General and Oriental Linguistics*. (pp. 207-220). Tokyo: TEC.

Hauser, I. (2019). Coarticulation with alveopalatal sibilants in Mandarin and Polish: Phonetics or phonology?. In *Proceedings of the Annual Meetings on Phonology*. Doi: 10.3765/amp.v8i0.4688

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech communication*, *47*(3), 360-378. Doi: 10.1016/j.specom.2005.04.007

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones.* New York: Cambridge University Press.

Hu, F. (2024). The Descriptive Tradition in Chinese Phonetics. In *Oxford Research Encyclopedia of Linguistics.* Doi: 10.1093/acrefore/9780199384655.013.900

Järvelä, S. (2006). Chapter 2. Personalized Learning? New Insights into Fostering Learning Capacity. In. OECD: *Personalising Education* (p. 31-46). Paris: OECD Publishing.

Jia, Z. (2017). *匈牙利初级汉语学习者语音学习策略研究(Research on phonetic studying strategies of elementary Hungarian learners of Chinese).* M.A. Dissertation, Beijing Foreign Studies University.

Jonassen, D.H., & Grabowski, B.L. (1993). *Handbook of Individual Differences, Learning, and Instruction (1st ed.)*. New York: Routledge. Doi: 10.4324/9780203052860

Józsa, S. (1988). 几十年来匈牙利汉语教学的几点体会(Several ideas about Chinese teaching in Hungary). Selected Papers *of the 7th International Symposium on Teaching Chinese as a Foreign Language*: 653-654.

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*(2), 1-16. Doi: 10.18460/ANY.2020.2.007

Karlgren, B. (1915-1926). *Études sur la Phonologie Chinoise.* Upsala: K.W. Appelberg.

Kaye, J. (1995). Derivations and Interfaces. In Durand, J. & Katamba, F. (eds.): *Frontiers of Phonology: Atoms, Structures, Derivations* (pp. 289–332.). London: Longman.

Kiss, G. (1985). A magyar magánhangzók első két formánsának meghatározása szintetizált hangmintákat felhasználó percepciós kísérlet segítségével. *Nyelvtudomány Közlemények*, 87(1), 160-172.

Kohler, K. J. (1981). Contrastive phonology and the acquisition of phonetic skills. *Phonetica*, 38(4), 213-226. Doi: 10.1159/000260025

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850-11857. Doi: 10.1073/pnas.97.22.11850

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606-608. Doi: 10.1126/science.1736364

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software,* 82(13), 1–26. Doi: 10.18637/jss.v082.i13

Ladefoged, P. & Disner, S. F. (2012). *Vowels and consonants*. Oxford: Wiley-Blackwell.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lado, R. (1957*). Language across cultures*. Ann Arbor: University of Michigan Press.

Leather, J. (1999). Second-language speech research: an introduction’. In Leather, J. (ed.): *Phonological Issues in Language Learning* (pp. 1-58). Oxford: Wiley-Blackwell.

Leben, W. R. (1973). *Suprasegmental phonology.* Doctoral dissertation, Massachusetts Institute of Technology.

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

Lee, W., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association, 33*(1), 109-112. Doi: 10.1017/S0025100303001208

Lee-Kim, S.-I. (2014). Revisiting Mandarin 'apical vowels': An articulatory and acoustic study. *Journal of the International Phonetic Association, 44*(3), 261–282. Doi: 10.1017/S0025100314000267

Lehiste, I. (1988). *Lectures on language contact*. Cambridge, Massachusetts: MIT Press.

Lenth, R. V. (2021). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5. 5–1. 2021.

Levis, J. M. (2005). Changing Contexts and Shifting Paradigms in Pronunciation Teaching. *TESOL Quarterly, 39*(3), 369–377. Doi: 10.2307/3588485

Levy, E. S. (2009). On the assimilation-discrimination relationship in American English adults' French vowel learning. *The Journal of the Acoustical Society of America*, *126*(5), 2670-2682. Doi: 10.1121/1.3224715

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of phonetics, 36*(1), 141-157.

Li, D., Gao, J., & Wang, H. (2021). 匈牙利中文教学资源发展路径构建 (Construction of the development path of Chinese teaching resources in Hungary). *Journal of Yunan Normal University (Teaching & Studying Chinese as a Foreign Language Edition), 19*(4), 56-62.

Li, M. (2020). 一带一路'沿线国家匈牙利的汉语教学现状及发展策略研究 (Research on the Status and Development Strategy of Chinese Teaching in Hungary Along the Belt and Road). *International Chinese Language Education. 5*(2), 100-110.

Lin, S., & Wang, L. (2013). *语音学教程 (A Course in Phonetics)*. Beijing: Beijing university press. 2013. 280p.

Lin, Y. H. (1990). Prenuclear glides in Chinese. In: Prenuclear glides in Chinese. Paper Presented at the Mid-America Linguistics Conference, University of Kansas.

Lin, Y-H. (1989). *Autosegmental treatment of segmental processes in Chinese phonology*. Ph.D. dissertation, The University of Texas in Austin.

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

Loewen, S. (2009). Recasts in multiple response focus on form episodes. In A. Mackey & C. Polio (Eds.), *Multiple perspectives on interaction: Second language research in honor of Susan M. Gass* (pp. 176–196). New York: Routledge.

Major, R. C. (1987). Phonological similarity, markedness, and rate of L2 acquisition. *Studies in second language acquisition*, *9*(1), 63-82. Doi: 10.1017/S0272263100006513

Major, R.C. & Kim, E. (1999). The Similarity Differential Rate Hypothesis. *Language Learning*, *49*(s1), 151-183. Doi: 10.1111/0023-8333.49.s1.5

Markó A. (2017). Hangtan. In: Tolcsvai Nagy G. (szerk.): *Nyelvtan*. A magyar nyelv kézikönyvtára 4. (pp 75-206). Budapest: Osiris Kiadó.

Marlett, S. (2001). An introduction to Phonological Analysis. Summer Institute of Linguistics and University of North Dakota. Fall 2001 Edition.

McCarthy, J. J. (2007). Derivations and levels of representation. In: de Lacy, P. (ed.): *The Cambridge handbook of phonology*. (pp. 99-118) Cambridge: Cambridge University Press.

Mei, L. (1990). 匈牙利的汉语教学及罗兰大学的汉学 (Chinese Language Teaching in Hungary and Sinology at Eötvös Loránd University). *世界汉语教学 (Chinese Language Teaching in the World), 1990(2*), 127-128.

Mitterer, H., Reinisch, E., & McQueen, J. M. (2018). Allophones, not phonemes in spoken-word recognition. *Journal of Memory and Language*, *98*, 77-92. Doi: 10.1016/j.jml.2017.09.005

Mora, J. C., & Mora-Plaza, I. (2019). Contributions of cognitive attention control to L2 speech learning. In Nyvad, A. M., Hejna, M., Højen, A., Jespersen, A. B., & Sørensen, M. H. (Eds.): *A sound approach to language matters - In honor of Ocke-Schwen Bohn* (pp. 477-499). Aarhus: Dept. of English, School of Communication & Culture, Aarhus University.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and speech*, *36*(1), 39-66. Doi: 10.1177/002383099303600103

Neuberger, T., Gyarmathy, D., Gráczi, T. E., Horváth, V., Gósy M., & Beke, A. (2014). Development of a large spontaneous speech database of agglutinative Hungarian language. In: Sojka, P., Horák, A., Kopeček, I. & Pala, K. (Eds.): *Proceedings of TSD* (pp. 424-431). New York – Berlin – Heidelberg: Springer. Doi: 10.1007/978-3-319-10816-2_51

Neufeld, G. G. (1988). Phonological asymmetry in second language learning and performance. *Language Learning 38*(4), 531-559. Doi: 10.1111/j.1467-1770.1988.tb00166.x

Norman, J. (1988). *Chinese.* (Cambridge Language Surveys.). Cambridge: Cambridge University Press. Doi: 10.1017/S0022226700012196

Perre, L., & Ziegler, J. C. (2008). On-line activation of orthography in spoken word recognition. *Brain research*, 1188, 132-138. Doi: 10.1016/j.brainres.2007.10.084

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of Phonetics 18*, 375–94. Doi: 10.1016/S0095-4470(19)30380-8

Polka, L., & Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467-478. Doi: 10.1016/j.wocn.2010.08.007

Posit team (2025). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. Retrieved from: http://www.posit.co/.

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

R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: https://www.R-project.org/

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

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In Strange, W. (Ed.): *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 379-410). Timonium, MD: York Press

Saito, K., Sun, H., & Tierney, A. (2019). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. *Bilingualism: Language and Cognition*, *22*(5), 1123-1140. Doi: 10.1017/S1366728918000895

Schumann, J. (1979). The acquisition of English negation by speakers of Spanish: A review of the literature. In Andersen, R. (Ed.), *The acquisition and use of Spanish and English as first and second languages* (pp. 3–32). Washington, DC: Teachers of English to Speakers of Other Languages.

Sekiyama, K., & Tohkura, Y. I. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, *21*(4), 427-444.

Selinker, L. (1972). Interlanguage. *IRAL: International Review of Applied Linguistics in Language Teaching*, *10*, 209-231. Doi: 10.1515/iral.1972.10.1-4.209

Selinker, L. (1983). Language transfer. In Gass, S. & Selinker, L. (Eds.), *Language transfer in language learning* (pp. 33-68). Rowley, MA: Newbury House.

Simay, A. E., & Fan, L. (2020): A kínai nyelv magyarországi tanításának rövid története és jelene (Brief Historical Overview and Present State of Chinese Studies in Hungary). *Külügyi Szemle 19*(2), 3-20.

Siptár, P., & Törkenczy, M. (2000). *The phonology of Hungarian*. Oxford: Oxford University Press.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: does musical ability matter? *Psychological science*, *17*(8), 675-681. Doi: 10.1111/j.1467-9280.2006.01765.x

Stevens, K. N. & Keyser, S. J. (1989). Primary Features and their Enhancement in Consonants. *Language*, *65*(1), 81-106. Doi:10.2307/414843.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of phonetics*, *17*(1-2), 3-45. Doi: 10.1016/S0095-4470(19)31520-7

Stevens, K. N., Keyser, S. J., & Kawasaki, H. (1986). Toward a Phonetic and Phonological Theory of Redundant Features. In Perkell, J. and Klatt, D. (Eds.*): Invariance and Variability in Speech Processes* (pp. 426-463). New York: Psychology Press.

Stevens, K. N., Li, Z., Lee, C. Y., & Keyser, S. J. (2004). A note on Mandarin fricatives and enhancement. In Fant, G. (Ed.): *From traditional phonology to modern speech processing* (pp. 393-403.). Beijing: Foreign Language Teaching and Research Press.

Stockwell, R. P., Bowen, J. D., & Martin, J. W. (1965). *The grammatical structures of English and Spanish* (Vol. 4). Chicago: University of Chicago Press.

Strange, W. (1992). Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. In Tohkura, E., Vatikiotis-Bateson, E. & Sagisaka, Y. (Eds.): *Speech perception, production, and linguistic structure* (pp. 197-219). Tokyo: Ohmsha.

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

Strange, W. (2007). Cross-language phonetic similarity of vowels: Language experience in second language speech learning. In Bohn, O. S. & Munro, M. J. (Eds.): *Language experience in second language speech learning: In honor of james Emil Flege* (pp. 35-55). Berlin: John Benjamins Publishing.

Szende, T. (1994). Hungarian. *Journal of the International Phonetic Association*, *24*(2), 91-94. Doi: 10.1017/S0025100300005090

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*(4), 2745-2765. Doi: 10.1121/10.0001012

Taft, M. (2001). Processing of orthographic structure by adults of different reading ability. *Language and Speech*, *44*(3), 351-376. Doi: 10.1177/00238309010440030301

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., Conover, L. A., & Reynolds, T. (2003). Differentiating Instruction in Response to Student Readiness, Interest, and Learning Profile in Academically Diverse Classrooms: A Review of Literature. *Journal for the Education of the Gifted, 27*(2-3), 119-145. Doi: 10.1177/016235320302700203

Tusor, N. (2016). A kínaiul tanuló magyar anyanyelvűek tipikus kiejtési hibái. Szakdolgozat, *Eötvös Loránd Tudományegyetem, Budapest.*

Tyler, M. D. & Burnham, D. K. (2006). Orthographic influences on phoneme deletion response times. *Quarterly Journal of Experimental Psychology, 59*(11), 2010-2031. Doi: 10.1080/17470210500521828

Vago, R. M. (1980), *The Sound Pattern of Hungarian.* Washington: Georgetown University Press.

Van den Bosch, A., Content, A., Daelemans, W., & De Gelder, B. (1994). Measuring the complexity of writing systems. *Journal of Quantitative Linguistics*, *1*(3), 178-188. Doi: 10.1080/09296179408590015

Venables W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S, Fourth edition*. New York: Springer.

Wang, L. J. (2018), *语言学论文集 (Linguistics Papers)*. Beijing: 人民教育出版社 (People's education Press).

Wang, P. (2008). 北京话二合元音/ai/, /au/, /ia/, /ua/韵腹/a/的听感实验研究(A perceptual Experiment of the Nuclei /a/ of /ai/, /au/, /ia/, /ua/ in Beijing Mandarin). *Applied Linguistics. 3*(10), 98-107.

Wang, W. S. (1973). The Chinese language. *Scientific American, 228*(2), 50-63. Doi: 10.1038/scientificamerican0273-50

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.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2013). Age-related sensitive periods influence visual language discrimination in adults. *Frontiers in Systems Neuroscience, 7*, 86. Doi: 10.3389/fnsys.2013.00086

Wickham, H (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer Cham. Doi: 10.1007/978-3-319-24277-4

Wiese, R. (1997). Underspecification and the description of Chinese vowels. In J. Wang. & N. Smith (Eds.), *Studies in Chinese phonology* (pp. 219–249). Berlin: Mouton de Gruyter. Doi: 10.1515/9783110822014.219

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

Ye, Q. (2013): *Pronunciation Problems of Hungarian University Students Learning Chinese: Analysis, Causes and Possible Solutions* (Ph.D. Dissertation ELTE)

Ye, Q., & Bartos, H. (2017). The comparison of Hungarian and Chinese phonological systems: A pedagogical perspective. In. Kecskes, I. (Ed.): *Explorations into Chinese as a Second Language,* (pp 31-63.). Dordrecht: Springer International Publishing.

Zhang, H., Chen, F., Yan, N., Wang, L., Chen, Y., & Shi, F. (2016). The effects of tone categories on the perception of Mandarin vowels. In *10th International Symposium on Chinese Spoken Language Processing (ISCSLP)* (pp. 1-5). Tianjin: IEEE. Doi: 10.1109/ISCSLP.2016.7918413

Zhang, Z. (2015). *匈牙利大学生汉语语音偏误分析(Analysis of Chinese Phonetic Errors for Hungarian University Students – A Case Study of beginners in Karoli University).* M.A. Dissertation, Beijing Foreign Studies University.

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.

Appendix

Appendix A

Questionnaire to investigate the perception of Hungarian learners of Chinese for their ability to pronounce Chinese vowel finals.

Amikor a kínai nyelvet tanulja, akkor az alábbi hangok megtanulása általában mennyire szokott nehéznek bizonyulni? Kérem, minden piros hanghoz ill. hangkapcsolathoz adja meg az Ön tapasztalata szerinti nehézségi fokot számértékkel egy 1-7 skálán, ahol 1 = ’egyáltalán nem jelent nehézséget’, 2 = ’nagyon könyű’,3= ’kicsit könyű’,4 = ’semleges’,5= ’kicsit nehéz’,6= ’nagyon nehéz’,7= ’rendkívül nehéz’

Nem:	□Férfi □Nő
Mennyi ideje tanul kínaiul (hónap):	

When you learn Chinese, how difficult is it to learn the following sounds in general? For each red sound please choose the degree of difficulty based on your experience, the scale is from 1 to 7, where 1 = ’very, very easy’, 2 = ’very easy’,3= ’easy’,4 = ’not easy not difficult’,5= ’difficult’,6= ’very difficult’,7= ’very, very difficult’

Sex:	□Male □Female
How long have you learned Chinese (months):	

1.	b/p/g/k/d/t/zh/ch/z/c + **ɑ**	1	2	3	4	5	6	7
2.	b/p/m/f + **o**	1	2	3	4	5	6	7
3.	g/k/h + **e**	1	2	3	4	5	6	7
4.	b/p/g/k/d/t/j/q/y + **i**	1	2	3	4	5	6	7
5.	j/q/x/d/t/n/l + **ie**	1	2	3	4	5	6	7
6.	j/q/x/d/t/n/l + **iɑ**	1	2	3	4	5	6	7
7.	b/p/g/k/d/t/zh/ch/z/c/w + **u**	1	2	3	4	5	6	7
8.	d/t/n/l/g/k/h/zh/ch/sh/r/z/c/s + **uo**	1	2	3	4	5	6	7
9.	g/k/h/zh/ch/sh/r/z/c/s + **uɑ**	1	2	3	4	5	6	7
10.	j/q/x/l/n/y **+ ü**	1	2	3	4	5	6	7
11.	j/q/x/y **+ üe**	1	2	3	4	5	6	7
12.	n/l/g/h/zh/sh/b/p/m/f/z + **ei**	1	2	3	4	5	6	7
13.	d/t/n/l/g/k/h/zh/ch/sh/r/z/c/s + **ou**	1	2	3	4	5	6	7
14.	d/t/n/l/g/k/h/zh/ch/sh/r/b/p/m/f/z/c/s + **ɑi**	1	2	3	4	5	6	7
15.	d/t/n/l/g/k/h/zh/ch/sh/r/b/p/m/f/z/c/s + **ɑo**	1	2	3	4	5	6	7
16.	j/q/x/d/t/n/l + **i(o)u**	1	2	3	4	5	6	7
17.	j/q/x/d/t/n/l + **iɑo**	1	2	3	4	5	6	7
18.	d/t/g/k/h/zh/ch/sh/r/z/c/s + **u(e)i**	1	2	3	4	5	6	7
19.	g/k/h/zh/ch/sh/r/z/c/s + **uɑi**	1	2	3	4	5	6	7
20.	zh/ch/sh/r + **i**	1	2	3	4	5	6	7
21.	z/c/s + **i**	1	2	3	4	5	6	7

Appendix B

Sample interview questions for the semi-structured interviews with teachers

1. 请问，您在匈牙利教汉语教多长时间了？您是否会说匈牙利语？

How long have you taught Chinese in Hungary? And do you know Hungarian?

2. 您的学历及专业是什么？

What is your educational qualification and major?

3.您是怎么教授语音的？音节到声韵母到字母，还是声韵母到音节，还是字母到声韵母到音节？还是其他？为什么？

How do you teach Pinyin? On a syllabic basis, or from initials and finals to syllables, or from letters to initials and finals to syllables? Or any other way? Why? For example, “ge’ is a syllable, “g’ and “e’ are an initial and a final respectively.

4. 您觉得哪些元音韵母或者字母对学生来说最难，换句话说，学生最容易出错？为什么？您能具体描述一下吗？(比如找不到发音位置，混淆，或者发得不好)

Which vowel finals are the most difficult to teach, in other words, which are the ones learners always make mistakes? And why, how would you describe the problems and difficulties? (For example, they miss the articulation place, or they mix them up, or they pronounce something inappropriately)

5. 您对纠音怎么看？

Do you think the pronunciation errors by learners should be corrected? Why or why not?

6. 您觉得哪些因素会影响到学生的汉语语音学习？比如学生的专业、学生知道更多语言等。为什么？

Which factors could influence the students’ Chinese learning, particularly on their perception and pronunciation?

7. Do you teach Pinyin together with a Hungarian（Chinese） teacher? How to influence Pinyin learning of the students (perception and production)?

您和匈牙利老师或者中国老师一起教学生语音？您觉得会对学生的语音学习产生哪些影响(听辨与发音)？

8. Do you have any further suggestion on Chinese language teaching and learning for pronunciation?

对于汉语语音教学与语音学习您是否还有其他建议

Appendix C

请您听完这几个音节，然后判断您觉得他或者她的发音离普通话的远近。

Listen to the speech samples from six native speakers of Chinese, and evaluate how close the pronunciation is to ideal Mandarin Chinese variety.

	1 非常不好 Very bad	2 不好 bad	3 不好也不坏 not good, not bad	4 好 good	5 非常好 Very good
Speaker1					○
Speaker2				○	
Speaker3				○	
Speaker4				○	
Speaker5			○		
Speaker6					○