The perception and production of Mandarin citation tones by prelingually deaf adults

by

Yu Chen

B.A., Henan Normal University, 2003
M.A., Nankai University, 2006
Ph.D., Nankai University, 2011

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Linguisitics

University of Victoria

The perception and production of Mandarin citation tones by prelingually deaf adults

by

Yu Chen

B.A., Henan Normal University, 2003
M.A., Nankai University, 2006
Ph.D., Nankai University, 2011

Supervisory Committee

Dr. R. Sonya Bird, Supervisor
(Department of Linguistics)

Dr. M. John Archibald, Departmental Member
(Department of Linguistics)

Dr. Xiaohu Yang, Affiliate Member
(Department of English, Tongji University)

ABSTRACT

While considerable attention has been paid to the study of speech perception and production by prelingually deaf children, little is known about prelingually deaf adults' performance in such tasks many years after rebuilding hearing and using spoken language. For prelingually deaf people who have Standard Chinese as their target language, because tonal information is hard to process by hearing devices due to technological limitations, it is especially important to investigate whether they can perceive and produce Mandarin tones correctly. This dissertation aimed to contribute to this knowledge by investigating Mandarin tone perception and production by three participant groups, namely the HA (hearing aid) group, the CI (cochlear implant) group, and the NH (normal hearing) group, through three experiments—synthesized tone perception, lexical tone perception, and lexical tone production.

In the experiment of synthesized tone perception, we addressed whether prelingually deaf people could categorically perceive synthesized Mandarin tones, using identification and discrimination tasks. The results showed that while not all NH participants perceive all the synthesized tones categorically, the NH group surpassed the deaf groups both in tone identification and discrimination. Inside the deaf groups, the HA group performed better than the CI group: while a few HA participants could categorically perceive the synthesized tones, almost none of the participants in the CI group could do this. Thus, the results of this experiment might indicate that: first, Mandarin tones are not categorically perceived, or at most quasi-categorically perceived, even by NH people; second, it is hard to process fine-grained tone information even for prelingually deaf people who have abundant experience using spoken language; third, the CI devices do not convey the acoustic details needed to perceive tone as well as the HA devices do.

In the experiment of lexical tone perception, we checked the deaf participants' performance in identifying Mandarin lexical tones by analyzing their mouse movements during the decision-making process. The results showed that both of the deaf groups could identify Mandarin lexical tone with quite high degrees of accuracy (around 70%) under the distractions of competing tones and segmental distractors, indicating that the groups' performance in this experiment was much better than in the experiment of synthesized tone perception. In addition, the CI participants reached the same level as the HA participants in identifying Mandarin lexical tones, revealing that the CI participants were able to gather valuable tonal information indirectly from real human-produced speech sounds. However, the deaf groups still performed much worse than the NH group. The results showed that the deaf

participants were vulnerable to the effect of target tones and of rhyme complexity in the syllables that these tones were embedded in: they performed much worse in identifying T2T3 and T2T4 than other tone pairs, and in identifying tones in nasal rhymes, compared to other rhyme types.

In the experiment of lexical tone production, we investigated how deaf participants behaved in Mandarin tone production tasks using acoustic analysis and subjective assessment with multiple judges. The results confirmed that the deaf participants could produce Mandarin citation tones quite well, and the CI participants even performed a little better than the HA participants in tone production. Nevertheless, although the deaf groups had set up similar tone patterns as the NH group, the deaf participants still performed worse than their NH counterparts. Similar to the results of lexical tone perception, the deaf participants' performance was also impacted by target tone and rhyme complexity. In the current study, T2 and T3 (especially T3) were much harder to produce than T1 and T4 for the deaf participants. Tones were also harder to produce in syllables with nasal rhymes although the impact of rhyme complexity was not as obvious as in the lexical tone perception experiment.

Overall, the current study indicated that, while all three groups performed much better in perceiving and producing Mandarin lexical tones than in perceiving synthesized tones, the NH group performed much better than the two deaf groups in all three experiments; inside the two deaf groups, although the CI group performed much worse than the HA group in the synthesized tone perception experiment, the two groups performed similarly both in the lexical tone perception and the production experiments. Compared with the performance of the NH group, the performance of the deaf groups revealed that the acoustic characteristics of tones themselves, the types of rhymes the tones are embedded in, and the different hearing devices were important factors that impact the prelingually deaf adults' Mandarin tone perception and production. Under the theoretical framework of the speech chain theory, these results demonstrated that the underlying mechanisms responsible for prelingually deaf Mandarin-speaking adults' challenges in tone perception and production were their deficits of information processing in the three levels (the acoustic level, the physiological level, and the linguistic level) over the speech chain. That is, because the deaf participants could not clearly hear the acoustic signals associated with the tones (the acoustic level), they showed weaker mental representations for these tones (the linguistic level), and thus experienced more difficulties in realizing them both in perception and production (the physiological level).

From the results of the synthesized tone perception experiment, we can see that the

deaf groups, especially the CI group had difficulty perceiving synthesized tones categorically because they could not directly access tone acoustic information. In contrast, both deaf groups performed much better in the lexical tone perception and in the production experiments, indicating their hearing devices provided access to indirect acoustic information on tones and therefore unexpected benefits to these deaf participants in terms of tone perception. In particular, considering the CI device has worse performance in processing tone information than the HA device, and the CI participants have worse original hearing loss than their HA counterparts, the current results indicate that: (a) the CI device is a useful tool for rebuilding hearing and developing spoken language even when it comes to aspects of speech that one would not expect them to help with, and (b) other cues to tones exist beyond those assumed to be the most important for speech processing (duration, amplitude, overtones of rhymes, etc.). Future studies should explore what cues to tone are available to CI users, and to Mandarin speakers more generally, that help them acquire tonal structures in the language.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

When I start to input the following words, I know an epoach for me is going to end and a new one is coming soon! 14 years ago, when I came to Uvic to start my second Ph.D, I never thought it could be so long and so hard! Now, when I am looking back this wonderful journel of knowledge-pursuising, I know I owe too much to many people. Especially, I would like to thank:

my Supervisor Dr. Sonya Bird, for mentoring, support, encouragement, and patience. It was Sonya who gave me another opportunity to go on with my study when I almost decided to quit! Without your kindness and help, this dissertation would be a mission impossible!

my committee members Dr. John Archibald and Dr. Xiaohu Yang. In the past few years, they sacrificed a lot of time and provided precious support for this work. From their comments and suggestions, I have learned a lot to make this dissertation a better one!

all the participants invovled in the experiments. Because of the long experiment time and the lockdown due to the Pandemic, it is hard for the participants to follow the experiment procedures. Fortunately, almost all participants finished the experiments in time, which makes the study available. So, I would give my special thanks to all the participants for their cooperation, patience, and understanding.

my family members and close friends. My wife Yanting is always the right person to resort to when I have any difficulties in life; my little boy Xianlin is the source of happiness for me that I had never imaged before his birth; my parents, although they may don't know what I am doing and why I am doing, are always encouraging me to do what I want to do. Special thanks also go to some of my close friends, Hongge Cao, Weimin Ge, and Tianlin He, Your help has never been absent to support me in the low moments.

LI Bai, one of the greatest poets in Chinese history. His great poems had inspired me to keep moving in those sleepless nights and would continue to inspire me in the future. So, at last, I would like to share one of his best poem as the end of the acknowledgements.

Invitation to Wine
By Li Bai(701-762AD)

Translated by Yuanchong Xu Do you not see the Yellow River come from the sky,
Rushing into the sea and ne'er come back?
Do you not see the mirrors bright in chambers high,
Grieve o'er the snow-white hair though once it was silk-black?
When hopes are won,
Oh! Drink your fill in high delight,
And never leave your wine-cup empty in moonlight.
Heaven has made us talents, we're not made in vain.
A thousand gold coins spent, more will turn up again.
Kill a cow, cook a sheep and let us merry be,
And drink three hundred cupfuls of wine in high glee.
Dear friends of mine,
Cheer up, cheer up!
I invite you to wine.
Do not put down your cup!
I will sing you a song, please hear,
O, hear! Lend me a willing ear!
Do not care for bells and drums, rare dishes you take!
I only want to get drunk and never to wake.
How many great men were forgotten through the ages?
But real drinkers are more famous than sober sages.
The Prince of Poets feast'd in his palace at will,
Drank wine at ten thousand a cask and laughed his fill.
Why should a host complain of money he is short,
To drink with you I will sell things of any sort.
My fur coat worth a thousand coins of gold,
And my flower-dappled horse may be sold.
To buy good wine that we may drown the woes age-old.

Yu Chen

DEDICATION

To all deaf people that I met and will meet, taught and will teach, and worked and will work with!
Deafness is not an impairment, but a journey of miracle!

Chapter 1

Introduction

This chapter provides a general introduction to this dissertation that concerns prelingually deaf adult's Mandarin tone perception and production. First, we will introduce some background information related to the research topic. Then, we will state the research questions, the purpose, and the significance of this thesis. After presenting the general methodology of the current study, we will end this chapter by briefly discussing the structure of the thesis.

1.1 Background

1.1.1 The lexical tones in Standard Chinese

Standard Chinese is a tone language that has four lexical tones: T1 (level), T2 (rising), T3 (dipping) and T4 (falling) (Sun, 2006, p. 39). In this language, the four tones are used to differentiate word meanings. A well-known example is mā *mother*, má *hemp*, mǎ *horse*, and mà *scold*, in which the same CV syllable [ma] under different tones conveys different lexical meanings. Figure 1.1, adopted from a classic study on Mandarin tones by Xu (1997), displays the pattern of lexical tones in Standard Chinese. It is noted that Figure 1.1 displays mean F0 contours and mean durations of four Mandarin tones in the monosyllable /ma/ produced in isolation by 48 Mandarin native speakers.

From Figure 1.1, one can see that the four tones can be differentiated in two dimensions: pitch contour and duration. In Standard Chinese, each tone is associated with a particular shape and direction of the pitch trajectory: T1 is at the high end of the tone register and remains level throughout its duration; T2 starts at the middle of the register and rises throughout its duration[1]; T3 starts toward the lower range, dips even lower, and then

---

[1] There is dispute about the duration difference between T1 and T2. As we will discuss later, some studies

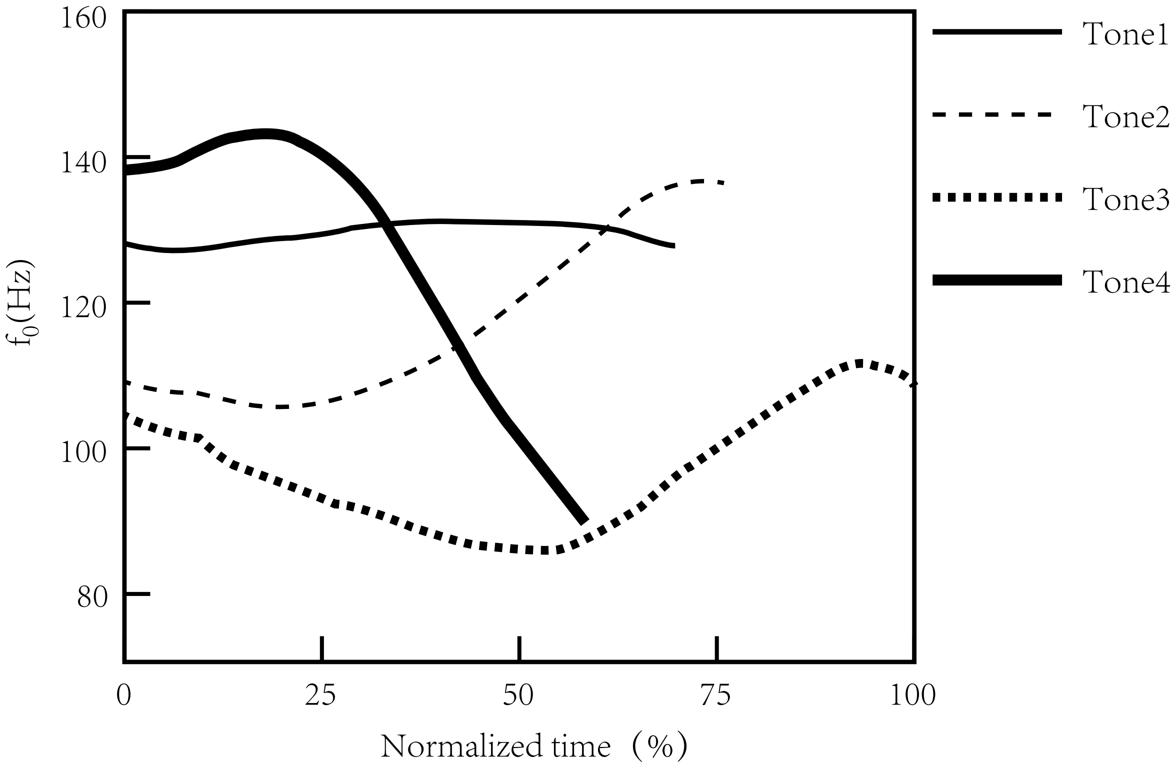


Figure 1.1: The Pattern of Lexical Tones in Standard Chinese

rises toward the end; T4 starts at the high end of the register and then shows a steep fall. Besides pitch contour, the four tones are also different in duration. For instance, according to Yu (2010), T4 is the shortest tone in Standard Chinese; T1 is a little longer than T4 but shorter than T2; and T3 has the longest duration among the four tones.

It should be noted that Mandarin is a typical contour tone language in which pitch contours are more important for distinguishing the four tones than duration and other acoustic parameters like intensity. Therefore, Mandarin speakers consider pitch contour as the primary cue for differentiating the four tones, and rely on this primary cue to produce and perceive these tones in most situations. However, in some other situations, such as when the primary cue is degraded, people may use secondary cues (mainly duration) to fulfill the tasks (Deroche et al., 2019).

---

reported that the duration of T1 is a little longer than that of T2. In our opinion, the two tones have no significant difference in duration, and the divergence might reflect the speakers' personal styles.

1.1.2 The basic terms of deafness

Deafness is a type of hearing impairment due to problems in the auditory system. With such an impairment, people may not be able to make use of sound signals for communicating effectively. According to the standard of hearing loss defined by the American Speech-Language-Hearing Association (see Table 1.1), if people can not hear a sound with a sound pressure level at or beyond 90 dB, they are deaf people; if peoples hearing loss is below 90 dB, they are hard of hearing people (Hersh and Johnson, 2003). In practice, people often use the term deafness to refer to hearing impairment at or above moderately severe hearing loss.

Table 1.1: Degree of hearing loss

Degree of hearing loss	Hearing loss range (dB HL)
Normal	0 to 15
Slight	16 t0 25
Mild	26 to 50
Moderate	41 to 55
Moderately severe	56 to 70
Severe	71 to 90
Profound	>90

Deafness is a complicated phenomenon that can be categorized from different angles. Based on the place of impairment, deafness can be categorized as conductive deafness (the deafness occurs in the outer or middle ear), sensorineural deafness (the deafness occurs in the inner ear), mixed deafness (a combination of conductive deafness and sensorineural deafness), and central deafness (the deafness occurs within the pathways of the central auditory nervous system) (Hersh and Johnson, 2003).

For those professionals who are extremely concerned about language development (e.g. linguists, speech rehabilitators, and educators), the timeline of language acquisition with respect to the onset of deafness is crucial. Thus, people in these fields often classify deafness into two types: prelingual deafness and postlingual deafness. In particular, prelingual deafness refers to deafness that occurs at birth or early in life, before speech and language development, and postlingual deafness is deafness that occurs after the development of speech and language (Kauffman et al., 2013). It is noted that experts differ in their views with respect to the dividing point between prelingual and postlingual deafness (Gleason and Ratner, 2017; Meadow-Orlans, 1987), they don't always provide a good rationale for their decisions. In this study, we don't need to concern with these discrepancies of the cut-off age of prelingual and postlingual deafness because the age of implantation is not a factor in our experimental design. Therefore, following Martines et al. (2013) and Li et al. (2014b), we will classify those who suffer deafness at or before five years old as prelingually deaf people, and those who suffer deafness afterward as postlingually deaf people.

After childhood and adolescence, prelingually deaf people would enter the adult stage. For these people, we call them prelingually deaf adults (PDAs for short in this study). In fact, PDAs are a highly heterogeneous population that can be divided into different subgroups. The PDAs who have no access to any hearing assistance and speech rehabilitation in life are significantly differentiated from those who have received hearing assistance. In the latter population, we can also differentiate three groups: the population who received CI (cochlear implant) at younger ages (early-implanters), the population who started to use HA (hearing aid) at younger ages and never switched to CI (constant HA users), the population who received HA at younger ages but switched to CI in adolescence or adulthood (delayed-implanters). It is noted that all three groups (early-implanters, constant HA users, and delayed-implanters) can access oral language at some level and thus are appropriate subjects of speech production and perception research. Because age of implantation is not a factor in the experimental design, we would not distinguish between early-implanters and delayed-implanters in the current study.

HA and CI are two completely different types of hearing devices: the HA device is basically a sound amplifying system to help the human cochlear "hear" the sound signal; in contrast, the CI device is a system which replaces the function of human cochlea by the electrode array on the implanted prosthesis. This fundamental difference between HA and CI leads to two outcomes. First, the two types of hearing devices are applicable to different deaf populations. That is, while HA is suitable for people who suffer mild to moderate sensorineural hearing loss and conductive hearing loss, CI is a better option (or only option in most situations) for deaf people with severe-to-profound hearing loss (often co-occurring with sensorineural hearing loss). Second, due to the two types of hearing devices have different mechanisms of sound signal processing, HA and CI users should "hear" totally different sounds after the same sound signals have been modulated by their hearing devices. Therefore, although both HA people and CI people access speech sounds after device using, they should be impacted by their hearing devices in speech perception and production, including tone perception and production. In particular, the CI device could not process pitch information directly as the HA device do, which should make CI people even harder in perceiving and producing tones.

1.1.3 The prelingually deaf population in China

Hearing impairment is a relatively common disorder around the world. According to the Ministry of Health, each year more than 30,000 babies are born deaf in China (Dai et al., 2006). From these data, one can see that China has a large deaf population that requires medical intervention, speech rehabilitation, and other assistance.

Unfortunately, however, the government's assistance was extremely scarce at the time of Dai et al. (2006)'s publishing. In 2007, China's healthcare spending amounted to less than 1% of the national GDP, placing it 144th among the 195 surveyed countries in the world in terms of health funding (Liang and Mason, 2013). Without the government's support, most families could not afford CI and the related long-term intervention for their children who are suitable for receiving implantation. From 1995 to 2011, only about 10,000 people in China received CI, 85% of which were children under 7 years old. Due to technological limitations of CI at the time, as well as a shortage of rehabilitation facilities, even some early-implanters did not develop good speech skills (Chen and Liu, 2010), not to mention those delayed-implanters who experienced more years of hearing deprivation. Besides this small recipient population, many prelingually deaf children had to resort to HA—inefficient devices for profound deafness—to try to rebuild their auditory pathways, and others had to give up the possibility of accessing oral language in their lifetimes. In terms of institutional help, about 50% of deaf children had the opportunity of receiving preschool institution-based rehabilitation. Among them, only 50% were then able to enter the ordinary education system. As a result, until 2013 and according to a survey by the Chinese disabled person's federation, most (75%) prelingually deaf children had to enroll in special schools for the deaf (Xie and Xia, 2018).

With the rapid development of China's economy, the deaf population has received increasing benefits including improved accessibility and affordability of devices and services (including hearing professionals) and increased government support (e.g. the universal newborn hearing screening project, the CI reimbursed program, the development of hearing infrastructure) in the past ten years. For instance, in 2015, the Chinese government alone funded implants for 18,600 children (Chen et al., 2016b). Meanwhile, spurred by potential commercial profits in the Chinese market, mainstream hearing device manufacturers are trying to develop new technology with better ability to process tonal information (Rader et al., 2016; Schatzer et al., 2014). Nowadays, cochlear implantation has become a more common intervention to help Chinese prelingually deaf children to become part of the hearing world. Interestingly, even some prelingually deaf adolescents or young adults, mainly those whose families could not afford implants a few years ago, choose implantation long after the best window for developing oral language is assumed to have passed.

1.1.4 The research gaps

Different from postlingually deaf people who can resort to the tonal knowledge that they already have to rebuild tone categories with the sparse signal transferred by hearing devices (Huang et al., 1995; Wei et al., 2004; Wang et al., 2011, 2012), lexical tone processing is much harder for prelingually deaf populations because they lack the necessary knowledge of lexical tones when they start to use hearing device. Thus, it is very interesting to understand how prelingually deaf people acquire tones. Nevertheless, while considerable studies had been done on prelingually deaf children in the past, researchers had not paid adequate attention to PDA's speech production and perception, not to mention considering the differences between different PDA populations such as HA and CI users (see Chapter 2 for details). Moreover, since past studies on PDA's speech processing have mainly been on non-tone languages, the academic circle knows relatively little specifically about tone production and perception in PDAs. Thus, more studies are needed to understand HA and CI users' characteristics in perceiving and producing synthesized and lexical tones, and the factors that impact tonal processing in the prelingually deaf population.

1.2 The goals and methods of this dissertation

1.2.1 The goals

The purpose of this study is to investigate how Chinese PDAs (born deaf or become deaf before five years old) produce and perceive citation tones in Standard Chinese. By collecting data from prelingually deaf adults who have used hearing aids (HA) and/or cochlear implant (CI) and have received speech rehabilitation for many years, this study is designed to see how these people produce and perceive Mandarin citation tones, to examine the relation between tonal perception and production, and to discuss the possible linguistic and non-linguistic factors that impact deaf people's processing of tonal perception and production in Standard Chinese.

1.2.2 The theoretical framework

At present, the mainstream theories of second language phonology such as Perceptual Assimilation Model (Best and Tyler, 2007), Speech Learning Model (Flege et al., 2003; Flege and Bohn, 2021), and Second language linguistic perception (Van Leussen and Escudero, 2015) all focus on mono-modal L2 users whose first and second languages are spoken languages. However, as we will see shortly, all the HA and CI participants in the current study are sign-spoken bi-modal speakers whose first and second languages are Chinese sign language and Mandarin respectively. The existing L2 phonology theories aren't appropriate in terms of predicting the deaf participants' L1 influence on their L2 because they focus on how the L1 sound system influences the L2 system. As a result, we decided to adopt the speech chain theory—a more general theory— as the theoretical framework to account for the deaf participants' performance in Mandarin tone perception and production.

1.2.3 The research questions

The current study will focus on prelingually deaf adults' performance in perceiving and producing Mandarin citation tones with three experiments: the synthesized tone perception experiment that investigates deaf participants' abilities in categorically perceiving synthesized Mandarin tones under the influence of fine-grained acoustic cues, the lexical tone perception experiment that checks prelingually deaf adults' performance in perceiving real human-produced lexical Mandarin tones embedded in rhymes with different complexities, and the tone production experiment that determines the differences between HA and CI participants in producing Mandarin tones. By analyzing the results of these experiments, this study will address the following three general research questions:

- Firstly, how do prelingually deaf adults perform in perceiving and producing Mandarin tones?
- Secondly, what are the impacts of different factors (tones themselves, rhyme types, and hearing devices) on prelingually deaf adult's Mandarin tone perception and production?
- Thirdly, what underlying mechanisms are responsible for the challenges to prelingually deaf adults' Mandarin tone perception and production?

By answering the above questions, we want to better understand the challenges to deaf people in perceiving and producing tones, explore the implications for improving the practice of speech rehabilitation and the design of hearing devices, and ultimately help prelingually deaf people's spoken language development.

1.2.4 The methodology of the dissertation

This study consisted of three experiments—two perception experiments and one production experiment. The methodological details specific to each experiment will be presented in subsequent chapters. Those aspects of the methods that are common to all experiments are presented here, so as to avoid having to repeat them in each experiment-specific chapter.

1.2.4.1 The participants

To realize the research purpose, we need to collect data from three groups of participants. The first experimental group is the CI group in which all participants currently use CI devices. Because age of implantation is not a factor in our experimental design, the current study does not distinguish early-implanted CI users from delayed-implanted CI users. The second experimental group is the HA group in which all participants are early and constant hearing aid users. The third group—the control group—is the NH group in which all participants are normal hearing adults. Please see Section 3.1.1 for more demographic information of the three groups.

Besides the subjects who participate the experiment tasks, the production study also involves a judge group in which all participants are normal hearing and L1 Mandarin-speaking people who only work as the judges in the subjective assessment experiment. Interested readers can check Section 5.1.3 for a detailed introduction of the judges.

1.2.4.2 The experiments

With the three participants groups, this study involved three experiments.

The first experiment was the synthesized tone perception experiment, in which we used synthesized stimuli to test participants' ability in identifying and discriminating synthesized Mandarin tones relying on fine-grained tone acoustic features only. Based on four real-human produced Mandarin tones embedded in a same syllable, we created synthesized stimuli for all six tone pairs in Standard Chinese. With the synthesized stimuli, we checked participants' performance on different tones and tone pairs. Combing the results of the identification and discrimination tasks, we determined the participants' capability in categorically perceiving Mandarin tones, as well as the impact of the acoustic features (tonal duration and F0) and device type (CI, HA, and NH) on the participants' performance.

The second experiment was the lexical tone perception experiment. In this experiment, we used the lexical stimuli to test the participants' ability in categorically identifying Mandarin tones. This experiment adopted a set of single-syllable words and their distractors as the materials (Mok et al., 2017), and recorded the movement trajectory of the mouse as participants scrolled to the tone they perceived, on a computer screen with Mouse-tracking technology (Freeman, 2018; Kieslich et al., 2020). By statistically analyzing parameters such as the accuracy of the choice, reaction time, and the maximum deviation of the mouse trajectory, we tested the participants' ability to identify Mandarin citation tones produced by a real human being. By doing this, we should determine the difference between the participants' capability in perceiving the synthesized and the real human-produced Mandarin tones, explore the reasons for this difference, and answer the questions of whether the ability of categorically perceiving synthesized tones is the basis of categorically perceiving lexical tones.

The third experiment was the lexical tone production experiment. In this experiment, we asked the participants to produce a set of single-syllable words, and then adopted two types of analysis: the acoustic analysis and the subjective assessment. With the acoustic analysis, we extracted the acoustic parameters from Mandarin tone data produced by the participants. By statistically analyzing these parameters, we checked the difference between the deaf and normal hearing participants and determined the influence of syllable complexity on the production of Mandarin tones. By inviting multiple judges to evaluate tone production data, the subjective assessment tested participants' performance in producing the four Mandarin tones. With the results of the acoustic analysis and the subjective assessment, we determined the ability of the participants in producing Mandarin tones. It is noted that the same word list is involved in the two lexical tone experiments. By doing this, we would like to compare the participants' performance in producing and perceiving Mandarin lexical tones, and to check the impact of the rhyme type and hearing device on deaf people's perception and production of Mandarin tones.

1.2.4.3 The apparatus, Procedures, and safety arrangements

The apparatus for the tone perception and production experiments included a Zoom H4N digital recorder, a HP-800 headphone, and a Lenovo laptop. The software needed for the experiments and their operation documents, and the video clips for using the device were pre-installed on the laptop. To avoid face-to-face contact between experimenters and participants due to COVID-19, we sent the apparatus to the participants one after another by Meituan delivery. Besides the apparatus, we also included a hard copy of the consent form in the delivery package.

The procedures of data collection for each participant were as follows. First, the experimenter mailed the apparatus to the participants. Then, the experimenter contacted the participants by instant message apps (Wechat or QQ) and instructed them how to fulfill the tasks by themselves remotely at their own places, which included helping them to get familiarize themselves with the materials, the equipments, and data collecting procedures. After that, the participants did the experiment under the constant monitoring of the experimenter. If the participants had any problems, they would get instant response from the experimenter using instant message apps. After finishing the experiments, the participants sent back the apparatus and the signed consent form by mail. The cost, covered by the experimenter, was paid upon receipt of the package. Finally, the experimenter sterilized the apparatus, collected the data, cleaned the programs, and cleaned the device again for the next usage.

Device sterilization is the key safety arrangement for the experiment. Before sending the apparatus to the participants, as well as after getting it back from them, the experimenter sterilized the instruments with surgical alcohol and ultraviolet radiation. The experimenter also prepared some alcohol prep pads for the participants to clean the device and their hands before and after using the device.

1.3 The significance of this dissertation

The significance of this study is as follows. On a narrow level, this study can broaden our knowledge about PDA's performance in processing Mandarin tones, about the factors that impact their tone perception and production, and about the end point of prelingually deaf people's tone acquisition. On a broader level, this study can help us understand more about the mechanisms underlying speech perception and production across populations. Practically, this study will provide some references for specialists such as otolaryngologist, speech therapists, and hearing device engineers, for their work supporting PDAs.

1.4 The structure of the dissertation

This dissertation consists of six chapters. Chapter 1 outlines the background, importance, methodology, and structure of the dissertation. Chapter 2 provides a brief introduction to the literature closely related to this thesis, including the overview of the main research methods on the topic of Mandarin tones, tone acquisition of Standard Chinese by normal hearing people, and Mandarin tone production and perception by prelingually deaf people. Chapter 3 covers the experimental design, procedures, and results of the synthesized tone perception experiment, and provides some discussion based on the results. Chapter 4 presents the experimental design, procedures, and results of the lexical tone perception experiment. Chapter 5 reports the experimental design, procedures, and results of the lexical tone production experiment. After summarizing the previous chapters' findings, Chapter 6 provides a general discussion on these findings before concluding the dissertation by presenting the implications, limitations, and future directions of the current study.

Chapter 2

Literature Review

Unlike other major languages such as Spanish, English, and Arabic, Mandarin is a tone language in which tone is crucial in distinguishing lexical meanings. Naturally, Mandarin tone production and perception are of interest to researchers in Speech Sciences and relevant fields such as audiology, speech rehabilitation, and cognitive neuroscience. In this chapter, we first provide a general review of the methods used in studying Mandarin tone production and perception respectively by Chinese Phonetics and related fields. Then, we will introduce the research on Mandarin tone production and perception of four specific populations—normal hearing children (NHC), Mandarin-speaking seniors (MSS), non-native adult speakers (NAS), and prelingually deaf children (PDC)—before introducing our pilot studies on prelingually deaf adults (PDA). It is noted that NHC, NAS, PDC, and MSS are not the target populations but closely relevant populations to our experimental participants, so the literature review of these populations will provide some reference for the current study. Considering the hearing device should also play an important role in PDA's speech processing including tone perception and production, we will briefly introduce HA and CI, the hearing devices used by the two deaf participant groups. Finally, we will conclude this chapter by making research predictions based on the reviewed literature, advancing the specific research questions for each chapter, and providing predicted answers under the theoretical framework of the speech chain theory.

2.1 A general review of research methods

Tone production and perception are located at the two ends of the speech chain and involve different psychophysical processes. Thus, the research methods used in studying tone production and perception are also different: To study tone production, researchers need to record the participant's speech and then analyze the speech signal using acoustic analysis, or evaluate the recordings using a subjective assessment; to study tone perception, researchers often need to invite the participants to listen to synthesized stimuli or real-human produced materials and analyze the participants' responses to these stimuli.

With these in mind, this section will briefly review the methods Chinese Phonetics and related fields have adopted in studying Mandarin tone production and perception respectively. By doing this, we hope to provide a foundation to introduce the research development of Mandarin tone production and perception for particular population groups.

2.1.1 The research methods of Mandarin tone production

As one of the most prominent features of Mandarin Chinese, tone has received long and enduring attention in the history of Modern Chinese linguistics. In the 1920s, the first generation of Chinese phoneticians started to study Mandarin tones using the methods of Experimental Phonetics. For instance, Fu Liu, the founder of the first Phonetics Laboratory in China, had published the first monograph *Record of Experiments on the Four Tones* that studied the tone systems of Beijing dialect and 11 other Mandarin dialects (Liu, 1924). In the following decades, the method of acoustic analysis has been becoming increasingly popular. Considering a tone is defined by a pitch contour (fundamental frequency and its change across the time dimension) and its duration, pitch contour is always the center of attention in the study of Mandarin tone production and perception. In those early studies, researchers often used musical notes to represent the pitch contour (Chao, 1956; Lin, 1965; Chao, 1968). With the application of spectrograph technology, the F0 value in Hertz became more widespread, but transforming the F0 from Hertz into semitone was and is still popular among Chinese Phoneticians when defining Mandarin tone contours (Lin, 1985; Wu and Lin, 1989; Tupper et al., 2020).

It is noted that, although the mean F0 is an important parameter in describing Mandarin tones in the early studies in Experimental phonetics, many phoneticians have noticed its shortcomings in describing the F0 changes across the time dimension. As a result, the past few decades have witnessed efforts of developing some more sophisticated parameters to account for Mandarin tone contours. Some studies tried to use a couple of distinguished points such as the beginning point, the end point, or/and the range (the distance between the highest point and the lowest point) to define the shape of tone contour (Jeng et al., 2006; Zhang et al., 2012a; Yang, 2015; Flemming and Cho, 2017; Hou et al., 2019). Other studies, realizing the possible deficiency of using a few points, developed some more sophisticated parameters like slope and curvature to describe the tone contour (Ghosh and Narayanan, 2009; Tang et al., 2019, 2021). In recent years, the analysis of the parabola—a bundle of parameters including mean F0, slope, and curvature of the tone contour—has become the most popular method in the studies of Mandarin tone production (Chen et al., 2017c; Li and Chen, 2016; Shih and Lu, 2015; Tupper et al., 2020).[1]

Considering the far-reaching influence of the “father of Chinese Modern Linguistics” Yuen-Ren Chao, his “Five level tone mark” was and is still the dominated tone representation system in describing the tone systems of Mandarin Chinese and other Chinese languages (Chao, 1956, 1968; Fon and Chiang, 1999). Under the framework of Chao’s tone representation system, Feng Shi and his colleagues developed Tone Pattern Theory (Shi, 1986; Shi and Liao, 1994)—a phonetics-based phonology of tone that is widely-used in China in recent years. The psychophysical basis of Tone Pattern Theory is that, although people from one language community have different F0 ranges, they can understand the tones produced by other people because all people in this language community have the ability to transform an absolute F0 value into a relative pitch, and assign it a tone value which belongs to the tone system in that language. Thus, after determining a person’s tone register that is defined by their largest F0 value and smallest F0 value, Tone Pattern Theory divides their tone register into five parts and transforms the absolute F0 values into T-values that range from 0 to 5. The correspondence between T-value and Chao’s “Five level tone mark” is displayed in Table 2.1 (Please see Section 5.1.2 for more details of T-value calculation).

Table 2.1: The correspondence between T-value and Chao’s mark

T-value	Chao’s mark
0.00~1.00	1
1.01~2.00	2
2.01~3.00	3
3.01~4.00	4
4.01~5.00	5

After normalizing the tone data by transforming the absolute F0 value into T-value, the tone data of people of different ages and genders could be taken together for statistical analysis. Moreover, after the T-value transformation, the results of acoustic experiments could

---

[1] In the current study, I will also calculate the parabola parameters in the lexical tone production experiment. Interesting readers could check Section 5.1.2 for more details about this approach.

be compared with the results of traditional transcriptions by phoneticians and dialectologists who are accustomed to Chao's tone representation system. Benefited from these merits, Tone Pattern Theory became one of the most successful phonetics-based phonology theories in China, and is widely applied to study the tone systems of Standard Chinese (Shi and Wang, 2006; Kang and Wu, 2020), Chinese Dialects (Gao, 2021; Qin and Yang, 2021; Shi, 2019), and other Chinese tone languages (Chang, 2021; Wei, 2014).

In related areas such as Clinical Phonetics, Audiology, and Speech Rehabilitation, although acoustic analysis is also used in studying Mandarin tone production (Shen et al., 2013c; Zhang et al., 2012a; Zhou et al., 2008), it seems the subjective assessment is a more popular method for researchers (Chen et al., 2017b; Li et al., 2018; Qiu et al., 2021). Different from the acoustic analysis that extracts and analyzes parameters from recordings produced by subjects, subjective assessment relies on judge's experience to evaluate the tone tokens produced by the subjects. Generally speaking, the subjective assessment has two apparent characteristics: first, the stimuli for data recording in these studies usually are Standardized speech intelligibility and audiometry materials; second, the judges in these studies are speech therapists who have abundant experience in subjective evaluation of Mandarin speech production, but the number of judges usually ranged from one to three. Compared with the acoustic analysis, subjective assessment has the merits of being easier to take place and faster to get results, but has the shortcomings that the results are more reliant on the experimenter's experience and harder to repeat and crosscheck.[2]

2.1.2 The research methods of Mandarin tone perception

Although not as long as that of Mandarin tone production, research on Mandarin tone perception was quite early in Chinese Phonetics. In 1976, Wang (1976) applied the first experiment using a "two-alternative forced choice" (2AFC) procedure to explore the categorical perception of Mandarin T1 and T2 by using synthesized stimuli that manipulate the height of the starting point. After that, many Chinese Phoneticians had paid substantial attention to the categorical perception of Mandarin tones with synthesized stimuli created by manipulating the starting point or/and the ending point (Gandour, 1983; Lee et al., 1996; Liu, 2004; Cao, 2010; Wang and Tan, 2015). Among these following studies concerning the categorical perception of Mandarin tones, Peng et al. (2010) is one of the most important papers. In this paper, Peng and his colleagues stipulated the equations for calculating

---

[2] Considering the existing subjective assessment only based on a few specialists' judgment of accuracy and intelligibility, the current study would develop a new method of subjective assessment involving multiple judges. Please see Section 5.1.3 for more details.

identifying scores[3] and discrimination scores that are widely used in subsequent studies (Wang and Tan, 2015; Chen et al., 2017a; Zhang et al., 2020b). According to Peng et al. (2010), the equation for calculating discrimination score is displayed in Equation (2.1):

$$P = P(''S''|S) \cdot P(S) + P(''D''|D) \cdot P(D) \tag{2.1}$$

where $P(''S''|S)$ denotes the percentage of “same” responses to all “same” pairs, $P(''D''|D)$ means the percentage of “different” responses to all “different” pairs, $P(S)$ and $P(D)$ are the percentages of “same” and “different” pairs, respectively. From Equation (2.1), one could see that the discrimination score, just like d-prime, is a sensitivity measure that is developed under the framework of Signal detection theory (Macmillan and Creelman, 2005, p. 5). However, different from d-prime which considers “hit rate” and “false alarm rate” simultaneously, the discrimination score involves “hit rate” and “correct rejection rate” at the same time.

Another important study is Blicher et al. (1990)’s paper in which the researchers investigated the categorical perception of T2 and T3 in Standard Chinese for the first time. Different from the other tones, T3 is a typically curved tone in Mandarin Chinese. Thus, the study of T3 involves a new possible factor that impacts the categorical perception—the turning point. After that, more studies discussed the impact of factors such as the height of the starting point, the place of the turning point, and the height of the ending point to the categorical perception of the two tones (Liu, 2004; Wang and Li, 2010; Rong, 2012; Shen et al., 2013a; Wang et al., 2014; Wang and Tan, 2015). More important, Blicher et al. (1990) is also the first study that investigated the influence of tone duration on the categorical perception of Mandarin tones. In this study, Blicher *et al.* synthesized two sets of stimuli with different durations: 350 ms and 450 ms, and then compared the categorical perception of T2 and T3 across the two duration conditions. The results of the study showed that “syllable lengthening appears to enhance auditorily the F0 cues for the Tone 3 category” (Blicher et al., 1990, p.37). After a period of silence, the last decade witnessed more attention to the effect of tone duration on the categorical perception of Mandarin tones (Wang and Peng, 2012; Rong and Shi, 2013; Huang et al., 2015b; Chen et al., 2017d; Wang et al., 2017b). It is noted that all these studies only differentiated two or three tone durations in creating synthesized stimuli, and thus could not tell the effect of duration to categorical perception of Mandarin tones in general.

---

[3] The original equation for calculating identifying scores was advanced by Hallé et al. (2004), Peng et al. (2010) replaced 0% with 0.1%, and 100% with 99.9% for individual identification curves at both ends, in order to fit the asymptotic property of probit function.

Although the method of tone categorical perception with synthesized stimuli has also been used to probe the categorical perception of Mandarin tones by special populations such as people with dyslexia (Zhang et al., 2012b), autism (Wang et al., 2017a), or deafness (Zhang et al., 2020b), the method of speech audiometry with lexical words is more widely used in areas like Clinical Phonetics, Audiology, and Speech Rehabilitation (Kong and Zeng, 2006; Chang et al., 2016; Liu et al., 2019). As a standardized hearing test that is widely used in above-mentioned fields, speech audiometry determines the client's speech-reception threshold and word-discrimination score by measuring the number of words the client can repeat after they are heard when delivered through earphones at precise decibel intensities (Chernecky and Berger, 2013, pp.84-180). With standardized materials embedded in specialized software or platform, studies in these fields often adopt the "four-alternative forced choice" procedure (4AFC) in which the participants are asked to choose which tone they heard among four possibilities (Wei et al., 2007; Wang et al., 2015; Lan et al., 2019). Using this method, the participant's ability of identifying the four tone categories of the Mandarin tones could be tested in an efficient and accurate way. Around ten years ago, some studies pointed out that more research is needed to build up an authoritative and effective set of materials for speech audiometry (Xi, 2013; Fu and Wu, 2014; Ji et al., 2015). Overall, while considerable research has been done on Mandarin speech audiometry in the past years, the work on Mandarin tone perception is still in the early stage which needs more attention in the future.

In the current study, we also used the 2AFC design in the synthesized tone perception experiment and the 4AFC design in the lexical tone perception experiment. However, as one could see in Section 3.1.2, we did not follow Peng et al. (2010)'s method to include all "same" and "different" trials but only used "different" trials to avoid too many discrimination trials in the synthesized tone perception experiment. As a result, we could not calculate the discrimination scores as Peng et al. (2010) did but had to analyze categorical discrimination based on the "hit" and "miss" results in the current study. Moreover, in the lexical tone perception experiment, we considered tones and rhymes/onsets simultaneously in the 4AFC design but not only tones as the common practice did. Thus, as one could see in Section 4.1, the four options in a trial only contained two contrasting tones in the present study.

2.1.3 The merits and shortages of the research methods

By briefly reviewing the research methods Chinese Phonetics and relevant fields had developed in studying Mandarin tone production, it might be safe to state that acoustic analysis and subjective assessment are the main methods used in Mandarin tone production studies. Chinese phoneticians often adopt the method of acoustic analysis, in which they either try to extract some special parameters and apply statistical analysis with these parameters, or they follow the Tone Pattern Theory that converts the F0 values into T-values from multiple points with equal time interval and directly observe the tone contours. For the specialists in the relevant fields such as clinical audiology and speech rehabilitation, subjective assessment is the main method used in Mandarin tone production studies. Obviously, all these methods have some potential risks: the method of analyzing special parameters might fail to reflect the characteristics of the whole tone contours, the method of T-value observation often fail to involve any statistical analysis (see Section 5.2.1.2 for details), and the method of subjective assessment generally only has one to three judges which might aggravate the biases caused by the judges.

As for the research methods of Mandarin tone perception, our brief review shows that the categorical perception experiment with synthesized stimuli and the lexical tone perception with the "four alternative forced choice" are the main methods adopted in Mandarin tone perception studies. The former method strengthens in probing the participants' abilities to tell the difference of the stimuli with fine-grained differences, and the latter one highlights convenience and efficiency. Nevertheless, the two methods also have their shortcomings: the experiment using the former method is complicated and time-consuming, so only a few studies have checked participants' performance in all Mandarin tone pairs[4], and no study had checked the effects of the durations with fine-grained differences; in the experiments using the latter method, it is hard to remove the risk of ceiling effect because the experiment tasks are often too simple and easy for the participants.

Overall, both the methods for studying Mandarin tone production and perception are having their own merits, and thus are good starting points for the future research on Mandarin tone production and perception. However, these methods also have their own apparent shortcomings. Therefore, we determine to combine these methods together to study PDA's Mandarin tone production and perception. By combining the results of the current experiments together and crosschecking the findings of different methods, we hope to achieve more robust findings in the current study.

---

[4] In Mandarin Chinese, there are four tones: T1, T2, T3, and T4. With the four tones, we can form six tone pairs, namely T1T2, T1T3, T1T4, T2T3, T2T4, and T3T4 in Mandarin.

2.2 The impact factors of Mandarin tone production and perception

In this section, we will briefly introduce the important factors—the tones themselves, the syllabic rhymes, and the hearing devices—that impact PDA's Mandarin tone perception and production. Considering the existing studies on PDA are limited, we would review the studies on Mandarin tone production and perception of four related populations—normal hearing children (NHC), Mandarin-speaking seniors (MSS), non-native adult speakers (NAS), prelingually deaf children (PDC)—before introduce our pilot studies on PDA.[5]

Among the five populations, NHCs and MSSs are typically developing and declining L1 users, NASs are typical L2 learners[6], and the constitution of PDCs and PDAs is more heterogeneous because the two groups are more like the mixture of L1, L2, and bilingual speakers. For deaf people, delayed L1 speakers may be deaf children who start to learn sign language as their L1 a few years after birth (Pénicaud et al., 2013), or hearing-rebuilding children who start to learn a spoken language as their L1 with a long period of delay (Mayberry and Kluender, 2018). Thus, PDCs nowadays in China are more like delayed L1 speakers of Standard Chinese benefiting from the increasing accessibility and earlier implementation of hearing reconstruction and speech rehabilitation. In contrast, at present PDAs are more like L2 speakers of Standard Chinese or bilinguals of Mandarin and Chinese sign language because of the deficiency of hearing and speech rehabilitation services at the time of their childhoods.

By reviewing the findings about the five populations, we should see the tones themselves, the syllabic rhymes, and the hearing devices are playing important roles in these populations' Mandarin tone production and perception. Based on the review, we would make predictions about the present study.

2.2.1 The effects of tones themselves

As we have introduced in Section 1.1.1, the four tones in Standard Chinese have different acoustic characteristics like pitch contours and durations. These differences in acoustic characteristics, on the one hand, are crucial cues for people to differentiate one tone from

---

[5] It is noted that normal hearing adults should be the baseline for comparison for these special groups. However, since most studies reported in Section 2.1 were focusing on normal hearing adults, we would not talk more about this population in this section.

[6] Typical L2 learners are normal-hearing people who start to learn another spoken language after mastering their L1s.

the other. On the other hand, these differences should also make some tones harder to produce and perceive than others. That is, tones themselves should be considered as an important factor that plays an important role in Mandarin tone production and perception. In this section, considering the literature about these populations' tone production and perception is abundant, we will briefly introduce studies on each population respectively, to see whether these populations have any common points in the difficulty levels of Mandarin tone production and perception.

2.2.1.1 Normal hearing children

Mandarin tone perception and production by NHCs—the typical L1 learners of Mandarin—is the heart of several fields related to this dissertation, including clinical audiology, speech rehabilitation, and cognitive neuroscience. More importantly, Mandarin tone studies focusing on NHC could also provide some reference for the studies on other populations including deaf people who have Mandarin as their first or second language. Intertwined with first language development, many studies have been conducted on NHC's Mandarin tone production and/or perception in the past decades. In the following, to make the review more clear, we will briefly introduce the studies focusing on the NHC's perception and production of Mandarin tones respectively.

Mandarin tone perception by NHC is a hot and in-progress topic. For normal hearing children, tone perception is acquired quite early in life. In a study focusing on English-, Cantonese- and Mandarin-exposed infants' early phonological development, Yeung et al. (2013) revealed that Cantonese- and Mandarin-exposed infants demonstrated tone discrimination abilities as early as four months, suggesting that the formation of tone categories evolves earlier than that for vowels and consonants. Using neurophysiological methods such as MMN (mismatch negative) and ERP (event-related potential), considerable studies had investigated NHC's Mandarin tone perception that reported the perceptual reorganization for Mandarin tone begins earlier than six months (Cheng and Lee, 2018; Lee and Cheng, 2020). When Mandarin-learning children are getting a little older, say around nine months, auditory tasks with simple experiment designs had been used more often to test their performance in Mandarin tone perception (Mattock and Burnham, 2006; Tsao, 2008; Mattock et al., 2008). An earlier study reported that Mandarin-exposed infants at ages 1;3 (year; month) to 1; 4 can differentiate T1 and T4 both in production and perception, indicating the two tones are relatively easy for NHC (Jeng, 1979). Contradicting this finding, a later study reported that Mandarin-learning infants from 10 to 12 months old have better performance in differentiating T1 and T3 than in the other two tones (Tsao, 2008). Similarly, Zhu and Dodd (2000) and Zhang (2014) also agreed that NHC could perceive all four tones correctly at two years old. For NHC who are four years of age or older, some more complex auditory experiments such as identification tasks and discrimination tasks were used to test their performance in Mandarin tone perception, and the results often revealed that they performed quite well in these experimental tasks (Chen et al., 2017a; Xi et al., 2009; Yang and Liu, 2012). At present, the majority of existing studies agreed that it happens quite earlier in life: at two-year-old, NHC can differentiate all four tones; at six-year-old, they get adult-like competence in Mandarin lexical perception (Li and Thompson, 1977; Zhu and Dodd, 2000; Xi et al., 2009; Chen et al., 2017a).

Compared with Mandarin tone perception, it seems that NHCs need longer and experience more difficulties in mastering Mandarin tone production. Using the methods of acoustic analysis or/and subjective evaluation, researchers investigated NHCs' performance in producing Mandarin tones. Some studies reported that two-year-old NHC could produce all four tones correctly, which is much earlier than vowels and consonants in Standard Chinese (Zhu and Dodd, 2000; Zhang, 2014). For example, in a study based on the spontaneous speech data of 129 Mandarin-speaking children aged 1;6 to 4;6, Li et al. (2000) found that typically developing Mandarin-speaking children acquire lexical tone production before 1; 6. In contrast, more researchers argued that a typical Mandarin-speaking child can produce lexical tones well around 3-year-old, which are also earlier than segments (Li and Thompson, 1977; Wu and Xu, 1979; Zhu, 2002; Wong et al., 2005). It is noted that, after mastering the four tones in production, NHCs still need a few years before they can produce adult-like tones (Chen, 2020; Wong, 2012, 2013; Xu Rattanasone et al., 2018; Yang et al., 2008). For instance, Chen (2020) reported that 4-year-old children produced falling or rising tones with flatter slope than adults while 7-and 9-year-old children produced tones with oversize contours, and these children produced tone with longer duration when compared with adult speakers. In line with Chen (2020), Yang et al. (2008) found that adult-like tone production is still developing for children from 5 to 12 years old, with tone duration and variability decreasing with age, as well as differences between T2 and T3 increasing with age.

Taking together, the existing literature on Mandarin-exposed NHC have revealed three things: First, the acquisition of lexical tones happens earlier than vowels and consonants; second, while the acquisition of tone perception is integrated with tone production, tone perception occurs earlier than tone production; finally, the acquisition of T1 and T4 happens earlier than T2 and T3 except one study showed the opposite.

2.2.1.2 Mandarin-speaking seniors

Different from the NHCs who are on the way of developing cognitive competence and speech ability, MSSs are degenerating in brain, hearing and cognitive processing which lead to their age-related deficits in speech perception and production. Nevertheless, compared with the abundant studies on NHC's Mandarin tone perception and production, only limited attention has been paid on MSS's Mandarin tone perception, and no attention has been paid on their tone production.

Yang et al. (2015) is the first study on MSS's Mandarin tone perception. In that study, the researchers recruited 16 elder Chinese native listeners with mild or mild-to-moderate hearing loss and 16 younger Chinese native listeners with normal hearing, checked these listeners' tone identification using the task of choosing one response from 20 real human-produced alternatives (5 vowels in row × 4 tones in column). The study found that these older listeners had significantly lower scores both in tone identification and vowel-plus-tone identification tasks than their younger counterparts in a quiet condition. Considering noisy backgrounds play an important role on age-related deficit in speech recognition, Chang Liu et al. (2021) used the same experiment paradigm as Yang et al. (2015) and checked MSS's Mandarin tone identification with and without background noise. The results reported that younger listeners outperformed older listeners in all listening conditions, whereas the younger-older listener difference became greater in noise than in quiet, indicating a more detrimental effect of noise for older listeners than for younger listeners.

Besides above-mentioned experiment paradigm that involves real human-produced stimuli, the experiment design of using synthesized tone stimuli is also used in the remaining behavioral studies such as Wang et al. (2017b), Feng et al. (2020), Yaru Meng et al. (2022), and Yan Feng et al. (2022) and Cognitive Neuroscience studies such as Xiao et al. (2020) and Feng (2022). Similarly, these studies also demonstrated that MSS's tone categorical perception are degenerated due to age-related deficits like structural and functional brain senescence, hearing loss, and degeneration in cognitive processing (Ning, 2018; Feng, 2022). It is noted that, however, the synthesized stimuli used in most of these studis are only from one tone pair T1T2, and a few of them used one more tone pair T1T4 or T2T4. Since these studies did not involve T3—probably the hardest tone for listeners (Ning, 2018), MSS's tone categorical perception found in these results might not be without doubt before all tone pairs are tested in further studies.

2.2.1.3 Non-native adult speakers

Since a normal hearing NAS is the typical subject of L2 language acquisition, studies on this population are closely relevant to that of prelingually deaf adults who are L1 signers of Chinese sign language and L2 Mandarin speakers.[7] Different from the case of NHC, for whom lexical tones are much easier to master than segments, lexical tones are seeming much harder than consonants and vowels for NAS. In the past, a wide range of research has been focused on the mechanism of NAS's L2 tone production and perception (Chen, 2006; Ke, 2012; Lin, 1996). In a bibliometric study on the evolution of Chinese language pedagogy, for example, the researchers found that, of the 745 articles published between 1966 and 2013 in the *Journal of Chinese Language Teachers Association*, about 62% are concerned with L2 Chinese tones (Casas-Tost and Rovira-Esteva, 2015). In fact, among the considerable number of studies on NAS of Standard Chinese, many studies concerned the difficulty degrees of tone production and perception (Elliot, 1991; Hao, 2012; Kiriloff, 1969; Miracle, 1989; Xue, 2013; Yang, 2015). At present, although no consensus has been reached on which tone is the hardest one, the majority of researchers agreed that T2 and T3 are much harder than the other two tones to produce and perceive for NASs from various home countries (NB: "from various home countries" means "with different mother tongues".) (Ding and Ji, 2020). As a result, some studies even focused on the production and perception of T2 and T3 by NASs from different L1 backgrounds (Wang and Li, 2010; Xue, 2013; Ding, 2020; Yin, 2021). These results found that, because of the acoustic features and/or pedagogical misunderstanding of T3, NASs from different backgrounds were similar having more difficulties in producing and perceiving T3 even in citation form, not to mention in tone sandhi.

Recent years have witnessed the development of research in Mandarin tone perception by NAS (Hallé et al., 2004; Leather, 2011; Stagray and Downs, 1993; Wang et al., 2003; Wu, 2006). Results from various studies display a complex scenario: while Stagray and Downs (1993) reported that NASs might be sensitive to some small acoustic differences which native speakers ignore, Leather (2011) argued that NASs are less sensitive to small acoustic pitch differences than native speakers of Mandarin; moreover, different from the native speakers who perceive Mandarin tones in a quasi-categorical way, some studies argue that those NASs whose L1s are not tonal languages do not process tones linguistically and that their perceptual judgments are based on a general psychophysical function (Hallé

---

[7] It is noted that not all prelingually deaf adults are L1 signers and L2 speakers. In fact, some prelingually deaf adults in China are monolinguals of Mandarin and some others are monolinguals of Chinese sign language.

et al., 2004; Burnham and Jones, 2002).[8]

Compared with Mandarin tone perception, Mandarin tone production by NASs has received earlier and wider attention (Ding and Ji, 2020). Impacted by tone feature geometry theory (Yip, 1980), some L2 researchers have developed a component approach of analyzing tone production errors (Chen, 1997; Leather, 1990; Shen, 1989; Wang et al., 2003; Bent, 2005). That is, as summarized by Bent (2005), researchers divided the lexical tones into separate components and classified tone errors into four categories: pitch range errors, pitch register errors, and tone contour errors. With this approach, these studies have indicated three things: first, NAS's pitch range is often smaller than that of native Mandarin speakers (Leather, 1990); second, NAS's tone targets fall short of both high and low targets (Miracle, 1989; Chen, 1997); and third, NAS produce wrong contours that involve incorrect pitch contour directions, using static tones to replace dynamic tones or vice versa (Wang et al., 2003).

The relationship of Mandarin tone production and perception in NASs has continued to be a topic of interest in recent years (Ding et al., 2011; Yang, 2012; Kang and Wu, 2020). Up to now, these studies reported different scenarios. That is, while all studies demonstrated the performance in Mandarin tone production is closely related with perception for the NASs with different first languages, some studies reached opposite conclusions. For instance, Yang (2012) argued that the NASs' performance in tone production was better than in tone perception by involving a tone identification task and a reading-aloud task; on the contrary, in line with the prediction of Speech Learning Model (Flege et al., 2003; Flege and Bohn, 2021), Kang and Wu (2020) reported that tone perception always precedes tone production because if the learners can correctly perceive a certain tone, they can produce the tone correctly but not vice versa. It is noted however that as we have said Section 1.2.2, these L2 phonology theories aren't appropriate in terms of predicting the deaf participants' Mandarin tone production and perception because PDA's L1-L2 interface is totally different from that of NAS.

Overall, the studies on NAS's Mandarin tone production and perception reported the following findings: Firstly, it seems that Mandarin tones are harder than vowels and consonants both in perception and production for NAS. Secondly, T2 and T3 are much harder to perceive and produce than T1 and T4 for NAS. Thirdly, it is still not clear whether NAS

---

[8] It is noted that it would be too simplistic to lump all NAS's Mandarin tone perception together because L2 speakers using tone and non-tone L1s, and L2 speakers using different tone L1s perform differently in perceiving Mandarin tones. However, considering this study would not involve any typical L2 group, we would not introduce their performance in the current chapter. Please see Toh et al. (2022) for a detailed review.

performs better in Mandarin tone perception than production or the reverse. From these results, one might note the difficult orders of tones and segmental components are different between NAS and NHC. However, the two groups share the same pattern that T2 and T3 are harder than T1 and T4 both in production and perception. Considering prelingually deaf adults involved in this study are also L2 learners of Mandarin, they might have similar problems in tone production and perception as NASs.

2.2.1.4 Prelingually deaf children

The existing studies on PDCs could provide direct insight on that of PDAs because PDAs were PDCs before entering adolescence. In fact, PDC's tone production and/or perception is a "hot topic" in the relevant fields and thus tremendous studies had been published in the past.[9] In general, the existing studies depict a complicated picture about PDC's tone production and perception (Huang et al., 1995; Chen and Wong, 2017; Li et al., 2014a; McGarr and Osberger, 1978; Morton et al., 2008; Lee and van Hasselt, 2004; Mao et al., 2020). On the one hand, due to the different research designs and participants' background, the existing literature has reported varying or even contradictory conclusions. For example, there is a dispute about which tone is the hardest one for prelingually deaf children. At present, different from the finding that T3 is the hardest tone for NHC and NAS, most studies agree that T2 is hardest both in production and perception for PDC (Han et al., 2007; Zan and Tang, 1998; Zhu et al., 2014a). Interestingly, there is one paper reporting T4 to be the hardest one to perceive for PDC (Cui, 2011), contrary to the viewpoint of most studies that T4 ranks the easiest tone for PDC (Han et al., 2009; Lee et al., 2010; Li et al., 2014a; Tang et al., 2019).

On the other hand, results of these studies have reached some consensuses. First, while tone perception and production are developed quite early for normal hearing children, tone acquisition seems particularly difficult for PDC. That is, unlike normal hearing children who usually acquire tones ahead of consonants and vowels, many PDCs still cannot master Mandarin tones years after acquiring consonants and vowels (Zhang, 2005). Second, compared to NHCs and NASs who primarily use pitch contour to discriminate tones, PDCs rely more on temporal information to differentiate tones both in perception and production (Peng et al., 2004; Chen and Liu, 2010; Deroche et al., 2019). Finally, tonal production and/or perception among PDCs has been linked to degree of hearing loss (Yin et al., 2016;

---

[9] Considering the main topic of this study, we decide not to provide a detailed review here. The interested readers could read the recent reviewing papers like Tan et al. (2016), Chen and Wong (2017), and Gao et al. (2021).

Huang et al., 2021), type of hearing device (Bi et al., 2008; Feng and Wu, 2017), length of device use (Han et al., 2007; Zhou et al., 2013), quality of social support (Liang and Mason, 2013), and many other factors. That is, the earlier and longer the device using and the higher the quality of language rehabilitation and social support, the better NHC's abilities in tone production and perception. Overall, the existing studies have revealed that acquiring Mandarin tones—a task that is easy for NHC but hard for NAS—is tremendously difficult for PDC.

2.2.1.5 Prelingually deaf adults

Compared with the abundant literature on PDC's Mandarin tone perception and production, research on the PDA population is quite scarce (Duchesne et al., 2017). In our opinion, the reasons for this phenomenon lie in two assumptions and one reality. For PDA who received hearing assistance (HA or CI) at younger ages and thus successfully rebuilt the auditory pathway and neural circuitry (Kral and Sharma, 2012), it is natural to assume that they have similar (or slightly lower) speech skills than those of normal hearing populations. For those delayed-implanters who switched to CI at older ages (e.g. older childhood, adolescence, or adulthood) due to the ineffectiveness of HA and to relatively long-term auditory deprivation, people often assume that these PDAs could only obtain limited improvements in speech skills (Schramm et al., 2002; Wooi Teoh et al., 2004; Caposecco et al., 2012; Craddock et al., 2016). Besides the above two assumptions, there is also a reality that it is much harder to recruit PDA participants than PDC. As a result, it would be no surprise that PDA's speech has received little attention in the fields such as clinical linguistics, speech rehabilitation, and otolaryngology.

That is not saying, however, the research value is low to study PDA's speech production and perception. Taking adult CI users for example, the existing literature on non-tone, spoken-language-using deaf people has revealed that, although the speech production and perception of the delayed implantation PDAs remains poorer than that of PDCs and PDAs who received hearing assistance at younger ages, the delayed implantation PDAs do shown significant improvements years after CI using when compared to themselves (Caposecco et al., 2012; Craddock et al., 2016; Peasgood et al., 2003; Wooi Teoh et al., 2004; Straatman et al., 2014). Besides speech perception, previous studies have also revealed that prelingually deaf adolescents and adults could benefit from delayed implantation with respect to better abilities in social interactions (Cao et al., 2022; Dong et al., 2010; Liu et al., 2010; Xie et al., 2019; Zhang et al., 2017), improved communication skills (Caposecco et al., 2012; Ito et al., 2002; Lazard et al., 2010; Ren et al., 2022; Zhang et al., 2020c), and increased psychological wellness (Craddock et al., 2016; Sahli and Belgin, 2006; Sahli et al., 2009; de Sousa et al., 2018; Häußler et al., 2020) etc.

To improve the predictability of implantation outcomes in tone-language-speaking PDAs (especially delayed-implanters) as well as across languages, it would be very important to clarify their speech production and perception characteristics, as well as the crucial factors that impact their speech production and perception. More important, considering Chinese is a typical tone language, it is very important to understand PDA's characteristics in producing and perceiving Mandarin tones. Unfortunately, except for my own pilot studies, PDA's Mandarin tone production and perception have never been addressed in Speech Sciences and relevant fields.

In fact, our pilot studies focusing on Mandarin tones produced by PDAs have revealed that PDAs still face tremendous difficulties in speech production many years after receiving a hearing device (i.e., CI or HA), especially in tone production (Chen et al., 2016a; Zhou et al., 2016; Chen et al., 2017e,f; Hou et al., 2019). Looking at apical syllables, Zhou et al. (2016) studied two PDAs' tongue movements during the production of Mandarin apical syllables ([tsɿ], [tsʰɿ], [sɿ], [tṣʅ], [tṣʰʅ], [ṣʅ], and [ẓʅ]) using ultrasonic technology, and found that they have similar problems in producing these syllables: they produce alveolar syllables as postalveolar syllables, they realize affricates as fricatives, and they are unable to pronounce some types of apical syllables which they can perceive correctly.

Chen et al. (2016a) investigated the tonal patterns of four PDA females (two using HA and two using CI) (see Figure 2.1 for details). Using the methods of acoustical analysis and hearing judgment, they found three things: first, tone durations are more distinctive than tone contours for the participants; second, T2 is the hardest tone to realize; finally, voice quality change (the change of phonation type from modal voice to creaky voice and from creaky voice to modal voice) is used as a strategy to distinguish T3 from other tones.

In two subsequent studies, we probed the acoustic characteristics of T1 (Chen et al., 2017e) and T3 (Chen et al., 2017f) respectively based on the data of 10 PDAs (five males and five females). The results of Chen et al. (2017e) indicated that PDAs usually make three types of errors in realizing T1, namely producing this high-level tone as a rising tone, a decreasing tone, or a flat tone with abnormal pitch height. In Chen et al. (2017f), we found PDA females and males apply different strategies to realize the dipping-rising T3.

---

[10] Upper-left: CI1, Upper-right: CI2, Bottom-left: HA1, Bottom-right: HA2. CI1, CI2, HA1 and HA2 are the first CI participant, the second CI participant, the first HA participant and the second HA participant respectively.

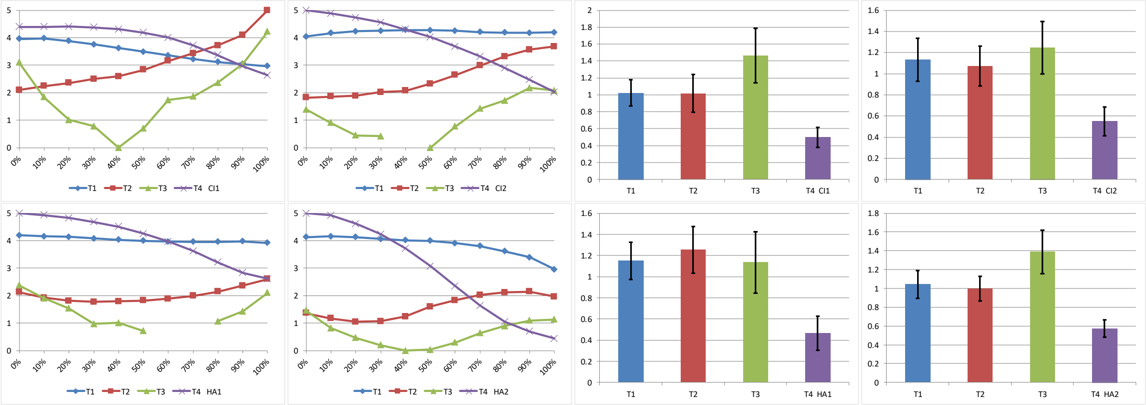


Figure 2.1: Tone contours (left) and duration (right) of four female PDAs [10]

That is, while PDA females tend to adopt creaky voice in producing this tone like normal hearing people, PDA males adopt a longer duration and a slower turning to distinguish T3 from other tones. Recently, by combing subjective transcriber judgments with acoustic analysis, Hou et al. (2019) found the error rate of early implanted PDA's tone production is 12.95%, with the error rates for T2 (30.70%) and T3 (19.85%) much higher than those of the other tones. This study also indicated that PDA males face greater challenges than PDA females in realizing Mandarin citation tones. Limited by the number of participants, the above pilot studies did not allow us to draw any conclusions about the differences between HAs and CIs. However, their results might shed some light on the current study and partly form the basis of our research predictions.

2.2.2 The effects of syllabic rhymes

Standard Chinese only has three types of syllabic rhymes, the simple rhyme that consists of a nucleus vowel, the open rhyme that consists of a nucleus vowel and another vowel (or glide), and the nasal rhyme that consists of a nucleus vowel and a nasal coda. Among the three rhyme types, the nasal rhyme is more complex than the open rhyme, and the open rhyme is more complex than the simple rhyme. In Standard Chinese, all tones are embedded in syllabic rhymes. Since tones always co-occur with syllabic rhymes, one might expect that the perception and production of Mandarin tones should be influenced by the complexity of the rhymes. Nevertheless, although some studies had focused on syllabic rhymes themselves, or the impact of nucleus vowels on tone perception and production, no existing study had paid any attention to the effects of rhyme complexity on Mandarin tone perception and production.

In the past, studies on NHCs had repeatedly proven that nasal rhymes are harder to acquire than open rhymes, and open rhymes are harder to acquire than simple rhymes for typically developing children (Li, 2018; Liu, 2007). As for PDCs, previous studies had also reported their difficulties in producing and perceiving Mandarin rhymes. For rhyme production, some studies on articulation intelligibility had reported that although deaf children apparently lagged behind normal hearing children, they followed the same trajectory in that the nasal rhymes were the hardest, followed by the open rhymes, and the simple rhymes were the easiest for deaf children (Xia et al., 2012; Huang et al., 2021). For instance, a recent study investigated the articulation intelligibility of four participant groups aged four to six years old with different deafness conditions: normal hearing, mild deaf, moderate deaf, and severe deaf (Huang et al., 2021). In this study, the authors found that, although the four participant groups performed differently on the articulation intelligibility of Mandarin rhymes, they followed the same order that the simple rhymes had higher intelligibility rates than the open rhymes, and the open rhymes had higher intelligibility rates than the nasal rhymes. Especially, the intelligibility rates of the children with severe deafness were 62.17%, 31.22%, and 23.39% on simple rhymes, open rhymes, nasal rhymes respectively, indicating that the open and nasal rhymes were far more difficult to articulate than the simple rhymes for this participant group. For rhyme perception, although some studies had investigated deaf children's perception of Mandarin rhymes in general (Mao et al., 2017; Yang et al., 2013; Zheng et al., 2016), none of them had assessed deaf children's different performance in the perception of the three rhyme types in detail, not to mention the relations between rhyme and tone perception.

As for the other populations, a few studies on NAS (Wen, 2019), MSS (Yang et al., 2015; Chang Liu et al., 2021), and even normal hearing adults (Zheng, 2014) also reported that Mandarin tone perception was impacted by nucleus vowels. It is noted that, in studies like Yang et al. (2015) and Chang Liu et al. (2021), the nucleus vowels are the sole components of syllabic rhymes. Thus, these studies had already investigated Mandarin tone perception under the impact of simple rhyme. However, since these studies did not include open rhymes and nasal rhymes, they could not provide any information about the influence of rhyme complexity to tone perception and production.

2.2.3 The effects of hearing devices

Nowadays, HA and CI are the most commonly used hearing devices to help deaf people access speech sounds.

As a device that is best for mild to moderate hearing loss (Conductive hearing losse or-/and Sensorineural hearing loss), the HA device is a sound amplifying system that mainly consists of a microphone to convert sound to electrical signals, a sound amplifier to increase the strength of the signal, and a receiver that is a miniature loudspeaker to convert the electrical signal back to sound (Dillon, 2008). Nothing is implanted into the HA user's inner ear, and thus the HA user's cochlea remains intact which keeps the capability of receiving low-frequency tone information. Deaf people who suffer conductive hearing loss could process fine-grained tone information after fitting HA because their inner hair cells are usually well functioning. For those who suffer sensorineural hearing loss, the audiograms of more than 90% of adults and 75% of children are downsloping, indicating the reason they could not hear some sounds clearly is that they could not process high-frequency information (Macrae and Dillon, 1996). Namely, their cochleae still retain the ability of processing low-frequency tone information. Therefore, it would not be surprising that HA users can process tone information well.

In contrast, CI is a hearing device for severe to profound sensorineural hearing loss that typically consists of an external device that is worn by the recipients and an internal device that is surgically implanted behind the ear (Kral et al., 2016). Since there is an implantation in the inner ear to bypass damaged hair cells and stimulate the auditory nerve directly, the function of the human cochlea has been replaced by the implanted electrode array. Because the size of the human cochlea is very small, an electrode array should be very tiny to be implanted in the cochlea. Limited by the current level of science and technology, the contemporary electrode array has limited implanted depth and could only carry 12 to 24 stimulation channels (Dhanasingh and Jolly, 2017). Because the low-frequency part of sound is processed at the apex of human cochlea, the implanted electrodes with limited depth could not encode very low-frequency cues such as tone information. Moreover, to avoid interactions of adjacent channels, electrical stimulation should be activated by short and interleaved pulses, which makes the representation of fine temporal information especially poor in current CI technology (Schatzer et al., 2015). Since pitch and temporal information are the two most important cues for Mandarin tones, it is reasonable to assert that the CI technology is not able to process tone information directly, especially fine-grained tone information (Feng and Wu, 2020). In addition, the implantation surgery might cause some intracochlear damage like apical trauma that leads to the loss of residual low-frequency hearing (Adunka et al., 2006; Von Ilberg et al., 2011), which might further reduce CI people's ability in processing low-frequency tone information. Considering the shortcomings of the CI device in processing low-frequency and temporal information, tone perception via CI has been found to be substantially less successful than that of segmental components (Wang et al., 2011; Zeng et al., 2014)

Overall, HA and CI are two totally different types of hearing devices, in which HA is basically a sound-amplifying system and CI is an electric-acoustic stimulating system. Due to the different mechanisms of the two technologies in speech signal processing, the HA device has some advantages over the CI device in directly perceiving tonal information. To know more about the two devices and their effects on Mandarin tone perception, please read Ma et al. (2014), Tang et al. (2019), Wei et al. (2004), and Peng et al. (2017)'s review articles.

2.2.4 Summary

From the literature review, three points could be summarized as follows:

- Firstly, the tones themselves should be considered as an important factor that impacts Mandarin tone production and perception. From the existing studies on various populations, one can say that the four tones have different difficulty levels both in perception and production (T2 and T3, especially T3 are harder than the other two tones), indicating that some tones are naturally more difficult than others due to their acoustic characteristics.

- Secondly, rhyme complexity is an important factor that affects Mandarin tone production and perception. At present, although no study has directly investigated the influence of rhyme complexity on tone production and perception, relevant studies indicate that rhyme complexity should influence the processing of Mandarin tones. That is, less complex rhymes should make tone processing easier and more complex rhymes should make tone processing harder.

- Thirdly, for people who suffer from hearing loss, the hearing device is also a crucial factor that affects Mandarin tone perception and production. That is, technically speaking, the HA device should be better than the CI device in processing Mandarin tones.

2.3 The current study

From Section 2.2.1, one might see that PDA's Mandarin tone production and perception has received much less attention than the other populations. That is, only a few pilot studies on PDA's Mandarin tone production have been conducted by us; and to our knowledge, no study has been done on PDA's Mandarin tone perception in the past. In contrast, using the methods introduced in Section 2.1, a lot of studies have been conducted on NHC, NNS, and PDC's Mandarin tone production and perception, and increasing attention has been paid on MSS's Mandarin tone perception.

Considering PDA is closely related to these four populations, the current study predict that PDAs should face those difficulties that we surmarrized in Section 2.2.4. In detail, the current study has the following predictions:

- First, T2 and T3 are more difficult than T1 and T4 both in PDA's tone production and perception. Up to now, although no study has been done on PDA's Mandarin tone perception, our pilot studies have reported that T2 and T3 are also harder to produce for them. Considering tone production is closely related to tone perception, we can predict that T2 and T3 are also harder for PDA to perceive than other two tones.
- Second, since Mandarin tones are embedded in syllabic rhymes, the complexity of rhymes might play an important role in the production and perception of Mandarin tones. In the current study, we predict that the more complex the rhymes are, the worse the deaf participants perform in perceiving and producing Mandarin tones.
- Third, since CI is worse than HA in processing fine-grained tone information[11], the present study predicts that the CI participants will perform worse than the HA participants in producing and perceiving Mandarin tones, especially in perceiving synthesized tone stimuli created by manipulating fine-grained acoustic information.

With the above predictions, we conduct three experiments, namely the synthesized tone perception experiment, the lexical tone perception experiment, and the tone production experiment with the same groups of participants. More important, as we have said in Section 1.2.2, we would explain the participant groups' performance in Mandarin tone perception and production under the framework of the speech chain theory (Denes et al., 1993). According to the speech chain theory, the course of speech production and perception involves three levels (acoustic level, physiological level, and linguistic level) and follows the order of "linguistic level→physiological level→acoustic level→physiological level→linguistic level" (see Section 6.2.3 for more details). Thus, we would account for the participants' difficulties in Mandarin tone perception and production on the deficits in the three levels.

---

[11] The fine-grained information refers to the manipulated frequencies or durations with subtle differences. Please see Section 3.1.2 for more details.

The research purposes, questions, and predicted answers for each experiment are laid out in the following paragraphs.

Considering acoustic cues could impact tone perception, the synthesized tone perception experiment is designed to investigate PDAs' performance in perceiving synthesized Mandarin tones by manipulating fine-grained acoustic cues. Meanwhile, since most existing tone categorical perception studies were only based on easier tone pairs T1T2 or T1T4, this experiment would involve all tone pairs to solve the dispute on whether Mandarin tones could be categorically perceived by the participants including normal-hearing participants. In detail, the specific questions and predicted answers for this experiment are as follows.

- **Question 1:** Can the participants categorically perceive the synthesized Mandarin tones?

  The predicted answer: If the deductions of the existing research results based on a few easier tone pairs are right, both the normal-hearing and deaf participants should be able to categorically perceive the synthesized Mandarin tones. However, the deaf participants, especially the CI participants would perform worse than the normal-hearing participants because of their deficits on acoustic, physiological, and linguistic levels.

- **Question 2:** Are certain synthesized tones harder to categorically perceive than the others for the participants?

  The predicted answer: Yes, since T2 and T3 (especially T3) have more complex acoustic characteristics than T1 and T4, the participants should experience more difficulties to categorically perceive synthesized T2 and T3. Among the three participant groups, the CI group should have more difficulties than the HA group because the CI device could not directly process the acoustic signals of tones.

- **Question 3:** In the cases where the participants could categorically perceive a given tone pair, what impacts do frequency and duration have in the perception of the synthesized tones?

  The predicted answer: Since frequency and duration are primary and secondary tone characteristics respectively, frequency should play a decisive role in tone categorical perception; meanwhile, duration should also impact tone categorical perception, especially for the deaf participants.

Considering the deficits on linguistics level also impair Mandarin tone perception, the lexical tone perception experiment checks the prelingually deaf adults' performance in perceiving real human-produced lexical Mandarin tones carried by rhymes with different complexities. In particular, the experiment tries to answer three research questions.

- **Question 1:** Do the participant groups perform better in perceiving the lexical tones than the synthesized tones?

  The predicted answer: Yes, all participant groups should perform much better in perceiving lexical tones because the participants have more cognitive resources to resort to in fulfilling the lexical tone perception than in the synthesized tone perception tasks.

- **Question 2:** What impacts do the factors such as rhyme complexity and hearing device have in perceiving the lexical tones?

  The predicted answer: Considering more complex rhymes contain more complex acoustic cues and weaker mental representations, the deaf participants should have worse performance in perceiving lexical tones embedded in more complex rhymes; the normal-hearing participants should experience limited difficulties in lexical tone perception tasks due to abundant cognitive resources in lexical tones. Because of the differences between CI and HA, the CI group's performance is still worse than the HA group, but the divergence between them in the current experiment should be smaller than in the synthesized tone perception experiment.

- **Question 3:** Are certain lexical tone pairs harder to identify than the others for the participants?

  The predicted answer: For deaf participants, T2T3 and T2T4 are harder to perceive than other tone pairs; for normal-hearing participants, there are no difference among the tone pairs.

The tone production experiment dives into the participants' performance in producing the four citation tones in Standard Chinese from two angles, the acoustic analysis and the subjective assessment. The three particular questions this experiment try to answer are as follows.

- **Question 1:** What acoustic characteristics do the deaf participants have in producing Mandarin tones?

The predicted answer: Compared with the normal-hearing participants, the lexical tones produced by the deaf participants are harder to differentiate from each other.

- **Question 2:** Can the subjective assessment by multiple judges tell the difference of Mandarin tone production among the three participant groups?

  The predicted answer: Since native speakers have stable mental representations for Mandarin tones, the subjective assessment based on multiple native speakers' grammar sense should be able to tell the difference of Mandarin tone production among the participant groups.

- **Question 3:** Can the results of acoustic analysis be used to predict the scores received in the subjective assessment?

  The predicted answer: The acoustic signals and the judges' assessments should be closely connected, the acoustic analysis should be able to predict the subjective assessment if the acoustic analysis is valid.[12]

By comparing the experiment results and the predicted answers for the questions listed above, this study should be able to address the general questions advanced in Section 1.2.3, and shed some light on developing the theoretical framework of speech perception and production, improving the experiment paradigms and analytic methods, and guiding the practice of perlingually deaf people's spoken language development.

---

[12] Under the framework of the speech chain theory, if the participants wanted the judges "correctly" understand whatever they had said, the acoustic signals they produced should be "right" enough to activate the judges' mental representations (the linguistic forms) of the target words.

Chapter 3

The synthesized tone perception experiments

This chapter focus on the synthesized tone perception experiments. To begin with, we introduce the procedures of stimuli making. Then, we report the methods and results. After that, a brief discussion is provided before ending this chapter.

The synthesized tone perception experiments consist of two sets of experiments: Tone identification and Tone discrimination.

In Tone identification experiment, we tested categorical identification of the four Mandarin tones by applying two-alternative forced choice (2AFC) paradigm on all trials of the six tone pairs (see Figure 3.4 for the procedures of the identification task). By analyzing the experiment results, we tried to determine whether the three participant groups could categorically identify the synthesized tones in Standard Chinese, and to check how frequency and duration impacted their categorical identification of Mandarin tone pairs.

In Tone discrimination experiment, a 2AFC paradigm was also applied to test the participants' performance in discriminating two similar but different stimuli (see Figure 3.5 for the procedures of the discrimination task). By analyzing this experiment results, we determined whether the three participant groups could categorically discriminate the synthesized tones in Standard Chinese, and to decide how frequency or duration impacted their categorical discrimination of Mandarin tone pairs.

As further step, combining the results of the two experiments, we addressed whether prelingually deaf adults could categorically perceive synthesized Mandarin tones.[1] Besides

---

[1] There are two reasons why we check PDA's ability to categorically perceive synthesized Mandarin tones. For one thing, considering many studies of tone categorical perception have been focused on PDC but no study on PDA, this study could help us know more about prelingually deaf people's tone categorical percep-

answering this main question of the current study, we provided some suggestions for future research.

3.1 Methods

3.1.1 The participants

To realize the research purpose, we needed to collect data from three groups of participants. The first experimental group was the CI group in which all participants are CI users, the second experimental group was the HA group in which all participants are early and constant hearing aid users, and the control group was the NH group in which all participants are normal hearing adults who participate all the experiment tasks the same as the deaf groups.

The inclusion criteria of the deaf participants were as follows. Firstly, all deaf participants were PDAs who, at least in one ear, had been diagnosed as suffering severe to profound hearing loss congenitally or since before five years old. Secondly, all deaf participants had participated in oral language rehabilitation for at least three years and had reached Level-3 or above in assessment of Mandarin speech and hearing (Sun et al., 2009). Thirdly, all deaf participants use Mandarin as their first oral language and speak Standard Chinese in most social situations. Finally, all deaf participants are right-handed college students who reported normal or corrected-to-normal vision.

The inclusion criteria of the control group were as follows. Firstly, they are normal hearing participants, reporting no history of speech and hearing impediment. Secondly, they have normal or corrected-to-normal vision. Thirdly, they are right-handed college students. Finally, they come from Beijing or north part of China, speaking Standard Chinese as their first language without any experience in sign language.

With these standards, this study recruited 76 participants in total for the three participants groups. As shown by Table 3.1, the first experimental group (the CI group) consisted of 20 (11 M, 9 F) Mandarin-speaking PDAs who are CI users. Specifically, eight of them are early CI users (at or before 7 years), and 12 people started to wear HA at young age (at or before 6 years) but switched to CI at older age (9 years or later).[2] The

---

tion and their development from youth to adulthood. For another, more relevant to the purpose of the thesis, this study could answer the links between PDA's tone categorical perception and lexical tone perception and production.

[2] To get enough statistical power, a study involving human beings as subjects often has 30 people in one participant group. Since we only have 8 early and 12 delayed CI users in this study, it would be reasonable

second experimental group (the HA group), as displayed in Table 3.2, were 26 (12 M, 14 F) Mandarin-speaking PDAs who are early and constant HA users (at or before 5 years). The third and the control group (the NH group), were 30 (10 M, 20 F) Mandarin-speaking normal hearing adults.[3] All participants were university students from some universities located at Tianjin, China.

Table 3.1: The demographic information of the CI group

Subject	Gender	CA (yrs)	SM	AHL (yrs)	ADU (yrs)	ADS (yrs)	PTA (dB HL)
CI_P01	Male	23	B	0	2.0	12	100 | 100
CI_P02	Female	24	R	0.6	4.0	18	118 | 120
CI_P03	Male	23	R	2.0	3.9	13	120 | 120
CI_P04	Male	23	B	0	2.4	20	98 | 110
CI_P05	Female	24	R	0	6.0	21	120 | 120
CI_P06	Male	23	R	0	3.0	16	120 | 120
CI_P07	Female	23	L	5.0	5.9	9	90 | 90
CI_P08	Male	20	B	1.5	3.5	10	100 | N.P.
CI_P09	Male	20	L	0.5	2.0	19	120 | 120
CI_P10	Female	21	R	3.0	5.8	17	110 | 110
CI_P11	Female	21	R	0.3	4.8	12	90 | 120
CI_P12	Female	24	R	1.3	3.1	19	90 | 110
CI_P13	Male	22	L	2.0	3.0	N/A	120 | 120
CI_P14	Female	21	L	3.5	5.9	N/A	120 | 120
CI_P15	Female	22	L	0	6.3	N/A	120 | 120
CI_P16	Male	20	L	2.0	2.0	N/A	90 | 100
CI_P17	Male	20	R	0	2.0	N/A	120 | 120
CI_P18	Female	20	R	0	2.3	N/A	110 | 110
CI_P19	Male	19	B	3	3.4	N/A	120 | 120
CI_P20	Male	21	R	0	6.6	N/A	120 | 120

Note. CA: chronological age. SM: Stimulation Model, the place(s) of hearing device using (B, using hearing devices at both ears; L, using hearing device at left ear only; R, using hearing device at right ear only). AHL: age of hearing loss. ADU: age at device using. ADS: age at device switching. PTA: unaided three-frequency pure-tone average at 125, 250, and 500 Hz for the unaided (left | right) ear. N.P.: not applicable. N/A: not available because the participants had never used HA before receiving cochlear implantation.

The results of one-way ANOVA indicated that the mean chronological ages (CAs) for the CI group, HA group, and NH group were 21.70 (*SD* = 1.59), 22.27 (*SD* = 2.29), and

---

not to separate the CI participants into two groups.

[3] To save space, this study did not provide a table to list the information of the control group.

Table 3.2: The demographic information of the HA group

Subject	Gender	CA (yrs)	SM	AHL (yrs)	ADU (yrs)	PTA (dB HL)
HA_P01	Male	24	B	1.0	3.0	90 | 95
HA_P02	Male	22	B	2.0	3.5	80 | 60
HA_P03	Female	25	B	0	2.6	60 | 75
HA_P04	Female	21	B	0	1.2	80 | 75
HA_P05	Male	23	B	4.0	4.7	80 | 80
HA_P06	Male	24	R	2.0	4.0	90 | 85
HA_P07	Female	23	B	1.0	1.0	100 | 95
HA_P08	Male	23	B	0	1.0	100|90
HA_P09	Female	22	B	2.0	4.6	80 | 70
HA_P10	Male	22	B	2.2	4.4	75 | 80
HA_P11	Female	24	B	0	3.5	120 | 90
HA_P12	Female	20	B	0	2.0	96 | 106
HA_P13	Male	21	B	0	2.5	110 | 110
HA_P14	Female	22	B	1.0	2.0	120 | 120
HA_P15	Male	20	B	3.6	5.3	75 | 80
HA_P16	Male	21	B	0	2.0	90 | 100
HA_P17	Female	20	B	1.0	5.0	100 | 80
HA_P18	Female	20	B	0	1.0	80 | 100
HA_P19	Female	22	B	5.0	5.5	80 | 100
HA_P20	Female	19	B	1.1	3.0	90 | 90
HA_P21	Female	23	B	0.3	6.0	100 | 100
HA_P22	Male	20	B	0	3.5	90 | 103
HA_P23	Female	20	B	3	3.8	115 | 120
HA_P24	Male	24	B	0.6	1.0	95 | 105
HA_P25	Female	30	B	3.0	3.6	95 | 95
HA_P26	Male	24	B	0	2.8	90 | 75

Note. CA: chronological age. SM: Stimulation Model, the place(s) of hearing device using (B, using hearing devices at both ears; L, using hearing device at left ear only; R, using hearing device at right ear only). AHL: age of hearing loss. ADU: age at device using. PTA: unaided three-frequency pure-tone average at 125, 250, and 500 Hz for the unaided (left | right) ear.

20.43 (*SD* = 1.50) years old respectively, and no significant difference was found among the three groups. Meanwhile, the results of simple T-test reported that the mean AHLs for the CI and HA group were 1.23 (*SD* = 2.18) and 1.26 (*SD* = 2.12) years old, and the mean ADUs for the CI and HA group were 3.90 (*SD* = 2.73) and 3.37 (*SD* = 3.07) years old, no significant difference was found on these two parameters between the two groups. Furthermore, this study also compared the PTAs of the two deaf groups. Considering the PTAs of the left and right ear might be different, the current study calculated the average PTA value for each participant to apply the simple T-test. The results reported that the mean PTAs for the CI and HA group were 111.65 (*SD* = 105.92) and 91.54 (*SD* = 179.32) dB HL, the CI group's PTA was significantly higher ($t = 5.565$, $p < 0.001$) than the HA group's PTA, indicating that the CI group's hearing loss was much worse than the HA group. From this result, we predicted the CI participants would experience more difficulties in all three experiments.[4]

3.1.2 Materials

3.1.2.1 The anchoring sounds

Considering the aim of the synthesized tone perception experiments is to investigate prelingually deaf adults' performance on tone perception in Standard Chinese, we chose syllable [tʰa]—the onset [tʰ] is an aspirated voiceless stop, and the formants of the final [a] are much higher than fundamental frequency (F0) values—as the carrying segmental string for synthesized tones, in order to avoid any possible impacts of the segments to the embedding tones.

To set up the anchoring sounds[5] for stimuli synthesizing, we first embedded the four

---

[4] In fact, people choose CI because they have profound deafness that could not be treated by HA. Therefore, almost all studies concerning CI and HA people have similar situations to the current study that CI people have worse hearing loss than HA people. That is, the impact of the degree of hearing loss and the type of hearing device might be confounded in these studies. Considering one main purpose of the current study is to investigate the impact of the type of hearing device on PDA's tone production and perception, one might suspect that the difference in hearing loss between the CI group and the HA group might make it hard to determine the impact of hearing device. However, we don't think the intertwining of the two factors is a big issue in the current study. In this study, all participants participated in all three experiments. Thus, if the two factors played roles in the synthesized tone experiment, they did play roles in the other two experiments. Namely, when we compared the results of the three experiments, we could counteract the effect of the degree of hearing loss when checking the impact of the type of hearing device.

[5] The anchoring sounds in this paper are the originally synthesized syllables with four tones respectively created by stipulating natural sounds produced by a female speaker. After setting up the duration, the F0, and amplitude for these four syllables, we use them as the start points to synthesize the stimuli between each two of them. That is, every two anchoring sounds are the two extremes of the stimuli spectrum of a given tone

syllables tā [t^ha^{55}], tá [t^ha^{35}], tǎ [t^ha^{214}], and tà [t^ha^{51}] in the carrying sentence wǒ zài shuō X zhè gè zì (*I am saying the character X*. X represents one of the embedded syllables)[6] and invited a female native speaker[7] of Standard Chinese to produce them for ten times in a sound-treated booth. Then, we extracted all [t^ha] sounds, measured their acoustic parameters such as F0, duration, and amplitude. Finally, after calculating the average values of these parameters, we synthesized four anchoring sounds, namely tā, tá, tǎ, and tà, for further use. To make the four anchoring sounds differ from each other in F0 values and durations only, we adopted the same formant pattern and average amplitude for the four synthesized [t^ha] syllables.

We determined the F0 values of the anchoring sounds as follows.

First, based on the average F0 values of the embedded sounds, we set the ceiling and floor of the tone register as 308 Hz and 178 Hz respectively.

Next, following Whalen and Levitt (1995), we converted the values from Hertz to Semitone by Equation (3.1):

$$Semitones(x) = 12 * log(x)/log(2) \tag{3.1}$$

in which x is the F0 value in Hertz, and the reference frequency is 1 Hz. After conversion, the tone register were from 99.2 st to 89.7 st.

Then, by assigning the register ceiling as 5, and the register floor as 1, we converted the tone registers from Chao's numbers to F0 in Semitone. Since there were 2.375 st

---

pair.

[6] According to Chao and many Chinese phoneticians, the tones eliciting in isolation, the so-called called ”citation tones”, are not real tone forms in Standard Chinese. In isolation conditions, speakers often exaggerate the differences among the tones (larger tone register, longer tone durations, etc.) In fact, they believed, the real tone forms are embedded in the flow of speech. Ideally, in the flow of speech, the real forms of the four tones are 55,35, 211, and 53 in Chao's system. However, it is not always true because of the variances among the speakers. With this in mind, instead of trying to extract the ideally real forms of the four tones, we used a carrier sentence to determine the ceiling and floor of the tone register of the four tones and then synthesized the four tones (the anchoring sounds). To avoid the impact of the surrounding syllables, we used the carrier sentence that had been used in many previous studies. In this sentence, the syllable before the targets is always “shuō”. According to the prosodic rules in Standard Chinese, the targets could be impacted by its directly preceding syllable “shuō” only. However, because there is a 55 tone in this preceding syllable “shuō”, it would not cause tone sandhi of the targets in Standard Chinese. Overall, it is safe to use a carrier sentence to extract the targets and there is no danger of phonological changes caused by the surrounding syllables in this sentence.

[7] For most adult male speakers and some female speakers who have low registers, a creaky voice is very common in T3 especially when it is produced in isolation. Considering a creaky voice in an anchoring sound could make it hard to synthesize stimuli, we chose a female speaker with a high tone register and asked her to produce the four target syllables embedded in the carrier sentence. In the current study, we did not find a creaky voice in her production.

between every two adjacent numbers, the tone register of the anchors could be represented as: Chao's 1, 89.7 st; Chao's 2, 92.075 st; Chao's 3, 94.45 st; Chao's 4, 96.825 st; Chao's 5, 99.2 st.

Finally, to improve the naturalness of the synthesized stimuli, we stipulated T3 and T4 as 211 and 53 separately. The reason behind this was that, although the citation forms of the four tones are 55, 35, 214 and 51 in Chao's system, their values are 55, 35, 211 and 53 separately in most situations (Chao, 1933, 1968). Therefore, the F0 values of the four anchors were 99.2-99.2 st, 94.45-99.2 st, 92.075-89.7-89.7 st, and 99.2-94.45 st respectively in this study. Namely, the actual forms of the synthesized anchors could be represented as [t^{h}a^{55}],[t^{h}a^{35}], [t^{h}a^{211}], and [t^{h}a^{53}] (see Figure 3.1).[8]

Figure 3.1 also illustrated the differences of the durations of the anchoring sounds. Based on the measurement of the embedded sounds, we stipulated the syllable lengths of the four anchors as 500 ms, 415 ms, 560 ms, and 335 ms respectively using the overlap-add method (Moulines and Charpentier, 1990) by Praat 6.1.53 (Boersma and Weenink, 2021). To ensure the naturalness of the experiment stimuli, we have checked the four anchoring sounds and all synthesized tokens based on these anchoring sounds in the current study.

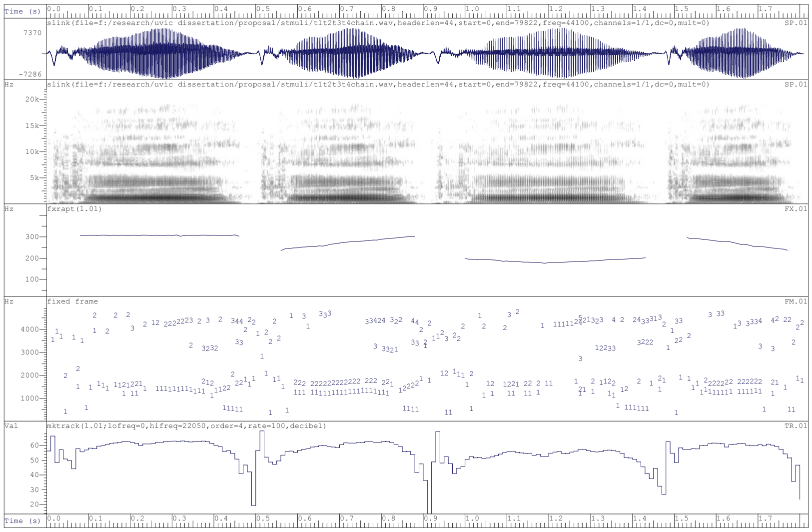


Figure 3.1: The anchoring sounds for stimuli synthesizing

---

[8] As a tone theory with a long history of about 100 years, Chao's system has two main shortcomings: from the aspect of Phonetics, it is oversimplified to describe Mandarin tones by five numbers; from the aspect of Phonology, on the contrary, it is too complex to classify tone heights by five numbers (five degrees). In fact, many experiment-based phonologists believe that three degrees (H, M, L) is enough for Standard Chinese. Considering people still follow Chao's 5 degrees to make stimuli in the synthesized tone perception studies, we still follow this line to make stimuli in the current study.

3.1.2.2 The values of manipulating parameters

Because the purpose of the current experiments was to investigate the impact of F0 and duration to the categorical perception of tone pairs in Standard Chinese, we had to set up the two parameters first, before synthesizing the experiment stimuli based on the four anchoring sounds.

(1) F0

The first parameter that we concerned was the F0. Based on the F0 values of the four anchoring sounds, we compared the F0 values of every two tones and divided the distances into 11 steps for each pair of the six tone pairs (T1T2, T1T3, T1T4, T2T3, T2T4, and T4T3). Considering T3 had a turning point that often locates in the middle of the tone contour, we also interpolated a middle value for other three tones, namely 99.2 st (T1), 96.825 st (T2 and T4), when they were in the same pairs with T3. Table 3.3 displayed the F0 values for the stimuli of the six tone pairs, and Figure 3.2 illustrated the F0 changes among the 11 steps of each tone pair.

Table 3.3: The stimuli F0 values of the six tone pairs (st)

Location	T1T2		T1T3			T1T4		T2T3			T2T4		T4T3		
	Start	End	Start	Middle	End	Start	End	Start	Middle	End	Start	End	Start	Middle	End
S01	99.200	99.200	99.200	99.200	99.200	99.200	99.200	94.450	96.825	99.200	94.450	99.200	99.200	96.825	94.450
S02	98.725	99.200	98.488	98.250	98.250	99.200	98.725	94.213	96.113	98.250	94.930	98.725	98.488	96.113	93.975
S03	98.250	99.200	97.775	97.300	97.300	99.200	98.250	93.975	95.400	97.300	95.400	98.250	97.775	95.400	93.500
S04	97.775	99.200	97.063	96.350	96.350	99.200	97.775	93.738	94.688	96.350	95.880	97.775	97.063	94.688	93.025
S05	97.300	99.200	96.350	95.400	95.400	99.200	97.300	93.500	93.975	95.400	96.350	97.300	96.350	93.975	92.550
S06	96.825	99.200	95.638	94.450	94.450	99.200	96.825	93.263	93.263	94.450	96.830	96.825	95.638	93.263	92.075
S07	96.350	99.200	94.925	93.500	93.500	99.200	96.350	93.025	92.550	93.500	97.300	96.350	94.925	92.550	91.600
S08	95.875	99.200	94.213	92.550	92.550	99.200	95.875	92.788	91.838	92.550	97.780	95.875	94.213	91.838	91.125
S09	95.400	99.200	93.500	91.600	91.600	99.200	95.400	92.550	91.125	91.600	98.250	95.400	93.500	91.125	90.650
S10	94.925	99.200	92.788	90.650	90.650	99.200	94.925	92.313	90.413	90.650	98.730	94.925	92.788	90.413	90.175
S11	94.450	99.200	92.075	89.700	89.700	99.200	94.450	92.075	89.700	89.700	99.200	94.450	92.075	89.700	89.700

(2) Duration

The other parameter that we manipulated is duration. Similar to that of F0, we also compared the durations and divided the differences into 11 steps for each of the six tone pairs. The results, as displayed by Table 3.4, were the duration values for the stimuli of the six tone pairs.

From Table 3.4, one can note that the six tone pairs display two patterns of duration changes. In T1T2, T1T4, and T2T4, the durations of stimuli are decreasing with equal time intervals from D01 to D11; In T1T3, T2T3, and T4T3, the durations of stimuli are steadily increasing with equal time intervals from D01 to D11. That is, there is a same number

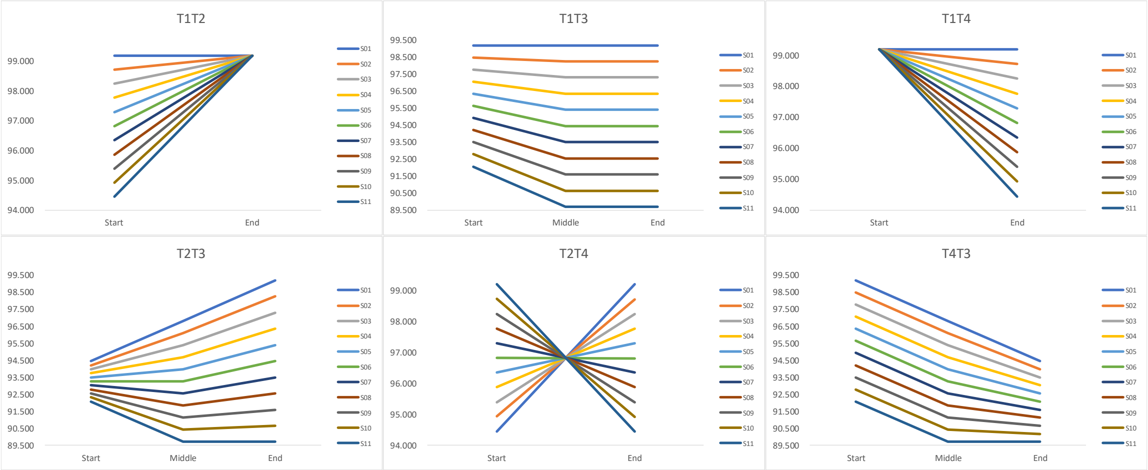


Figure 3.2: The F0 changes among the 11 steps of the six tone pairs

Table 3.4: The stimuli Duration values of the six tone pairs(ms)

Duration	T1T2	T1T3	T1T4	T2T3	T2T4	T4T3
D01	500	500	500	415	415	335
D02	491.5	506	483.5	429.5	407	357.5
D03	483	512	467	444	399	380
D04	474.5	518	450.5	458.5	391	402.5
D05	466	524	434	473	383	425
D06	457.5	530	417.5	487.5	375	447.5
D07	449	536	401	502	367	470
D08	440.5	542	384.5	516.5	359	492.5
D09	432	548	368	531	351	515
D10	423.5	554	351.5	545.5	343	537.5
D11	415	560	335	560	335	560

of stimuli created from the two manipulating orders of duration. To check the impact of duration on tone categorical perception, this study would include the duration order as an affecting factor in the following statistic analysis.

3.1.2.3 The created stimuli

With the above settings of F0 and Duration, we created the stimuli for all the six tone pairs respectively. For each tone pair, we created 11 F0 steps × 11 Duration steps = 121 stimuli. Overall, there are 121 × 6 pairs = 726 stimuli. As an example, Figure 3.3 displays 11 diagonal stimuli of T2T3. In Figure 3.3, the formant pattern and the average amplitude of these stimuli are remaining the same, but the F0 values and durations are continuously changing with an equidistant step. As a result, this [t^ha] syllable has changed from a rising tone (T2) to a dipping-level tone (T3), and its duration lengthened from 415 ms to 560 ms.

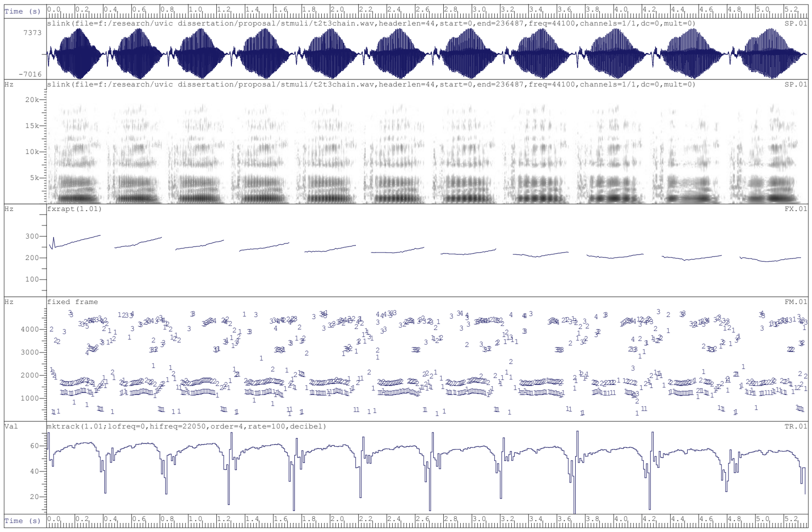


Figure 3.3: The F0 and duration changes among the 11 stimuli of T2T3

3.1.3 Experiment procedures

3.1.3.1 Tone identification experiment

All stimuli created for the six tone pairs involved in the tone identification experiment. Since only one stimulus was used in one trial, there were 726 trials in this experiment. Figure 3.4 sketches the experiment procedures for one trial of T1T2.

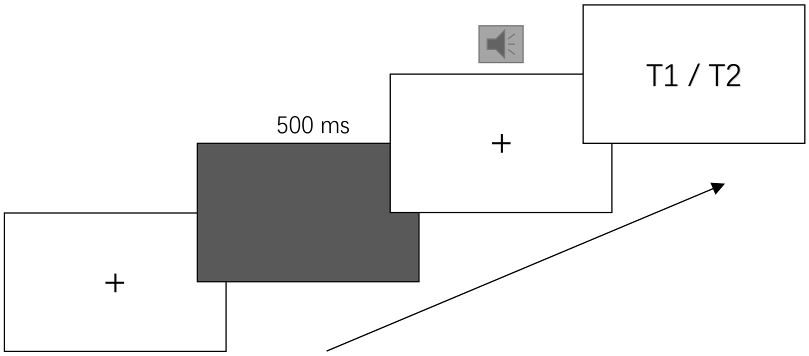


Figure 3.4: The sketch of the identification procedures for one trial

As displayed in Figure 3.4, in each trial, a cross is presented at the middle of the screen before the participants hitting any key to start the task. After the participants start the task, a dark gray screen is displayed for 500 ms and then a stimulus sound is played with the comfortable sound pressure level that the participants had set before the experiment. Immediately after the sound playing, the participants are asked to decide which tone (T1 or T2 in this example) the stimulus belongs to as quickly as possible. After making the choice by hitting *f* (one tone) or *j* (the other tone) on the keyboard, a cross is promoted on the screen again.[9] The participants can take a rest or hit the keyboard to start the task for the next trial. It is noted that all trials of the six tone pairs were mixed together and presented randomly in this experiment. Considering most participants can fulfill all the trials in one hour, as well as they can take a break at any time, only one experimental session was provided in this experiment.

In this experiment, deaf participants were required to use their own CI/HA devices with their daily settings. Due to COVID-19, all participants performed this experiment and all the following experimetns in their own places at their own paces. During the experiment, the experimenters had constantly provided remote support via Wechat or QQ.

3.1.3.2 Tone Discrimination experiment

With the 726 stimuli of the six tone pairs, we created six tonal continuum discrimination tasks. In each task, 99 pairs were created by combining two stimulus sounds separated by two frequency steps from S01 to S11.[10] Overall, there were 594 different pairs but no identical pair in this experiment. Figure 3.5 sketches the experiment procedures for one experiment trial.

As displayed in Figure 3.5, in each trial, a cross is presented at the middle of the screen before the participants hitting any key to start the task. After the participants start the task, a dark gray screen is displayed for 500 ms and then a stimulus sound is played with the comfortable sound pressure level that the participants had set before the experiment.

---

[9] In the current study, we did not counterbalance the answer keys because it would cause additional difficulties in collecting the experiment results. Because of the Pandemic, we could not work with the participants face in face, but had to distribute experiment tasks to the participants one by one through postal delivery. Since the procedure of experiment preparation had already caused too much work, we decided not to counterbalance the keystrokes and the answers across participants in all three experiments.

[10] In the current study, we did not create any discrimination trials by combining two stimulus sounds separated by two duration steps because that would involve too many experiment trials. Considering duration is a secondary feature for tones, we would just study its impact by comparing the choice results of the trials with the stimuli involving the same frequency condition.

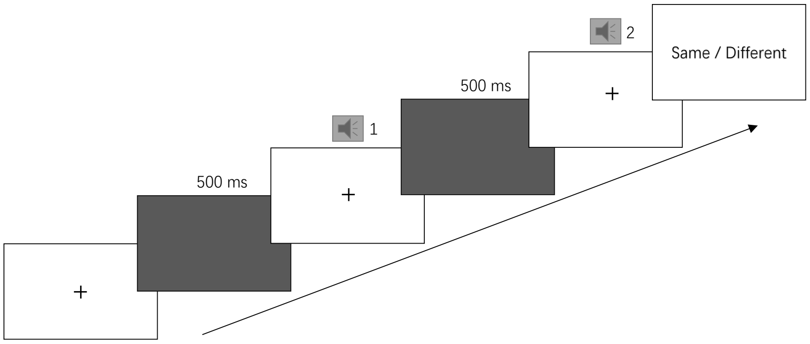


Figure 3.5: The sketch of the discrimination procedures for one trial

Then, a dark gray screen is displayed again for 500 ms before the second stimulus sound's playing.[11] Immediately after the second stimulus sound, the participants are asked to decide whether the two stimuli were the same or different as quickly as possible. After making the choice by hitting *f* (same) or *j* (different) on the keyboard, a cross is promoted on the screen again. The participants can take a rest or hit the keyboard to start the task for the next trial. It is noted that all trials of the six tone pairs were mixed together and presented randomly in this experiment. Considering most participants can fulfill all the trials in one hour, as well as they can take a break at any time, only one experimental session was provided in this experiment. Furthermore, all other settings of this experiment are the same as 3.1.3.1.

As mentioned in 1.2.4, all participants participated this experiment. However, one deaf subject in HA group (HA_P16) did not finish the discrimination experiment and thus his discrimination data was dropped in the following analysis.

3.1.4 Data analysis

3.1.4.1 Data processing

Both the identification and discrimination experiments collected two types of data: choice and reaction time (RT). Specifically, the former counts the choice, and the latter involves the reaction time of each and every trial. Considering our research purpose is to determine

---

[11] For normal hearing adults, the ISI of the past studies was often set as 200 ms (Gerrits and Schouten, 2004; Jiang et al., 2012). In such a short period, people may only have enough time to process the auditory information of the first syllable. For special groups such as children, the ISI was often set as 500 ms (Chen et al., 2017a; Zhang et al., 2020b). The rationale behind this is that the special group may need more time to process the auditory information of the first syllable. Considering the deaf groups may need more time to process the auditory information of the first syllable, we also set the ISI as 500 ms in this study.

the participant's categorical perception of Mandarin tones, this study only focused on the analysis of choice data, namely, the tone choice data for Tone identification experiment and the *same/difference* choice data for Tone discrimination experiment.[12]

In the past, many studies had discussed the characteristics of categorical perception (Xu et al., 2006; Wang et al., 2017b; Yan Feng et al., 2022). For example, Wang et al. (2017b) had summarized the characteristics of Mandarin tone categorical perception as follows:

> *"For native Mandarin listeners, Mandarin lexical tones are perceived categorically; that is, (a) the identification of one tone category (e.g., Tone 1) shifts to another category (e.g., Tone 2) sharply; (b) the discrimination of tone pairs across tone categories shows a peak, whereas the discrimination of tone pairs within one tone category is relatively poor; and (c) discrimination score is predictable from identification data."*

Nevertheless, no previous study had talked about the standards of categorical/non-categorical identification, discrimination, and perception explicitly. Therefore, following the above characteristic of Mandarin tone categorical perception, we have to set up standards to differentiate categorical data from non-categorical data both for the identification data and discrimination data, and then decide on the categorical/non-categorical perception data in the current study.

According to the characteristics of Mandarin tone categorical perception listed above, the standards to decide categorical/non-categorical identification and discrimination were as follows:

(1) The criteria for categorical tone identification

The standards of deciding tone categorical identification are as follows:

- The curve of one tone has an increasing tendency that starts below 50% and ends beyond 50%, the curve of the other tone displays an opposite mirror image;
- The curves of the two tones should roughly correspond to the orders of creating the stimuli;

---

[12] Choice data tell whether someone can identify and/or discriminate a given tone or not, and RT data tell how quickly make the decision. Since the main research purpose is to see whether deaf participants can categorically identify and/or discriminate Mandarin tones, this study focus on the choice data only.

- The two curves must meet each other, but no more than three times at somewhere between S2 to S10.[13]

(2) The criteria for categorical tone discrimination

Considering the two stimuli in one trial are always different, this experiment considered 50% of *same* choice as the bottom of judgement. With that in mind, the following standards were involved to decide whether a tone pair was categorically discriminated or not by the shape of its discrimination curve:

- A discrimination curve should start below 50% and end below 50%;
- There is at least one peak in a curve that is at or above 50%.
- If there are more than two peaks in a given curve, the lowest point of the dip between every two peaks should be at or above 50%.[14]

For each tone pair, we plotted the identification and discrimination results for each and every participant, to decide whether these participants categorically identified and/or discriminated the tones using the above standards. If some participants did not identify and/or discriminate a given tone pair categorically, their data on that tone pair would be excluded from further statistical analysis. As revealed shortly later, the size of categorical discrimination data was too small to be statistically analyzed. Thus, this study only statistically analyzed the categorical identification data.

As a further step, the current study checked categorical perception based on the categorical identification and discrimination results. In line with the thinking of past studies that identification and discrimination should share the same categorical boundary (Huang et al., 2015a; Chen et al., 2017a), we stipulated the standard of categorical perception as

---

[13] Ideally, the two curves meet only once in a categorical identification. However, if we use that standard, it would be too strict, especially for deaf people. We understand that people, especially deaf people, would hesitate at or near a given boundary and consequently jump back and forth between the two choices. Deciding how many times is a hard job. Among the 11 steps in one tone pair, to meet Standard 1 (starting below 50% and ending beyond 5%, or vice versa), theoretically speaking, the two curves could cross 1, 3, 5, 7, or 9 times. After checking the result data thoroughly, we found it would be too strict to use the standard of crossing one time, and too loose to use the standard of crossing five times. As a result, we adopted the standard of crossing no more than three times in this study.

[14] The location of the peak is also important for discrimination curves. Ideally, its location should be the same as the location of the identification boundary. Later on, we would compare the two locations to decide whether the participants could categorically perceive the tone pairs or not.

the difference between the identification boundary and the peak of the discrimination curve should be less than one frequency step. Using the standard, we checked the overlapping between the identification boundary and the peak of the discrimination curve for each participant's categorical identified and discriminated tone pairs and decided the categorical perception of all categorical data.

3.1.4.2 Statistic analysis for the all data

Based on the standards advanced in Chapter 3.1.4.1, we visually checked the choice data to decide categorical/non-categorical identification data, categorical/non-categorical discrimination data, and categorical/non-categorical perception data. With the results of direct observation, we adopted a battery of chi-square tests using *basic* package of R 4.0.5 (Team, 2021). To thoroughly decide the impacts of tone pairs (TonePair), the participant groups (Group), and their interactions on categorical identification, and categorical discrimination, we also applied multiple comparisons with the chi-square test method using *rcompanion* package (Mangiafico, 2023).

3.1.4.3 Statistic analysis for the categorical identification data

As displayed in Section 3.2.1, a minor part of categorical identification and limited categorical discrimination and perception data were left after visually checking the data. Thus, the following statistics only focused on the remaining categorical identification data to check the impacts of frequency and duration on synthesized tone categorical perception.

With the categorical identification data, we visualized the data to directly inspect (Schoonahd et al., 1973) the impact of frequency and duration on tone identification results and decided what parameters would be extracted for statistic analysis, as well as the methods of the following statistic analysis.

(1) Data visualization

To visualize the categorical identification data, we considered frequency, duration, and the identification rates as X dimension, Y dimension, and Z dimension individually, and calculated the 3D matrixes using the inverse square method for each and every tone pair. Figure 3.6 was the visualizing results of T1T2 for the three participant groups. In Figure 3.6, the Z dimensions represent the identification rates of the terminal tone T2, the black dots are the original identification results, and the 3D surfaces are the smoothed results calculated by the inverse square method.

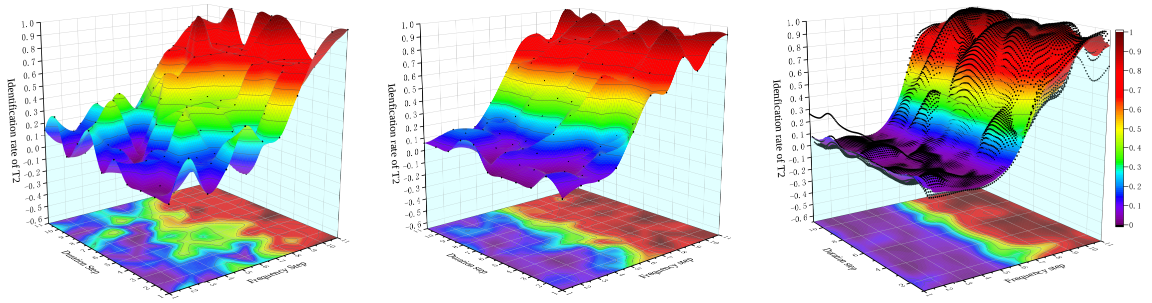


Figure 3.6: The smoothed matrix of results on T1T2 for CI (left), HA (middle) and NH (right)

Observing Figure 3.6, one can note that frequency seems playing a regular and consistent impact on identification results. That is, for all three participant groups, the identification rates of T2 are steadily increasing from the start-point stimulus to the end-point stimulus. Meanwhile, from the shadows on the floor of Figure 3.6, we can see that the middle points at the beginning duration steps are a little to the right compared with those ending duration steps for all three groups, indicating that longer duration (the first few duration steps) delays the categorical boundary of T1T2.[15] However, also from these shadows, we can see that the impact of duration seems inconsistent among the three groups. Namely, duration seems playing little impact on NH and HA, but does play apparent impact on CI group.[16] Overall, data visualization of the six tone pairs revealed a uniformed pattern. That is, for all six tone pairs, frequency has a much stronger impact on identification rate than that of duration, and there are no apparent interactions between the two factors.

Thus, the current study considered frequency as the main dimension of statistical analysis. Since we could not do statistical analysis with the identification curves directly, in the procedures of parameter extraction, we would extract parameters from the identification rate curves of the frequency dimension and analyze the impact of frequency on synthesized tone identification. Meanwhile, considering stimuli duration also impacts synthesized tone perception, we would also check its impact based on the extracted parameters in the fol-

---

[15] From the shadow on the floor, we can see that there are blue to yellow bays at the beginning duration steps that get a little deeper into the red zones, indicating that the boundaries are to the right of the middle point. In contrast, there are no such deep bays at the ending duration steps.

[16] From this figure, we could see the changes in duration are inconsistent for the CI group, and no clear effect patterns for HA and NH people. Since data visualization could not provide more precise details, we will check the impact of duration on tone categorical identification through statistical analysis later.

lowing statistical analysis.

(2) Parameter extraction

As illustrated by Figure 3.7, this study extracted two key parameters, Position and Slope, from the identification rate curves. Specifically, Position was the 50% crossover point of the two identification curves that indicated the location of the categorical boundary. Considering the positions of the categorical boundaries could be impacted by many factors such as the anchoring sounds, the tone pairs, and the manipulating steps, this study considered the NH group's Position values as the standards to evaluate the deaf groups' data, to see whether they had set up similar categorical boundaries to the normal hearing ones.[17]

The other parameter, Slope, was a parameter of assessing the wellness of tone categories. In this study, We defined Slope as the slope of the linear line that was fitted by the last stable point from the start-point direction and the first stable point from the end-point direction on the identification rate curves. Theoretically speaking, if two tones in a given tone pair were better categorically identified, its Slope value would be larger because the slope of the linear fitted line was steeper (Peng et al., 2010; Chen et al., 2017a; Zhang et al., 2020a,b).

When two tones in a given tone pair were well identified by the participants, just like the one illustrated by Figure 3.7, the slope line crossed the Position point of the identification rate curves. However, if the two tones were not categorically identified perfectly, the fitted slope line might not cross the Position point like the examples displayed in Figure 3.8, or we could not find the Position point because the two identification rate curves had more than one crossing points like NH_P28's case in Figure 3.9.[18] In that case, we had to calculate a Position value with the fitted slope line of the two identification rate curves.

---

[17] Ideally, the boundary positions of all tone pairs should be located in the middle of their stimuli continua. However, considering the acoustic features of synthesized stimuli are varied in every experiment, nobody can ensure that normal hearing people's boundary position of a given tone pair is located in the middle of its stimuli continuum. Therefore, for a given tone pair, if its two tones can be categorically perceived by native speakers with normal hearing, all members in this speech community should share a similar boundary position with that tone pair. In this study, if deaf participants have a similar performance to the NH participants, they should have the same or similar boundary positions in these tone pairs. Namely, the closer the deaf groups' boundaries are to the NH group's boundaries, the better the deaf groups performed in identifying the tone pairs.

[18] Although we excluded the cases that were not categorically identified, it is not saying that all the remaining cases are perfectly categorical. As shown in Figure 3.7, some cases like CI_P13's T1T3 are perfectly categorical; some others such as the T1T2 identification curves of HA_P12 (left), HA_P17 (middle), and HA_P25 (right), as shown in Figure 3.8, are categorical but not that perfect.

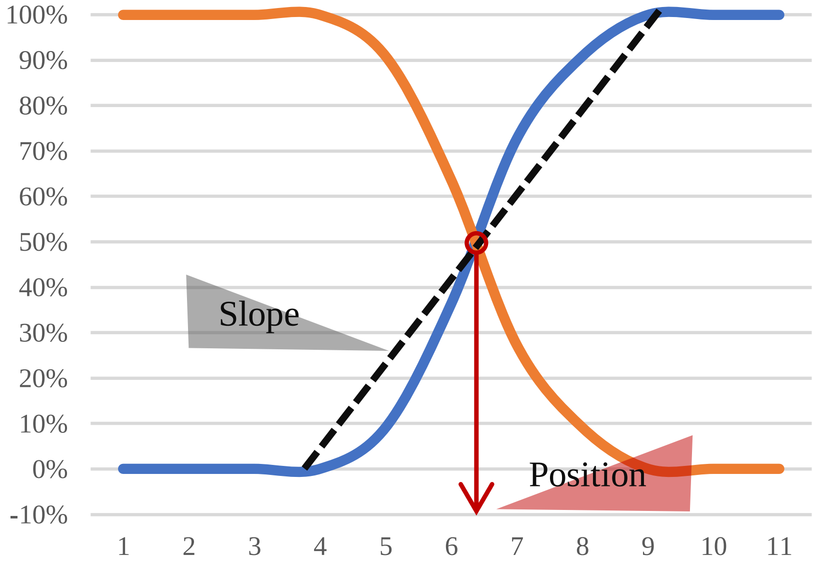


Figure 3.7: The extracted identification parameters: an example

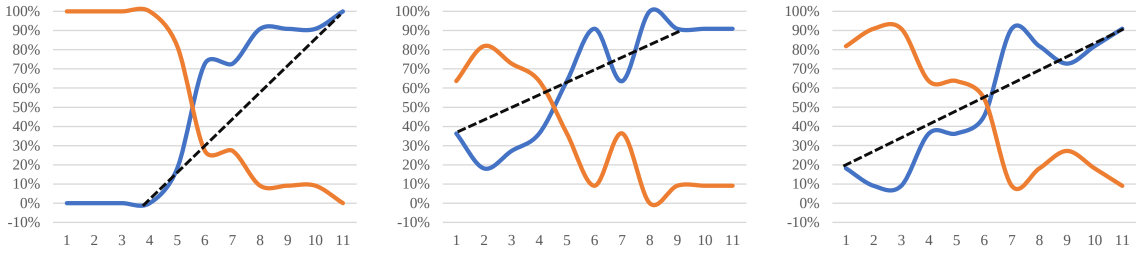


Figure 3.8: More Slope examples of identification rate curves

(3) Statistical analysis

With the extracted parameters, we used R 4.0.5 (Team, 2021) to analyze the three participant groups' performance on different tone pairs, and to check the impact of frequency and duration on tone categorical identification respectively.

In particular, we adopted the Linear mixed-effect models (Gałecki and Burzykowski, 2013) using *lme4* package (Bates et al., 2015) and *lmerTest* package (Kuznetsova et al., 2017) to check the impact of frequency. With R 4.0.5 (Team, 2021), we created the models using *lme4* package (Bates et al., 2015), compared all possible models with the full models to exclude the fixed factors in question using *lmerTest* package (Kuznetsova et al., 2017), and decided the final models with the lowest Akaike information criterions (AICs) and Bayesian information criterions (BICs) (Profillidis and Botzoris, 2019; Elosua and

De Boeck, 2020). Furthermore, besides the check of goodness-of-fit (minus two times of the logarithm of the likelihood) for the fixed and random effects, visual inspection of residual plots was applyed to ensure no obvious deviations from homoscedasticity and normality in all final models.

The final models were as follows:

- The final model of Position was Position ~ Group $*$ TonePair + (1 | Subject), in which Group and TonePair were the independent factors, 1 was the random intercept, and Subject was the random factor. It is noted that the simple coding method was used for all three independent factors in this final model, in which CI and T1T2 were set as the baselines for Group and TonePair respectively.
- The final model of Slope was Slope ~ Group $*$ TonePair + (1 | Subject), in which Group and TonePair were the independent factors, 1 was the random intercept, and Subject was the random factor. It is noted that the simple coding method was used for all three independent factors in this final model, in which CI and T1T2 were set as the baselines for Group and TonePair respectively.

We applied Growth curve analysis (Mirman, 2014) to check the impact of duration. For Position, no model with significant fixed effect was built up; for Slope, the final model was Slope ~ (ot1+ot2+ot3) $*$ Group $*$ Order + (1 | TonePair), in which ot1 to ot3[19], Group, and Order (Order1, in which the duration is steadily decreasing with equal time intervals from D01 to D11; Order2, in which the duration is steadily increasing with equal time intervals from D01 to D11) were the independent factors, 1 was the random intercept, and TonePair was the random factor.

3.2 Results

In this section, we first reported the results of categorical/non-categorical identification, discrimination, and perception based on direct observation and the following Chi-Square tests, and then reported the statistical results of tone categorical identification data.

---

[19] Namely orthogonal time order 1 to 3, please see (Mirman, 2014, pg. 53) for more details.

3.2.1 The results of all choice data

3.2.1.1 Categorical/non-categorical tone identification

(1) Direct observation

Based on the standards of categorical identification presented in 3.1.4.1, we checked each participant's tone identification data by direct observation. For example, Figure 3.9 illustrates the normal hearing participants' identification results of T4T3.

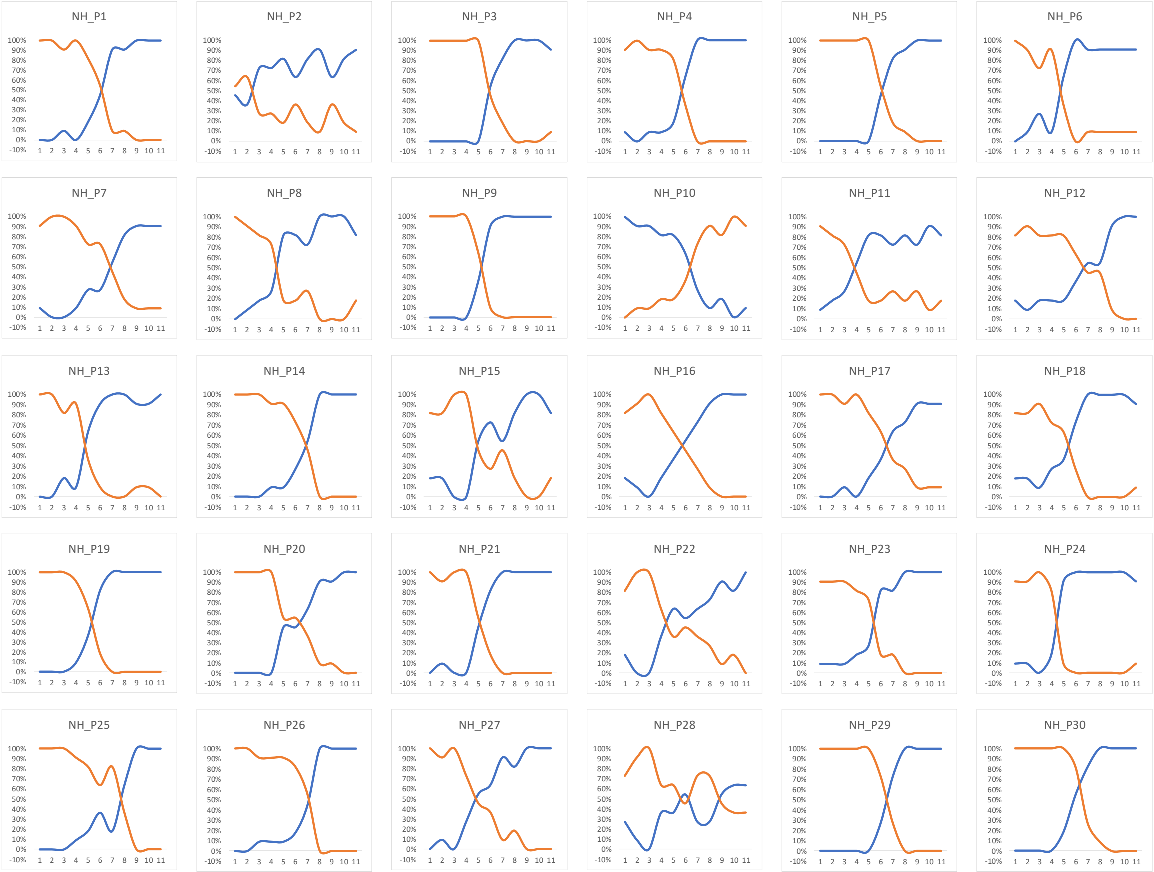


Figure 3.9: The identification results of T4T3 by normal hearing people

From Figure 3.9, one can note that NH_P10's data should be excluded because NH_P10's curves violate the second standard. Observing the two curves of NH_P10, we can see that, although they display a categorical identification between T3 and T4, the two curves violate the second standard, which requires T4 and T3 have a decreasing and an increasing tendency respectively.[20]

---

[20] It may happen that NH_P10 just flipped the keystrokes by accident. However, in this experiment, the

Furthermore, although other normal hearing participants' T4T3 data satisfy all three standards, NH_P2 and NH_P28's data should be excluded because they triggered the first criterion. That is, NH_P2 and NH_P28 chose T4 86 and 44 times respectively, all outside the range of mean±2std (45.13, 82.80; mean = 63.97, std = 9.42).

Table 3.5 displayed the classifying results of the tone identification data, in which the categorical data and non-categorical data were marked as *Cat* and *NonC* respectively.

Table 3.5: The classifying results of tone identification data

	CI			HA			NH		
Tone pair	*Cat*	*NonC*	% of *Cat*	*Cat*	*NonC*	% of *Cat*	*Cat*	*NonC*	% of *Cat*
T1T2	7	13	35.00%	17	9	65.38%	29	1	96.67%
T1T3	15	5	75.00%	19	7	73.08%	29	1	96.67%
T1T4	6	14	30.00%	18	8	69.23%	28	2	93.33%
T2T3	7	13	35.00%	11	15	42.31%	28	2	93.33%
T2T4	3	17	15.00%	7	19	26.92%	30	0	100.00%
T4T3	7	13	35.00%	16	10	61.54%	27	3	90.00%

As expected, Table 3.5 indicated that, while most NH participants could identify the six tone pairs categorically, a few of them failed to categorically identify some tone pairs. In contrast, the deaf participants, especially the CI users, had encountered more difficulties in categorical identification of the tone pairs, especially T2T4.

(2) Chi-Square tests

Based on the numbers listed in Table 3.5, we calculated the total numbers of categorical and non-categorical identification cases for each of the three groups and checked the three groups' differences by using a Chi-Square test with *bonferroni* method. Since the results reported a significant relationship between the three groups, $X^2(2, 76) = 118.350$, $p < 0.001$, multiple comparisons using Chi-Square test were also applied to determine the differences between every two groups. The results, as displayed in Table 3.6, revealed that there were significant differences between every two groups. Combing the results in Table 3.5, the current results demonstrated that the three groups followed the order of CI < HA < NH ("<" means "worse") in categorically identifying synthesized Mandarin tones.

---

trials of T4T3 are mixed with the trials of other tone pairs, and NH_P10 made no such errors for other tone pairs. So, it is probably she/he did a mistake with the two tones but did not just hit the wrong keys accidentally.

Table 3.6: Multiple comparisons of Chi-square tests for identification data: Group

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
CI : HA	8.970	1	0.002	0.007	0.003	0.008
CI : NH	115.000	1	0.000	0.000	0.000	0.000
HA : NH	68.300	1	0.000	0.000	0.000	0.000

Using the same method, we calculated the total numbers of categorical and non-categorical identification cases for each of the six tone pairs and checked their differences. The results reported significant effects among the six pairs, $X^2(5, 76) = 17.487$, $p = 0.004$. As displayed in Table 3.7a, multiple comparisons revealed that only T1T3 vs T2T4 reported significance and T1T3 vs T2T3 reported approaching significance. Considering it would be hard to read the comparisons in Table 3.7a and the following tables that involved many comparison pairs, we also extracted the cldLists (Compact letter display for lists of comparisons table) using *cldList* function in *rcompanion* package and displayed them in Table 3.7b. From Table 3.7b, one could see that the six tone pairs could be classified into three categories: T1T3, T2T4, and other tone pairs reporting no significant difference with both T1T3 and T2T4. Taking the direct observation and statistical results together, the current results indicated that the six tone pairs followed the order of T1T3 $\approx$ T1T2 $\approx$ T1T4 $\approx$ T4T3 $\gtrapprox$ T2T3 $>$ T2T4 ("$>$" means "better with significance","$\approx$" means "similar", and "$\gtrapprox$" means "better with approaching significance") in categorically identifying synthesized Mandarin tones.

Table 3.7: Multiple comparisons of Chi-square tests for identification data: TonePair

(a) Statistics numbers

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
T1T2 : T1T3	2.950	1	0.085	1.000	0.086	1.000
T1T2 : T1T4	0.000	1	1.000	1.000	1.000	1.000
T1T2 : T2T3	1.040	1	0.307	1.000	0.307	1.000
T1T2 : T2T4	3.990	1	0.045	0.681	0.046	0.687
T1T2 : T4T3	0.120	1	0.729	1.000	0.729	1.000
T1T3 : T1T4	3.570	1	0.058	0.868	0.059	0.882
T1T3 : T2T3	8.300	1	0.004	0.055	0.004	0.059
T1T3 : T2T4	14.600	1	0.000	0.002	0.000	0.002
T1T3 : T4T3	4.970	1	0.025	0.376	0.026	0.387
T1T4 : T2T3	0.718	1	0.397	1.000	0.397	1.000
T1T4 : T2T4	3.330	1	0.068	1.000	0.068	1.000
T1T4 : T4T3	0.030	1	0.863	1.000	0.863	1.000
T2T3 : T2T4	0.669	1	0.413	1.000	0.413	1.000
T2T3 : T4T3	0.254	1	0.614	1.000	0.614	1.000
T2T4 : T4T3	2.210	1	0.137	1.000	0.137	1.000

(b) ckdList

T1T2	T1T3	T1T4	T2T3	T2T4	T4T3
ab	a	ab	ab	b	ab

To further explore the interactions between Group and TonePair, we checked the differences among the three groups in each tone pair and the differences among the six tone pairs in each participant group respectively.

From the angle of the differences among the three groups in each tone pair, the results of Chi-square tests and the following multiple comparisons reported significant effects in all tone pairs except for the case of T1T3. Table 3.8 displays the results of multiple comparisons between every two groups for each tone pair. Combing the results of direct observation and Chi-square tests together, one could see that: first, the three groups had no significant difference in identifying T1T3 trials; both CI and HA were significantly different from NH in identifying T1T2, T2T3, and T2T4 cases; CI was significantly different from HA and NH in identifying T1T4 trials; CI and NH were significantly different from each other in identifying T4T3 tasks, and HA fell somewhere in the middle ground between CI and NH.

Table 3.8: Multiple comparisons of identification data between groups in each tone pair

(a) Statistics numbers

Tone pair	Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
TIT2	CI : HA	3.050	1	0.073	0.219	0.081	0.242
	CI : NH	19.700	1	0.000	0.000	0.000	0.000
	HA : NH	7.280	1	0.004	0.011	0.007	0.021
TIT3	CI : HA	0.000	1	1.000	1.000	1.000	1.000
	CI : NH	3.480	1	0.032	0.095	0.062	0.186
	HA : NH	4.550	1	0.019	0.057	0.033	0.099
TIT4	CI : HA	5.490	1	0.016	0.049	0.019	0.057
	CI : NH	19.300	1	0.000	0.000	0.000	0.000
	HA : NH	4.000	1	0.033	0.100	0.046	0.137
T2T3	CI : HA	0.040	1	0.763	1.000	0.842	1.000
	CI : NH	16.800	1	0.000	0.000	0.000	0.000
	HA : NH	14.800	1	0.000	0.000	0.000	0.000
T2T4	CI : HA	0.374	1	0.476	1.000	0.541	1.000
	CI : NH	34.900	1	0.000	0.000	0.000	0.000
	HA : NH	30.000	1	0.000	0.000	0.000	0.000
T4T3	CI : HA	2.210	1	0.136	0.408	0.137	0.411
	CI : NH	14.300	1	0.000	0.000	0.000	0.000
	HA : NH	4.830	1	0.024	0.072	0.028	0.084

(b) ckdList

Group	T1T2	T1T3	T1T4	T2T3	T2T4	T4T3
CI	a	a	a	a	a	a
HA	a	a	b	a	a	ab
NH	b	a	b	b	b	b

From the angle of the differences among the six tone pairs in each group, the results of Chi-square tests and the following multiple comparisons reported that CI and HA but not NH had significant differences among the six tone pairs. Table 3.9 displays the results of multiple comparisons between every two tone pairs for each participant group. When taking the results of direct observation and Chi-square tests together, one could see that: while no significant difference was found between the tone pairs, both CI and HA followed the general order of Mandarin tone identification. Namely, the two deaf groups followed the order of T1T3 $\approx$ T1T2 $\approx$ T1T4 $\approx$ T4T3 $\gtrapprox$ T2T3 $>$ T2T4 ("$>$" means "better with significance","$\approx$" means "similar", and "$\gtrapprox$" means "better with approaching significance") in categorically identifying synthesized Mandarin tones.

Table 3.9: Multiple comparisons of perception data between tone pairs in each group

(a) Statistics numbers: CI

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
T1T2 : T1T3	4.950	1.000	0.025	0.372	0.026	0.392
T1T2 : T1T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T2T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T2T4	1.200	1.000	0.273	1.000	0.273	1.000
T1T2 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T3 : T1T4	6.420	1.000	0.010	0.156	0.011	0.169
T1T3 : T2T3	4.950	1.000	0.025	0.372	0.026	0.392
T1T3 : T2T4	12.200	1.000	0.000	0.005	0.000	0.007
T1T3 : T4T3	4.950	1.000	0.025	0.372	0.026	0.392
T1T4 : T2T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T4 : T2T4	0.573	1.000	0.451	1.000	0.449	1.000
T1T4 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T2T3 : T2T4	1.200	1.000	0.273	1.000	0.273	1.000
T2T3 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T2T4 : T4T3	1.200	1.000	0.273	1.000	0.273	1.000

(b) Statistics numbers: HA

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
T1T2 : T1T3	0.090	1.000	0.764	1.000	0.764	1.000
T1T2 : T1T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T2T3	1.930	1.000	0.164	1.000	0.164	1.000
T1T2 : T2T4	6.270	1.000	0.011	0.172	0.012	0.184
T1T2 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T3 : T1T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T3 : T2T3	3.860	1.000	0.048	0.724	0.049	0.741
T1T3 : T2T4	9.310	1.000	0.002	0.030	0.002	0.034
T1T3 : T4T3	0.350	1.000	0.555	1.000	0.554	1.000
T1T4 : T2T3	2.810	1.000	0.093	1.000	0.094	1.000
T1T4 : T2T4	7.700	1.000	0.005	0.075	0.006	0.083
T1T4 : T4T3	0.085	1.000	0.771	1.000	0.771	1.000
T2T3 : T2T4	0.765	1.000	0.382	1.000	0.382	1.000
T2T3 : T4T3	1.230	1.000	0.267	1.000	0.267	1.000
T2T4 : T4T3	4.990	1.000	0.024	0.368	0.025	0.382

(c) Statistics numbers: NH

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
T1T2 : T1T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T1T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T2T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T2T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T4T3	0.268	1.000	0.612	1.000	0.605	1.000
T1T3 : T1T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T3 : T2T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T3 : T2T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T3 : T4T3	0.268	1.000	0.612	1.000	0.605	1.000
T1T4 : T2T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T4 : T2T4	0.517	1.000	0.492	1.000	0.472	1.000
T1T4 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T2T3 : T2T4	0.517	1.000	0.492	1.000	0.472	1.000
T2T3 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T2T4 : T4T3	1.400	1.000	0.237	1.000	0.236	1.000

(d) ckdList

Group	T1T2	T1T3	T1T4	T2T3	T2T4	T4T3
CI	ab	a	ab	ab	b	ab
HA	ab	a	ab	ab	b	ab
NH	a	a	a	a	a	a

(3) Summary

To summarize, the results of direct observation and Chi-Squre tests supported the following points. Firstly, the three groups were different in their ability to categorically identify Mandarin tones. That is, while the CI group was significantly worse than the HA group, both of them are significantly worse than the NH group in categorically identifying Mandarin tones. Secondly, the difficulty degrees of categorical identification were significantly different among the tone pairs. Namely, T2T4 was significantly harder than T1T3, and other tone pairs were lying somewhere in the middle. Finally, there were significant interactions between the participant groups and tone pairs. That is to say, while the NH group performed similarly well in all tone pairs, both of the two deaf groups experienced more difficulties in categorically identifying T2T4 and T2T3 than in other tone pairs.

3.2.1.2 Categorical/non-categorical tone discrimination

(1) Direct observation

Based on the standards of categorical discrimination presented in 3.1.4.1, we checked each and every participant's data by direct observation. For example, Figure 3.10 illustrates the normal hearing participants' discrimination results of T1T2.

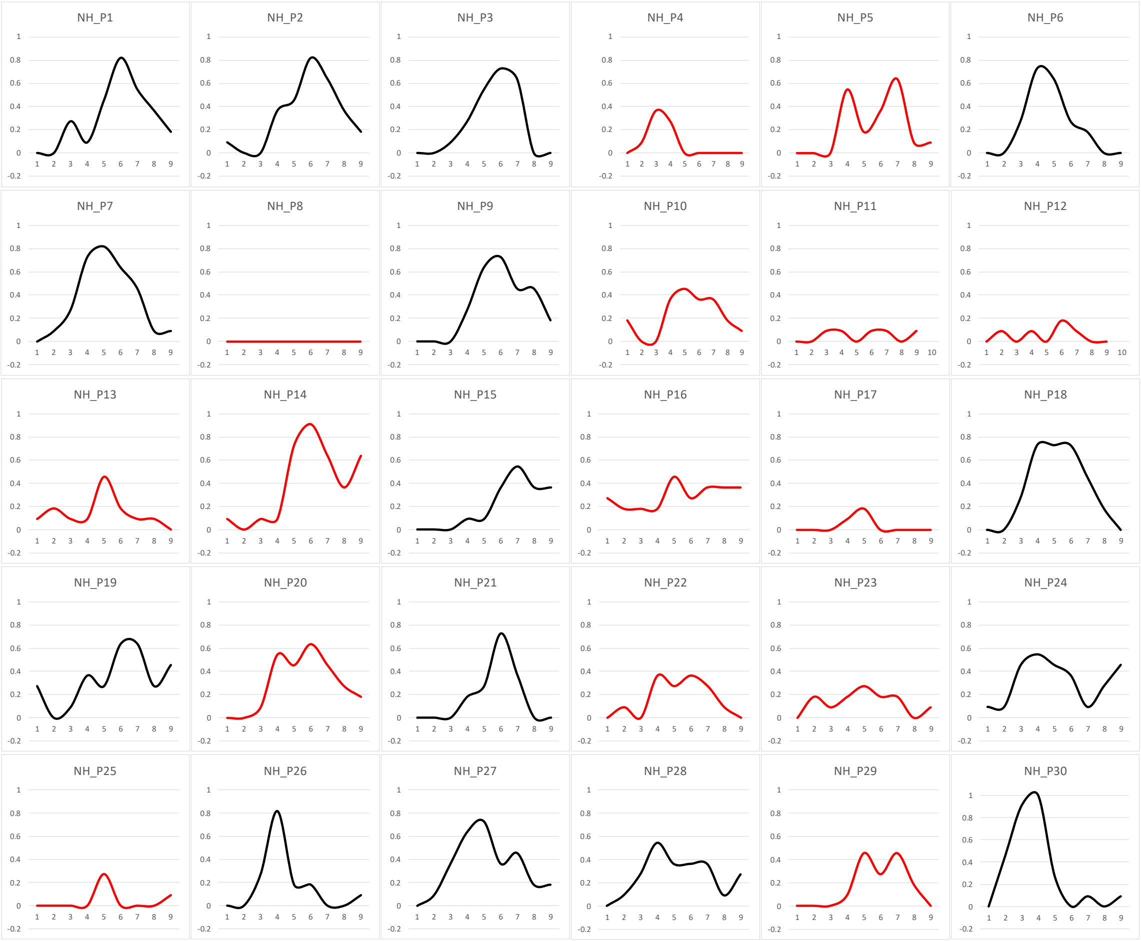


Figure 3.10: The discrimination results of T1T2 by normal hearing people

From Figure 3.10, one can see that 15 participants' curves (black) are considered as categorical discrimination cases. As for the other 15 participants' curves (red), because of at least violating one of the three standards, are judged as non-categorical cases.

Table 3.10 displayed the classifying results of the tone discrimination data, in which the categorical data and non-categorical data were marked as *Cat* and *NonC* respectively.

Table 3.10: The classifying results of tone discrimination data

	CI			HA			NH		
Tone pair	*Cat*	*NonC*	% of *Cat*	*Cat*	*NonC*	% of *Cat*	*Cat*	*NonC*	% of *Cat*
T1T2	5	15	25.00%	11	14	44.00%	15	15	50.00%
T1T3	1	19	5.00%	9	16	36.00%	20	10	66.67%
T1T4	4	16	20.00%	7	18	28.00%	17	13	56.67%
T2T3	6	14	30.00%	5	20	20.00%	6	24	20.00%
T2T4	4	16	20.00%	2	23	8.00%	15	15	50.00%
T4T3	3	17	15.00%	5	20	20.00%	13	17	43.33%

Compared with Table 3.5, Table 3.10 indicates that all participants experience more obstacles in categorically discriminating than identifying Mandarin tone pairs. From Table 3.10, one can see that, besides deaf participants, a considerable number of normal hearing participants could not categorically discriminate all tone pairs. Especially in T2T3, only six participants in NH group could categorically discriminate the two tones, which might indicate that categorically discriminate this tone pair is a hard task even for normal hearing people.

(2) Chi-square tests

After calculating the total numbers of categorical and non-categorical discrimination cases for each of the three groups, we checked the three groups' differences using a Chi-Square test and found a significant relationship between the three groups, $X^2(2, 76) = 32.926$, $p < 0.001$. Moreover, the results of multiple comparisons, as displayed in Table 3.11, revealed that both CI and HA were significantly different from NH in categorically discriminating synthesized Mandarin tones, but the two deaf groups had no significant difference in their performances. Combing the results of direct observation and statistical results, the three groups followed the order of CI ≈ HA < NH ("<" means "worse" and "≈" means "similar") in categorically discriminating synthesized Mandarin tones.

Table 3.11: Multiple comparisons of Chi-square tests for discrimination data: Group

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
CI : HA	1.820	1.000	0.145	0.435	0.177	0.531
CI : NH	25.800	1.000	0.000	0.000	0.000	0.000
HA : NH	15.600	1.000	0.000	0.000	0.000	0.000

As for the case of categorical and non-categorical discrimination data for each of the six tone pairs, both the Chi-Square test and the following multiple comparisons displayed in Table 3.12 failed to report any significance, indicating that the six tone pairs were generally similar in difficulty degrees.

Table 3.12: Multiple comparisons of Chi-square tests for discrimation data: TonePair

(a) Statistics numbers

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
T1T2 : T1T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T1T4	0.112	1.000	0.738	1.000	0.738	1.000
T1T2 : T2T3	6.070	1.000	0.013	0.199	0.014	0.206
T1T2 : T2T4	2.380	1.000	0.122	1.000	0.123	1.000
T1T2 : T4T3	2.970	1.000	0.084	1.000	0.085	1.000
T1T3 : T1T4	0.112	1.000	0.738	1.000	0.738	1.000
T1T3 : T2T3	6.070	1.000	0.013	0.199	0.014	0.206
T1T3 : T2T4	2.380	1.000	0.122	1.000	0.123	1.000
T1T3 : T4T3	2.970	1.000	0.084	1.000	0.085	1.000
T1T4 : T2T3	3.890	1.000	0.048	0.718	0.048	0.728
T1T4 : T2T4	1.090	1.000	0.296	1.000	0.296	1.000
T1T4 : T4T3	1.500	1.000	0.220	1.000	0.220	1.000
T2T3 : T2T4	0.574	1.000	0.449	1.000	0.449	1.000
T2T3 : T4T3	0.329	1.000	0.567	1.000	0.566	1.000
T2T4 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000

(b) ckdList

T1T2	T1T3	T1T4	T2T3	T2T4	T4T3
a	a	a	a	a	a

The results of Chi-square tests for the participant groups' performances in each tone pair, as displayed in Table 3.13, reported that the participant groups were only significantly different in categorically discriminating T1T3 and T2T4 trials. Combing the results of direct observation and Chi-square tests together, one could see that: firstly, the CI group performed significantly worse than the HA and NH groups in discriminating T1T3 trials; secondly, HA and NH were significantly different in discriminating T2T4 cases, and the CI group fell somewhere in the middle; finally, all three groups had no significant difference in other tone pairs.

The results of Chi-square tests for the differences of tone pairs in each participant group, as displayed in Table 3.14, reported that only the NH group's performances were significantly different among the six tone pairs. That is, the NH group's performance in T1T3 was significantly different from T2T3, and this group's performance in other tone pairs fell somewhere in the middle. Taking the results of direct observation and Chi-square tests, one could see that: while no significant difference in difficulty degree between the tone pairs was found in the CI and HA groups, the NH group followed the order of T1T3 $>$ T1T2 $\approx$ T1T4 $\approx$ T2T4 $\approx$ T4T3 $>$ T2T3 ("$>$" means "better with significance" and "$\approx$" means "similar") in categorically discriminating synthesized Mandarin tones.

(3) Summary

Table 3.13: Multiple comparisons of discrimination data between groups in each tone pair

(a) Statistics numbers

Tone pair	Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
TIT2	CI : HA	1.020	1.000	0.224	0.672	0.313	0.939
	CI : NH	2.170	1.000	0.140	0.420	0.141	0.423
	HA : NH	0.030	1.000	0.788	1.000	0.863	1.000
TIT3	CI : HA	5.600	1.000	0.012	0.037	0.018	0.054
	CI : NH	16.300	1.000	0.000	0.000	0.000	0.000
	HA : NH	2.910	1.000	0.061	0.183	0.088	0.264
TIT4	CI : HA	0.074	1.000	0.729	1.000	0.786	1.000
	CI : NH	5.200	1.000	0.018	0.055	0.022	0.068
	HA : NH	3.470	1.000	0.055	0.166	0.063	0.188
T2T3	CI : HA	0.002	1.000	0.731	1.000	0.968	1.000
	CI : NH	0.005	1.000	0.736	1.000	0.944	1.000
	HA : NH	0.000	1.000	1.000	1.000	1.000	1.000
T2T4	CI : HA	0.541	1.000	0.383	1.000	0.462	1.000
	CI : NH	3.400	1.000	0.041	0.123	0.065	0.196
	HA : NH	9.380	1.000	0.001	0.003	0.002	0.007
T4T3	CI : HA	0.000	1.000	1.000	1.000	1.000	1.000
	CI : NH	3.220	1.000	0.062	0.186	0.073	0.218
	HA : NH	3.580	1.000	0.041	0.123	0.059	0.176

(b) ckdList

Group	T1T2	T1T3	T1T4	T2T3	T2T4	T4T3
CI	a	a	a	a	ab	a
HA	a	b	a	a	a	a
NH	a	b	a	a	b	a

Table 3.14: Multiple comparisons of discrimination data between tone pairs in each group

(a) Statistics numbers: CI

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
T1T2 : T1T3	1.760	1.000	0.182	1.000	0.184	1.000
T1T2 : T1T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T2T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T2T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T4T3	0.156	1.000	0.695	1.000	0.693	1.000
T1T3 : T1T4	0.914	1.000	0.342	1.000	0.339	1.000
T1T3 : T2T3	1.760	1.000	0.182	1.000	0.184	1.000
T1T3 : T2T4	0.914	1.000	0.342	1.000	0.339	1.000
T1T3 : T4T3	0.278	1.000	0.605	1.000	0.598	1.000
T1T4 : T2T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T4 : T2T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T4 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T2T3 : T2T4	0.000	1.000	1.000	1.000	1.000	1.000
T2T3 : T4T3	0.156	1.000	0.695	1.000	0.693	1.000
T2T4 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000

(b) Statistics numbers: HA

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
T1T2 : T1T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T1T4	0.781	1.000	0.377	1.000	0.377	1.000
T1T2 : T2T3	2.300	1.000	0.128	1.000	0.130	1.000
T1T2 : T2T4	6.650	1.000	0.008	0.124	0.010	0.149
T1T2 : T4T3	3.430	1.000	0.062	0.933	0.064	0.962
T1T3 : T1T4	0.357	1.000	0.551	1.000	0.550	1.000
T1T3 : T2T3	1.520	1.000	0.217	1.000	0.217	1.000
T1T3 : T2T4	5.370	1.000	0.018	0.272	0.020	0.308
T1T3 : T4T3	2.480	1.000	0.114	1.000	0.115	1.000
T1T4 : T2T3	0.110	1.000	0.742	1.000	0.741	1.000
T1T4 : T2T4	2.170	1.000	0.138	1.000	0.141	1.000
T1T4 : T4T3	0.466	1.000	0.496	1.000	0.495	1.000
T2T3 : T2T4	0.664	1.000	0.417	1.000	0.415	1.000
T2T3 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T2T4 : T4T3	0.189	1.000	0.667	1.000	0.663	1.000

(c) Statistics numbers: NH

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
T1T2 : T1T3	1.100	1.000	0.295	1.000	0.295	1.000
T1T2 : T1T4	0.067	1.000	0.796	1.000	0.796	1.000
T1T2 : T2T3	4.690	1.000	0.029	0.438	0.030	0.456
T1T2 : T2T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T4T3	0.067	1.000	0.796	1.000	0.796	1.000
T1T3 : T1T4	0.282	1.000	0.596	1.000	0.595	1.000
T1T3 : T2T3	11.500	1.000	0.001	0.009	0.001	0.011
T1T3 : T2T4	1.100	1.000	0.295	1.000	0.295	1.000
T1T3 : T4T3	2.420	1.000	0.119	1.000	0.119	1.000
T1T4 : T2T3	7.050	1.000	0.007	0.109	0.008	0.119
T1T4 : T2T4	0.067	1.000	0.796	1.000	0.796	1.000
T1T4 : T4T3	0.600	1.000	0.439	1.000	0.439	1.000
T2T3 : T2T4	4.690	1.000	0.029	0.438	0.030	0.456
T2T3 : T4T3	2.770	1.000	0.095	1.000	0.096	1.000
T2T4 : T4T3	0.067	1.000	0.796	1.000	0.796	1.000

(d) ckdList

Group	T1T2	T1T3	T1T4	T2T3	T2T4	T4T3
CI	a	a	a	a	a	a
HA	a	a	a	a	a	a
NH	ab	a	ab	b	ab	ab

To summarize, the results of direct observation and Chi-Squre tests supported the following points. Firstly, the three groups were different in their ability to categorically discriminate Mandarin tones. That is, although there was no significant difference between the CI group and the HA group, both of them are significantly worse than the NH group in categorically discriminating Mandarin tones. Secondly, in general, there was no significant difference in the difficulty degree of categorical identification among the tone pairs. Finally, while all six tone pairs were similarly difficult for the two deaf groups, the NH group experienced more difficulties in categorically identifying T2T3 cases than other tone pairs.

3.2.1.3 Categorical/non-categorical tone perception

(1) Direct observation

Using the standard of categorical perception presented in 3.1.4.1, the current study checked the categorical perception of the six tone pairs by the three participants groups. Table 3.15 displays the numbers of categorical identification (*Iden*), categorical discrimination (*Disc*), and categorical perception (*Perc*) for all three participant groups.

Table 3.15: The results of categorical identification, discrimination, and perception

Tone pair	HA			CI			NH		
	Iden	*Disc*	*Perc*	*Iden*	*Disc*	*Perc*	*Iden*	*Disc*	*Perc*
T1T2	17	11	9	7	5	1	29	15	15
T1T3	19	9	9	15	1	1	29	20	20
T1T4	18	7	5	6	4	1	28	17	15
T2T3	11	5	1	7	6	3	28	6	6
T2T4	7	2	2	3	4	1	30	15	14
T4T3	16	5	4	7	3	0	27	13	7

From Table 3.15, besides the fact that there are apparently more *Iden* than *Disc*, one might note that the numbers of *Perc* are smaller than that of *Disc* for all participant groups. After checking the data thoroughly, we found that there are 17 cases in which the participants (HA: 4, CI: 4, and NH: 9) could not categorically perceive the tone pairs because they failed to overlap the identification boundaries and the discrimination peaks, and there are other 13 cases in which the participants (HA: 9, CI: 4, and NH: 0) who could categorically discriminate given tone pairs had failed to identify them categorically.

Based on the above results, we divided the perception data into categorical and non-cateogircal data listed in Table 3.16. From Table 3.16, it seemed that the deaf groups performed much worse than the NH group in categorical perception of synthesized Mandarin tones.

Table 3.16: The classifying results of tone perception data

	CI			HA			NH		
Tone pair	*Cat*	*NonC*	% of *Cat*	*Cat*	*NonC*	% of *Cat*	*Cat*	*NonC*	% of *Cat*
T1T2	1	19	5.00%	9	17	34.62%	15	15	50.00%
T1T3	1	19	5.00%	9	17	34.62%	20	10	66.67%
T1T4	1	19	5.00%	5	21	19.23%	15	15	50.00%
T2T3	3	17	15.00%	1	25	3.85%	6	24	20.00%
T2T4	1	19	5.00%	2	24	7.69%	14	16	46.67%
T4T3	0	20	0.00%	4	22	15.38%	7	23	23.33%

(2) Chi-square tests

Based on the data displayed in Table 3.16, we examed the three groups' general differences in categorical perception using a battery of Chi-Square tests. Since the reuslts [$X^2(2, 76) = 56.621$, $p < 0.001$] and Multiple comparisons displayed Table 3.17 reported significant effects, indicating the three groups were significantly different from each other in categorically perceving the synthesized Mandarin tones.

Table 3.17: Multiple comparisons of Chi-square tests for perception data: Group

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
CI : HA	9.360	1.000	0.001	0.004	0.002	0.007
CI : NH	46.900	1.000	0.000	0.000	0.000	0.000
HA : NH	20.300	1.000	0.000	0.000	0.000	0.000

As for the general differences among the tone pairs, the Chi-Square test reported a significant effect, $X^2(5, 76) = 21.754$, $p = 0.004$. The following multiple comparisons, as displayed in Table 3.18, the six tone pairs could be classified into three categories: T1T3, T2T3 and T4T3, and other tone pairs reporting no significant difference with the first two categories. Combing the results of direct observation and statisitcal analysis, the six tone pairs followed the order of T1T3 > T1T2 ≈ T1T4 ≈ T2T4 > T2T3 ≈ T4T3 (">" means "better with significance" and "≈" means "similar") in categorically perceiving synthesized Mandarin tones.

As displayed in Table 3.19, the participant groups performances in each of the six tone pairs demonstrated a complex pattern: in T1T2, T1T3, and T1T4, while CI and NH were

Table 3.18: Multiple comparisons of Chi-square tests for perception data: TonePair

(a) Statistics numbers

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
T1T2 : T1T3	0.456	1.000	0.500	1.000	0.500	1.000
T1T2 : T1T4	0.281	1.000	0.597	1.000	0.596	1.000
T1T2 : T2T3	7.280	1.000	0.006	0.097	0.007	0.105
T1T2 : T2T4	1.610	1.000	0.204	1.000	0.204	1.000
T1T2 : T4T3	6.150	1.000	0.012	0.188	0.013	0.196
T1T3 : T1T4	1.890	1.000	0.169	1.000	0.169	1.000
T1T3 : T2T3	12.200	1.000	0.000	0.006	0.000	0.007
T1T3 : T2T4	4.440	1.000	0.035	0.519	0.035	0.528
T1T3 : T4T3	10.800	1.000	0.001	0.013	0.001	0.015
T1T4 : T2T3	4.050	1.000	0.043	0.645	0.044	0.662
T1T4 : T2T4	0.316	1.000	0.575	1.000	0.574	1.000
T1T4 : T4T3	3.210	1.000	0.072	1.000	0.073	1.000
T2T3 : T2T4	1.620	1.000	0.202	1.000	0.203	1.000
T2T3 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T2T4 : T4T3	1.090	1.000	0.295	1.000	0.295	1.000

(b) ckdList

T1T2	T1T3	T1T4	T2T3	T2T4	T4T3
ab	a	ab	b	ab	b

significantly different, HA reported no significance both with CI and NH; in T2T4, HA and NH reported a significant difference between each other but both of them were not significantly different from CI; in T2T3 and T4T3, all three groups failed to report any significance between each other. Combing the results of direct observation and Chi-square tests together, one could see that: first, all three groups similarly performed bad in T2T3 and T4T3; second, HA performed signficantly worse than NH in categorically peceiving T2T4, and CI fell somewhere in the middle between HA and NH; third, CI performed signficantly worse than NH in categorically peceiving T1T2, T1T3, and T1T4, and HA lied somewhere in the middle ground between CI and NH.

Table 3.19: Multiple comparisons of perception data between groups in each tone pair

(a) Statistics numbers

Tone pair	Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
TIT2	CI : HA	4.220	1.000	0.028	0.083	0.040	0.120
	CI : NH	9.200	1.000	0.001	0.002	0.002	0.007
	HA : NH	0.791	1.000	0.288	0.864	0.374	1.000
TIT3	CI : HA	4.220	1.000	0.028	0.083	0.040	0.120
	CI : NH	16.300	1.000	0.000	0.000	0.000	0.000
	HA : NH	4.520	1.000	0.031	0.093	0.034	0.100
TIT4	CI : HA	0.959	1.000	0.212	0.636	0.328	0.984
	CI : NH	9.200	1.000	0.001	0.002	0.002	0.007
	HA : NH	4.480	1.000	0.025	0.075	0.034	0.103
T2T3	CI : HA	0.645	1.000	0.303	0.909	0.422	1.000
	CI : NH	0.006	1.000	0.724	1.000	0.940	1.000
	HA : NH	2.010	1.000	0.108	0.324	0.156	0.468
T2T4	CI : HA	0.000	1.000	1.000	1.000	1.000	1.000
	CI : NH	8.040	1.000	0.002	0.005	0.005	0.014
	HA : NH	8.550	1.000	0.002	0.007	0.003	0.010
T4T3	CI : HA	1.710	1.000	0.121	0.363	0.191	0.573
	CI : NH	3.660	1.000	0.033	0.098	0.056	0.167
	HA : NH	0.168	1.000	0.517	1.000	0.682	1.000

(b) ckdList

Group	T1T2	T1T3	T1T4	T2T3	T2T4	T4T3
CI	a	a	a	a	ab	a
HA	ab	ab	ab	a	a	a
NH	b	b	b	a	b	a

As for the differences of each participant group's categorical perception among the six tone pairs, as displayed in Table 3.20, the current results revealed the similar patterns to that of categorical discrimination: only the NH group's performance in T1T3 was significantly different from T2T3, and this group's performance in other tone pairs fell somewhere in the middle; in contrast, no significance was found in CI and HA's performance in all tone pairs. Taking the results of direct observation and Chi-square tests, one could see that: while no significant difference in difficulty degree between the tone pairs was found in the CI and HA groups, the NH group followed the order of T1T3 > T1T2 ≈ T1T4 ≈ T2T4 ≈ T4T3 > T2T3 (">" means "better with significance" and "≈" means "similar") in categorically perceiving synthesized Mandarin tones.

Table 3.20: Multiple comparisons of perception data between tone pairs in each group

(a) Statistics numbers:CI

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
T1T2 : T1T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T1T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T2T3	0.278	1.000	0.605	1.000	0.598	1.000
T1T2 : T2T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T3 : T1T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T3 : T2T3	0.278	1.000	0.605	1.000	0.598	1.000
T1T3 : T2T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T3 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T4 : T2T3	0.278	1.000	0.605	1.000	0.598	1.000
T1T4 : T2T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T4 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T2T3 : T2T4	0.278	1.000	0.605	1.000	0.598	1.000
T2T3 : T4T3	1.440	1.000	0.231	1.000	0.230	1.000
T2T4 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000

(b) Statistics numbers: HA

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
T1T2 : T1T3	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T1T4	0.880	1.000	0.349	1.000	0.348	1.000
T1T2 : T2T3	6.070	1.000	0.011	0.164	0.014	0.207
T1T2 : T2T4	4.150	1.000	0.038	0.578	0.042	0.624
T1T2 : T4T3	1.640	1.000	0.199	1.000	0.200	1.000
T1T3 : T1T4	0.880	1.000	0.349	1.000	0.348	1.000
T1T3 : T2T3	6.070	1.000	0.011	0.164	0.014	0.207
T1T3 : T2T4	4.150	1.000	0.038	0.578	0.042	0.624
T1T3 : T4T3	1.640	1.000	0.199	1.000	0.200	1.000
T1T4 : T2T3	1.700	1.000	0.191	1.000	0.193	1.000
T1T4 : T2T4	0.660	1.000	0.419	1.000	0.416	1.000
T1T4 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T2T3 : T2T4	0.000	1.000	1.000	1.000	1.000	1.000
T2T3 : T4T3	0.885	1.000	0.350	1.000	0.347	1.000
T2T4 : T4T3	0.188	1.000	0.668	1.000	0.664	1.000

(c) Statistics numbers: NH

Comparison	X^2	df	p.Fisher	p.adj.Fisher	p.Chisq	p.adj.Chisq
T1T2 : T1T3	1.100	1.000	0.295	1.000	0.295	1.000
T1T2 : T1T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T2T3	4.690	1.000	0.029	0.438	0.030	0.456
T1T2 : T2T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T2 : T4T3	3.520	1.000	0.060	0.894	0.061	0.912
T1T3 : T1T4	1.100	1.000	0.295	1.000	0.295	1.000
T1T3 : T2T3	11.500	1.000	0.001	0.009	0.001	0.011
T1T3 : T2T4	1.700	1.000	0.192	1.000	0.193	1.000
T1T3 : T4T3	9.700	1.000	0.002	0.024	0.002	0.028
T1T4 : T2T3	4.690	1.000	0.029	0.438	0.030	0.456
T1T4 : T2T4	0.000	1.000	1.000	1.000	1.000	1.000
T1T4 : T4T3	3.520	1.000	0.060	0.894	0.061	0.912
T2T3 : T2T4	3.680	1.000	0.054	0.808	0.055	0.828
T2T3 : T4T3	0.000	1.000	1.000	1.000	1.000	1.000
T2T4 : T4T3	2.640	1.000	0.103	1.000	0.104	1.000

(d) ckdList

Group	T1T2	T1T3	T1T4	T2T3	T2T4	T4T3
CI	a	a	a	a	a	a
HA	a	a	a	a	a	a
NH	ab	a	ab	b	ab	ab

(3) Summary

To summarize, the results of direct observation and Chi-Squre tests supported the following points. Firstly, the three groups were different from each other in their abilities to categorically perceive Mandarin tones. That is, CI performed worse than HA and HA performed worse than NH. Secondly, the six tone pairs followed the general order of T1T3 > T1T2 ≈ T1T4 ≈ T2T4 > T2T3 ≈ T4T3 (">" means "better with significance" and "≈" means "similar") in categorically perceiving synthesized Mandarin tones. Finally, while all six tone pairs were similarly difficult to perceive for the two deaf groups, it seemed that T2T3 and T1T3 were the hardest and the easiest tone pairs to categorically perceive respectively for the NH group.

3.2.2 Statistic results for the categorical identification data

Because the small numbers of participants, espeically the deaf participants could categorically discriminate and perceive synthesized Mandarin tones, this section only focused on categorical identification data to statisitcally analyze the imapct of stimulus frequency and duration to the participants' performance.

3.2.2.1 The impact of the stimulus frequency

(1) Position

Figure 3.11 displays the three groups' means and standard errors (SEs) of boundary position (the 50% crossover point) values in all six tone pairs. From Figure 3.11, one could note that the CI group's T2T4 has a much earlier boundary position than all other cases. Moreover, the boundary positions of the NH group's T1T4 and all three groups' T4T3 are also coming earlier when compared with the remaining cases.

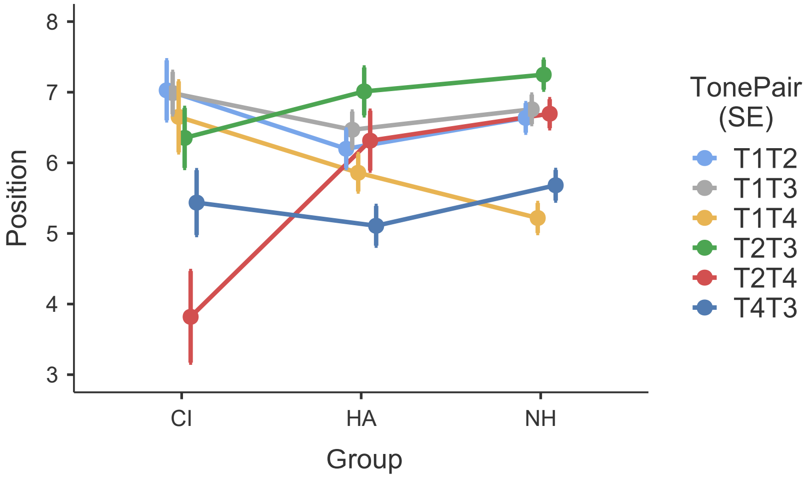


Figure 3.11: The means and standard errors of Position

Table 3.21 presents the ANOVA results of the final model for Position, in which the main effects of TonePair and the interaction of TonePair and Group are significant. The following check of the simple effects reports that the difference between NH and CI's T1T4, NH and CI's T2T4, and HA and CI's T2T4 are significant.

Table 3.21: The ANOVA table of the final model for Position

	F	Num df	Den df	*p*
TonePair	10.75	5	248.5	< 0.001
Group	1.48	2	64.2	0.234
TonePair × group	3.68	10	242.9	< 0.001

Note: Satterthwaite method for degrees of freedom.

The estimates of the final model for Position are displayed in Table 3.22. To save space, only the effects with significance or approaching significance are included in this table.

Table 3.22: The parameter estimates of the fixed effects for Position

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	6.1929	0.0886	6.01930	6.366	73.8	69.9357	< 0.001
TonePair2	T1T4 - T1T2	-0.7108	0.2677	-1.23558	-0.186	250.2	-2.6551	0.008
TonePair4	T2T4 - T1T2	-1.0110	0.3194	-1.63699	-0.385	260.1	-3.1655	0.002
TonePair5	T4T3 - T1T2	-1.2115	0.2614	-1.72384	-0.699	252.3	-4.6341	< 0.001
TonePair3 × Group1	T2T3 - T1T2 × HA - CI	1.4869	0.7307	0.05477	2.919	254.2	2.0349	0.043
TonePair3 × Group2	T2T3 - T1T2 × NH - CI	1.2868	0.6595	-0.00569	2.579	248.7	1.9513	0.052
TonePair4 × Group1	T2T4 - T1T2 × HA - CI	3.3229	0.9147	1.53019	5.116	263.7	3.6329	< 0.001
TonePair4 × Group2	T2T4 - T1T2 × NH - CI	3.2671	0.8187	1.66243	4.872	262.0	3.9905	< 0.001

From Table 3.22, one could note that the main effects of T1T4, T2T4, and T4T3 are significant and negative, indicating that their boundary positions are apparently to the left of the boundary position of T1T2. In contrast, all remaining tone pairs fail to report any significance, showing that they share similar boundary positions to T1T2. As for the interactions, the effect estimates of the HA and NH group's T2T4-T1T2, and the HA group's T2T3-T1T2 are significant and positive, as well as the NH group's T2T3-T1T2 are approaching significant and positive, revealing that the CI group's differences in boundary positions are much larger than the other two participant groups.

Overall, statistical analysis for Position indicated that the deaf participants who had set up similar categorical boundaries shared similar boundary positions to their normal hearing counterparts on all tone pairs. Among the three participant groups, while the NH and HA groups have more similar and stable boundary positions, the CI group experienced more varied boundary positions, and its T2T3 and T2T4 had much earlier boundary positions than the other two groups. That is, even for those who can categorically identify Mandarin tone pairs, the CI participants still perform worse than the NH and HA participants on Position.

(2) Slope

Figure 3.12 displays the three groups' means and standard errors (SEs) of curve slopes in all six tone pairs.

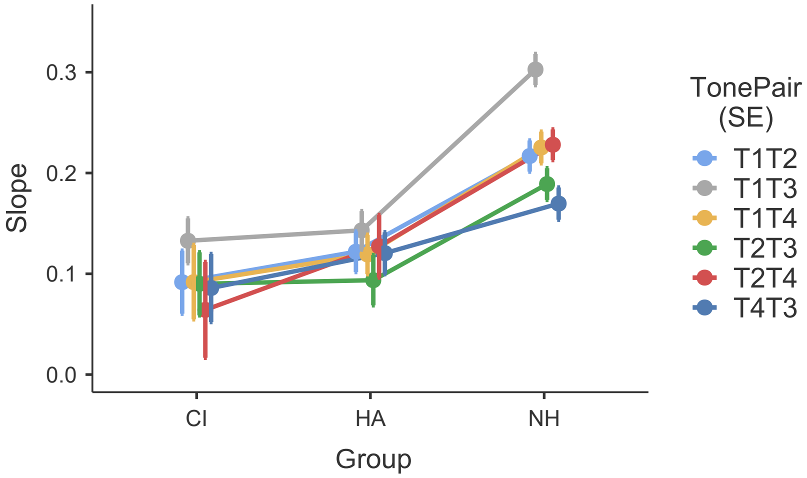


Figure 3.12: The means and standard errors of Slope

From Figure 3.12, one could note that the NH group's slopes are larger than that of the deaf groups, in which the NH group's T1T3 has the largest slope in all other cases.

Table 3.23 presents the ANOVA results of the final model for Slope, in which the main effects of Group and TonePair are significant, but no significance is found between the interaction of TonePair and Group.

Table 3.23: The ANOVA table of the final model for Slope

	F	Num df	Den df	*p*
Group	42.81	2	82.9	< 0.001
TonePair	4.78	5	255.7	< 0.001
Group × TonePair	1.31	10	251.2	0.223

Note: Satterthwaite method for degrees of freedom.

The estimates of the final model for Slope are displayed in Table 3.24. To save space, only the effects with significance or approaching significance are included in this table.

Table 3.24: The parameter estimates of the fixed effects for Slope

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	0.14519	0.00699	0.13149	0.1589	92.6	20.7773	< 0.001
Group2	NH - CI	0.12923	0.01782	0.09430	0.1642	97.7	7.2504	< 0.001
TonePair1	T1T3 - T1T2	0.04937	0.01665	0.01674	0.0820	252.4	2.9656	0.003

From Table 3.24, one could note that the main effect of the NH group is significant and positive, indicating that this group has much steeper slopes than the CI group. In contrast, no significance is reported between the HA and CI group, revealing that the two deaf groups have similar performance in curve slopes. Meanwhile, while the main effect of T1T3 is significant and positive, no significance is found in other tone pairs. That is, the slope of T1T3 is much steeper than the other tone pairs.[21]

Summarizing the above results, three points could be found about the impact of participant groups and tone pairs on Slope values:

Firstly, our three participant groups' performance were apparently different in the Slope values. That is, NH's Slope values were higher than the other two groups in all six tone pairs, which meant this group had better performance on tone categorical identification. In contrast, CI had the lowest Slope values in all tone pairs, and its Slope data were reporting significance in comparing with the NH group. Moreover, HA's Slope data located at between the two extremes, which indicated that this group's tone categorical identification was better than CI but worse than NH.

Secondly, tone pairs also impacted the categorical identification of Mandarin tones, but their impact does not contribute too much to the model. Some tone pair like T1T3 had higher Slope values than the other tone pairs, indicating that T1T3 was easier to categorically identify than the latter one. However, the Slope values, especially those of the deaf groups, were not changed too much among different tone pairs.

Finally, no interaction was found between participant group and tone pair, indicating that the three groups followed similar patterns in the difficulty orders in identifying the

---

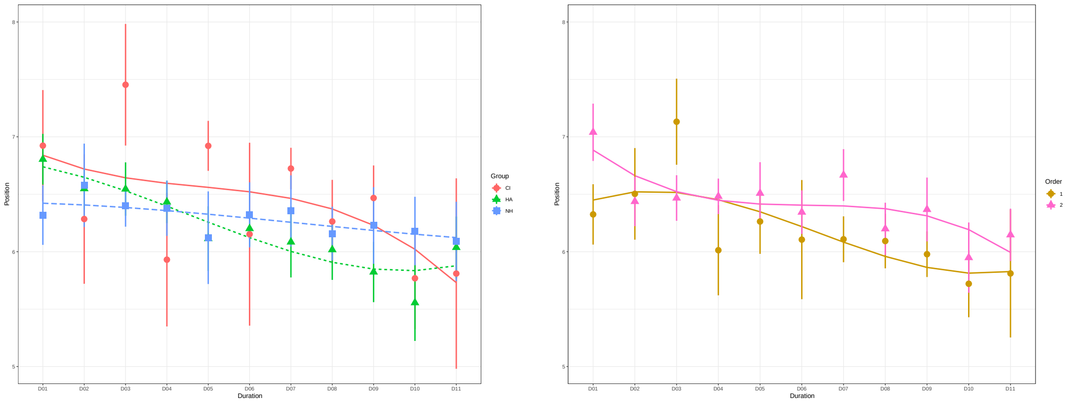
[21] A steeper slope means a clearer categorical boundary. In this study, the reason T1T3 has the steepest slope might be that the two tones have apparently different tone heights that make the participants easier to differentiate the two tones.

tone pairs. That is, no specific tone pair is exclusively and more difficult for a particular participant group.

3.2.2.2 The impact of the stimulus duration

(1) Position

Using Growth curve analysis, we tired to fit models to analyze the impact of Group and Order on the changes of Position values over the 11 duration steps. From linear to quartic order models, we failed to find any significant fixed effects. Figure 3.13 displayed the fitting results of cubic model, which indicated that no regular Position value changed over the 11 duration steps impacted by Group or Order. However, from this figure, we could see a general trend (non-sig) that longer durations are associated with right-ward shifts in tone boundaries.



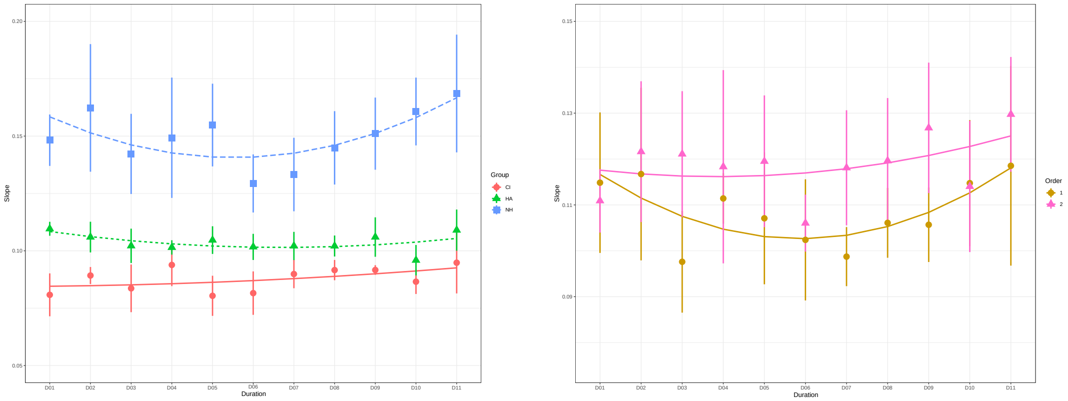
(a) Group (b) Order

Figure 3.13: The cubic fits for effect of Group and Order on Position

(2) Slope

Growth curve analysis was also applied to analyze Slope value changes over the 11 duration steps. From linear to quartic, the best fitting curves were modeled with second-order orthogonal polynomials and fixed effects of Group and Order on all duration steps.

In this model, CI group and Order1 were treated as the baseline to evaluate HA, NH and Order2 conditions. The model also included random effects of TonePair on all duration situations. The fixed effects of Group and Order were added individually and their effects on model fit were evaluated using model comparisons. Improvements in model fit were evaluated using –2 times the change in log-likelihood (–2 · Δ*LL*), which is distributed as χ^2 with degrees of freedom equal to the number of parameters (Mirman, 2014, pg.74). Figure 3.14 illustrated the observed data and growth curve model fits for effect of Group and Order on Slope.



(a) Group

(b) Order

Figure 3.14: The Quadartic fits for effect of Group and Order on Slope

Within this model, although the effect of Order on the intercept failed to report improvement ($\chi^2(1) = 0.715$, $p = 0.398$), the effect of Group on the intercept did improve model fit ($\chi^2(1) = 146.039$, $p < .001$). Moreover, the effect of Group on the quadratic term did improve model fit ($\chi^2(1) = 159.101$, $p < .001$), indicating that the three participant groups differed in the slope changes. Table 3.25 displayed the fixed effect parameter estimates and their standard errors along with *p*-values estimated using the normal approximation for the *t*-values.

Table 3.25: Parameter Estimates for Analysis of Effect of Group and Order on Slope

	Estimate	Std. Error	*t*	*p*
Intercept	0.0801	0.0089	8.957	< .001
Linear	0.0023	0.0139	0.163	0.871
Quadratic	-0.0064	0.0139	-0.458	0.647
GroupHA: Intercept	0.0200	0.0060	3.379	< .001
GroupNH: Intercept	0.0654	0.0060	11.039	< .001
Order2: Intercept	0.0152	0.0126	1.201	0.263
GroupHA: Linear	-0.0091	0.0197	-0.464	0.643
GroupNH: Linear	0.0068	0.0197	0.347	0.729
GroupHA: Quadratic	0.0149	0.0197	0.76	0.448
GroupNH: Quadratic	0.0559	0.0197	2.843	0.005
Order2: Linear	0.0123	0.0197	0.626	0.532
Order2: Quadratic	0.0164	0.0197	0.833	0.406
GroupHA:Order2	-0.0079	0.0084	-0.944	0.346
GroupNH:Order2	-0.0071	0.0084	-0.847	0.398
GroupHA:Order2: Linear	-0.0049	0.0278	-0.175	0.861
GroupNH:Order2: Linear	-0.0130	0.0278	-0.468	0.640
GroupHA:Order2: Quadratic	-0.0210	0.0278	-0.756	0.450
GroupNH:Order2: Quadratic	-0.0645	0.0278	-2.321	0.021

From Table 3.25, one can note that, among the three participant groups, only NH group, NH group × Quadratic, and Order × Quadratic reported significance. These results indicated that the two deaf groups did not display any sensibility to the changes of stimuli durations. In contrast, NH group was apparently impacted by duration changes in categorically identifying Mandarin tones. As displayed in Figure 3.14, normal hearing people have higher Slope values (N.B., better tone categorical identification) when the stimuli durations are at or approaching the anchoring sounds' durations (the prototypes), and have lower Slope values (N.B.: worse tone categorical identification) when the stimuli durations are at or near the intermediate states of the anchoring sounds' durations; and this pattern is more obvious under the Order2 condition (N.B.: Order2 means the duration is steadily increasing with equal time intervals from D01 to D11).

Taking together, the current model revealed that, although duration's impact was not as strong as frequency, it did have some impact on tone categorical identification. It is noted that the impact of duration was confined to normal hearing people but not deaf participants.

3.3 Discussion

Now, it is time to address the questions we advanced at the beginning of this chapter.

3.3.1 Can PDAs categorically perceive synthesized Mandarin tones?

The main question of the present chapter is whether PDAs can categorically perceive synthesized Mandarin tones. To answer this question, we need to check PDAs' performance in the tone identification and tone discrimination experiment separately, combine the results of the two experiments together, and set up some standards to judge these results before we can arrive at any sound conclusion.

3.3.1.1 The performance in tone identification experiment

The aim of Tone identification experiment is to test whether prelingually deaf adults can categorically identify the four tones in Standard Chinese, and how frequency and duration impact their identification results.

The experiment results revealed a mixed answer to the first question. Namely, while normal hearing people had no problem in categorically identifying Mandarin tones, our deaf participants experienced difficulties in these identification tasks. From the results of data sorting, we could see that almost all normal hearing participants had categorical identification for all six tone pairs. Meanwhile, deaf people had great difficulty in categorically identifying Mandarin tones, and the difficulty degrees varied among the six tone pairs. First, many deaf participants, especially the CI people, could not categorically identify the six tone pairs. From the results of data sorting, we could see both the two deaf groups reported high *NonC* ratios (the percentage of non-categorical identification perception). Specifically, CI people reported much higher *NonC* ratios than HA users in all pairs except for T1T3. In the meanwhile, the two deaf groups strikingly and uniformly reported that the *NonC* ratios of T1T3 were lowest (HA: 7 in 26, around 26.92%; CI: 5 in 20, around

25.00%), and the *NonC* ratios of T2T4 were the highest (HA: 19 in 26, around 73.08%; CI: 17 in 20, around 85.00%) among the six tone pairs. Second, the following statistic analysis revealed that they also had worse performance in tone categorical identification than their normal hearing counterparts. That is, for one thing, the numbers of participants in the two deaf groups were statistically smaller than the NH group; for another, for the participants who can categorically identify the tones, while the deaf participants had similar category boundaries with the normal hearing people, their performances were worse than their normal-hearing counterparts. In particular, CI people's performance were also worse than HA users. When comparing the tone pairs, all participant groups had the best and the worst performance on T1T3 and T2T4 respectively,[22] and the deaf participants had the smallest divergence with normal hearing people on these two pairs. Overall, these results displayed a similar pattern with the direct observation results that CI people had greater difficulty than HA people, but HA people also performed much worse than their normal hearing counterparts in categorically identifying Mandarin tones. In addition, even the participants had categorically identified these tone pairs, some pairs like T2T3 and T2T4 were seemingly harder to identify than other pairs such as T1T3.

As for the answer for the second question, namely how frequency and duration impact tone identification results, our statistic results of the categorical identification data indicated that frequency played more important impact than duration on tone categorical identifications. First, the current results clearly indicated that both the deaf participants and the normal hearing participants were highly sensitive to the frequency changes of the stimuli, which helped them categorically identify Mandarin tones and set up similar tone category boundaries for all six tone pairs even with the existence of the confounding factor–the duration changes of the stimuli. Second, the impact of duration varied among the participant groups. On one hand, duration changes did impact normal hearing participants' performance in tone categorical identification. On the other hand, the results of Growth curve analysis indicated that duration changes had no impact on the performance of the two deaf groups. The current result is quite different from the past studies that Chinese deaf children highlight duration but not only frequency to distinguish one tone from the others in Standard Chinese.

---

[22] We speculate the reason T1T3 is easier than all in of tone register is that they have the largest spaces in perception. These characteristics might make them easier to distinguish from each other. As for the case of T2T4, the middle stimuli of this tone pair are more similar to T1, which could be the possible reason that people feel confused and make more mistakes.

3.3.1.2 The performance in tone discrimination experiment

The aim of the discrimination experiment is to test whether prelingually deaf adults could categorically discriminate similar but different stimuli created from the six tone pairs in Standard Chinese. And if the answer for this question is yes, then we would address the second question of how frequency and duration impact their categorical discrimination of those tone pairs. Considering there was no clear and open definition of what a categorical tone discrimination is in the past, the current study also develop a set of criteria to decide tone categorical discrimination.

The current results, unfortunately, show that the majority of deaf participants and not too few NH participants fail to categorically discriminate the tone pairs.[23] As a result, the current results could only answer our first research question. As for the second question, we have to leave it for future research when enough categorical discrimination data have been collected. Overall, the answer for the first question is that the majority of deaf people could not categorically discriminate synthesized Mandarin tones. Compared with the results of identification experiment, we could see that the participants have performed much worse in tone categorical discrimination experiment, and deaf people's performance (especially the CI people's performance) is worse than that of normal hearing people.

In our opinion, these results reflect two facts.

First, the discrimination task is much harder than the identification task. In the identification task, there is only one synthesized stimulus the participants only need to process in one trial. For one identification trial, what the participants need to do is to decide whether the stimulus matches the prototype of a particular tone in their own conscious. In the discrimination task, there are two synthesized stimuli that are same or with subtle difference in on trial. For one discrimination trial, the participants need to compare the two stimuli and decide whether the two stimuli belong to one category or two different categories that involves the processing the prototypes of the two categories. For the participants, the discrimination task involves heavier work-load and more fine-grained acoustic information,

---

[23] One might suspect that our criteria for tone categorical discrimination were too stringent, which caused most participants failed to categorically discriminate the tone pairs. In our opinion, the criteria are quite undemanding. Ideally, a good categorical discrimination curve should have a peak that is the same as the identification boundary. However, after checking the current data, as well as the discrimination curves displayed in existing Mandarin tone perception studies, we only stipulate that the listeners should be insensitive inside a category and be sensitive between two boundaries. Thus, the results of past and current study might just reflect a fact that Mandarin tones might not be categorically perceived even by native speakers.

and thus is harder to process than the identification task.

Second, deaf people have more difficulty in processing synthesized stimuli with subtle difference. That is, confined by the limits of hearing device, the auditory system of deaf people could not hear speech signal as well as normal hearing people. Thus, even though the deaf participants in the current experiments have long history of spoken language usage and abundant experience with Mandarin tones, they still could not match their normal hearing counterparts in categorically discriminating synthesized Mandarin tones. This point could also be supported by the difference between the two deaf groups. Specifically, the CI people have more difficulty in tone categorical discrimination than the HA people. HA is a device to amplify sound that is mainly used by deaf people who have intact cochlea. Thus, HA people have no problem in processing fine-grained tone information in ideal situations. In contrast, CI is a device to stimulate the function of cochlea. Because of the small size of CI, it could not process tone information directly. Also limited by its small size, CI could include a few channels to process sound signal. That is, even under ideal situations, the information CI could process is still much sparser than HA. As a result, the tonal information CI could convey is also much sparser than HA. Moreover, the CI participants have worse hearing than their HA counterparts, which also contributes their worse performance in perceiving tonal information. Taken together, it would not be a surprise that CI people have much worse performance than HA people in categorical perception with synthesized tone stimuli. Interesting readers could read Section 6.2.2, where we would thoroughly discuss the impact of hearing devices on deaf people's performance in all three experiments including the current one.

3.3.1.3 The performance in tone categorical perception

Although the relation between identification task and discrimination is still under debate, the two tasks were often used simultaneously to determine whether two tones were categorically perceived or not (Gerrits and Schouten, 2004; Xu et al., 2006; Jiang et al., 2012). In particular, the identification task was used to ascertain the categorical boundary between a given tone pair, and the discrimination task was used to test whether people could show a greater sensitivity to the tone difference between adjacent stimuli at or near the categorical boundary than to similarly spaced stimuli within a given category (Gussenhoven and Van de ven, 2020). The basic hypothesis behind this is that, for a given tone pair, the results of the two tasks should share the same boundary: Only when the crossing point of the identification curves is overlapping with the peak of the discrimination curve, then one can say the participants did categorically perceive the tone pair.

Although the overlapping of the identification boundary and the discrimination peak is sufficient to decide a categorical perception, it must be based on two prerequisites: both the identification and discrimination should be categorical. Unfortunately, the past studies on tone categorical perception in Standard Chinese often did not mention the methods of data classifying (or data cleaning if more exact) that differentiate categorical data from non-categorical data. Without such information, we could not know whether these studies had adopted a procedure of data cleaning, what standards they had used to screen the data, and whether they had excluded non-categorical identification and discrimination data for further analysis. Because the identification boundary was mistakenly taken as the categorical boundary, the past studies often adopted the identification task alone to determine tone categorical perception (Jiang et al., 2012). Therefore, it is perhaps these past studies took it for granted that the participants could also categorically discriminate a given tone pair, or did not concern whether that tone pair could be categorically discriminated. As a result, the results of those studies would be doubtful because they had included too much non-categorical data (especially non-categorical discrimination data), or did not check the overlapping between the identification boundary and the peak of the discrimination curve.

With above consideration in mind, the current study set up the standards of categorical/non-categorical data both for the identification and discrimination data, and then combined their results together to decide whether the participants could categorically perceive Mandarin tones. And the results indicate that the majority of deaf participants and some normal hearing participants could not categorically perceive Mandarin tones. Thus, the results might reflect the fact that, just like the case of vowels (Ma et al., 2021), the perception of Mandarin tones is not categorical or quasi-categorical for native speakers of Mandarin Chinese.

More important, from the results of categorical perception, one might note that deaf people (especially the CI people) who have reconstructed hearing still experience tremendous difficulty in categorical perception of Mandarin tones many years after using Mandarin as their first spoken language. Among the three participant groups, NH people perform quite well in identifying all six tone pairs benefiting from their awareness of frequency and duration changes. Meanwhile, one might note that if the NH participants could discriminate some given tone pairs, their discrimination peaks are overlapping with the corresponding identification boundaries in most cases. Consequently, the NH participants perform much better in categorical perception of Mandarin tones except for T2T3 and T4T3. In contrast, the deaf participants face more challenges both in identification and discrimination tasks because of the disability in perceiving the details of frequency and duration changes. As a result, only a few of them could categorically perceive Mandarin tones. It is noted that the CI participants could barely perceive any of the tones categorically, which might indicate the limitation of cochlear implant as a device of processing tonal information.[24]

3.3.2 Is there any suggestion for future tone categorical perception studies?

From this study, we could arrive at two suggestions for future tone categorical perception studies.

First, data cleaning is a necessary procedure for tone categorical perception studies. From the current results, we could see that, no matter what the participants are–the normal hearing native speakers or deaf L2 learners of the Standard Chinese, the participant's tone perception is quite complicated: being able to categorically identify two tones in a given tone pair does not imply being able to categorically discriminate between the two tones; being able to categorically identify and discriminate a given tone pair simultaneously does not imply being able to categorically perceive the two tones (see Table 3.15). Thus, similar work in the future should also apply a procedure of data cleaning to ascertain whether the participants can identify or discriminate the targets categorically. Using data cleaning, researchers can check the overlap between the identification boundary and the discrimination peak for each participant to decide whether that participant can categorically perceive a given tone pair. With the categorical perception data, they can take further steps to analyze the characteristics of tone categorical perception by different participant groups.

Second, it is important to include all tone pairs in tone categorical perception studies. Until now, people often use only one tone pair like T1T2 or T1T4 to study categorically perception of Mandarin tone system (Wang et al., 2021a,b), no matter the involving participants were normal hearing adults (Xi et al., 2010; Wang et al., 2017b), children (Chen et al., 2017a), or special population such as congenital amusics (Huang et al., 2015a; Jiang et al., 2012) and deaf children (Zhang et al., 2020a,b). The logic behind was that, if the

---

[24] Interesting readers could read Section 6.2.2, where we would discuss the limits of the CI device and the consequence to to deaf people in depth.

participants can categorically perceive one given tone pair, they could also categorically perceive the other tone pairs. Based on the results of the current study that includes all six tone pairs, the participants might have different performance on different tone pairs. Taking normal hearing people for example, while some tone pairs such as T1T3 and T1T4 are easier to perceive categorically, some others like T2T3 and T4T3 are harder to perceive categorically even for normal hearing native speakers of Standard Chinese. Since the current results reveal that some Mandarin tones might be naturally harder to perceive even for native speakers with normal hearing, it would be not rigorous to study the categorical perception of the Mandarin tone system by only involving one particular tone pair, especially when involving special participants like deaf people.

3.4 Conclusion

Findings of the synthesized tone perception experiments confirmed that normal hearing people surpass deaf people in tone categorical perception of the Standard Chinese. In the deaf populations, HA adults performed much better than CI people who could barely perceive Mandarin tones categorically. Moreover, this study highlighted the importance of cleaning the data and including all possible tone pairs in tone categorical perception studies, because failing to do so might cause methodological deficits.

This study has its limitations. Firstly, limited by the data size, we could not adopt statistic analysis for the discrimination data after data cleaning. Secondly, considering the time consumed in the discrimination experiment, we did not involve same/same trials in the present study, which make it is impossible to measure the parameters such as d' or criterion C from the raw data (Macmillan and Creelman, 2005; Shao et al., 2019). Finally, because of the experiment design, the participants in the current study needed much longer time and experienced more difficulties to finish the experiment tasks when compared with the past studies. In this study, we tested all six tone pairs in Standard Chinese; in each tone pair, we set up 11 steps both in frequency and duration domain. As a result, it is not only that the participants need much longer time to process the big number of stimuli, but also that they need to tackle more factors (the duration of the stimuli, the involving tones of the stimuli) to fulfill the tasks.

To avoid the above limitations, future research could recruit more participants for each group, to see whether the categorical perception of Mandarin tone pairs could be found from a larger participant base. When enough clean data could be collected for all three participant groups, we would be able to thoroughly probe deaf people's characteristics of tone categorical perception and analyze the impact of hearing devices (HA, CI) on the categorical perception of Mandarin tones by using statistical methods. Furthermore, the following studies on deaf people could focus on those more difficult tone pairs such as T2T4, which might reveal more facts about the mechanism of deaf people's tone perception. Finally, although the current study failed to find any impact of duration on deaf people's tone categorical perception, we could not exclude the possibility of the impact of duration because we did find some impact on normal hearing people. Thus, studies in the future could use synthesized stimuli with fewer steps, which might help us to answer whether duration has any impact on deaf people's categorical perception of Mandarin tones in depth.

Chapter 4

The lexical tone perception experiment

The current chapter focus on the lexical tone perception experiment. Different from the previous chapter that investigated the impact of pure acoustic cues such as frequency and duration based on synthesized stimulus, this chapter involves in real human-produced monosyllables to check Mandarin tone perception by prelingually deaf adults, as well as the impact of different rhymes and tone pairs on their performance of Mandarin tone perception.

Specifically, the experiment of the current chapter tried to answer three particular questions. First, do the participant groups in the current task perform better than in the synthesized tone perception tasks? Second, does the complexity of syllabic rhyme impact the participants' Mandarin tone perception? Finally, are certain tone pairs harder to distinguish in the current experiment?

The organization of this chapter are as follows: to begin with, we will introduce the methodology of the experiment. Then, we will report the results of the experiment. After that, we will provide a general discussion before concluding this chapter.

4.1 Methodology

Similar to the experiment reported on in Chapter 3, all recruited participants participate in the current experiment. Considering the preparation of stimuli in the current experiment is not as complex as in the synthesized tone perception experiment, we will introduce the stimuli in this section but not in a separate independent section.

4.1.1 Materials

The word-list for making the target materials of the current experiment is displayed in Table 4.1. Observing Table 4.1, one might note that the 48 monosyllabic real words in Standard Chinese belong to three types of rhymes with different degrees of complexity (the simple rhyme V < the open rhyme VV < the nasal rhyme VN, "<" means "less complex"). The onsets of the syllables in Table 4.1 are mainly aspirated stops [pʰ], [tʰ], [kʰ], and the unaspirated [p] is used as the onset when no suitable syllable with aspirated stops could be found. Because the words in Table 4.1 are also used as the word list in the following tone production experiments, the arrangement of this word list should also consider the convenience for acoustic analysis. Considering the main aim of the current experiment, we will not discuss it for now, but shall return to this topic in the next chapter.

Table 4.1: The basis of making the materials for the lexical tone perception experiment

rhyme	T1	T2	T3	T4
e [ɤ]	kē [kʰɤ^{55}] *section*	ké [kʰɤ^{35}] *shell*	kě [kʰɤ^{214}] *thrist*	kè [kʰɤ^{51}] *special*
i [i]	tī [tʰi^{55}] *kick*	tí [tʰi^{35}] *lift*	tǐ [tʰi^{214}] *body*	tì [tʰi^{51}] *instead*
a [a]	tā [tʰa^{55}] *he*	pá [pʰa^{35}] *climb*	tǎ [tʰa^{214}] *tower*	tà [tʰa^{51}] *tread*
u [u]	tū [tʰu^{55}] *bare*	tú [tʰu^{35}] *map*	tǔ [tʰu^{214}] *soil*	tù [tʰu^{51}] *rabbit*
ai [ai]	tāi [tʰai^{55}] *tire*	tái [tʰai^{35}] *raise*	kǎi [kʰai^{214}] *model*	tài [tʰai^{51}] *too*
ei [ei]	bēi [pei^{55}] *cup*	péi [pʰei^{35}] *accompany*	běi [bei^{214}] *north*	pèi [pʰei^{51}] *match*
ao [au]	tāo [tʰau^{55}] *draw out*	táo [tʰau^{35}] *escape*	tǎo [tʰau^{214}] *beg*	tào [tʰau^{51}] *cover*
ou [əu]	tōu [tʰəu^{55}] *steal*	tóu [tʰəu^{35}] *head*	kǒu [kʰəu^{214}] *mouth*	tòu [tʰəu^{51}] *lucid*
an [an]	tān [tʰan^{55}] *beach*	tán [tʰan^{35}] *talk*	tǎn [tʰan^{214}] *blanket*	tàn [tʰan^{51}] *sigh*
en [ən]	pēn [pʰən^{55}] *puff*	pén [pʰən^{35}] *basin*	kěn [kʰən^{214}] *agree*	bèn [bən^{51}] *stupid*
ang [aŋ]	tāng [tʰaŋ^{55}] *soup*	táng [tʰaŋ^{35}] *sugar*	tǎng [tʰaŋ^{214}] *lie*	tàng [tʰaŋ^{51}] *hot*
eng [əŋ]	pēng [pʰəŋ^{55}] *cook*	péng [pʰəŋ^{35}] *friend*	pěng [pʰəŋ^{214}] *boost*	pèng [pʰəŋ^{51}] *touch*

Besides the targets displayed in 4.1, four types of distractors were included as follows:

- The tone distractor (TD), namely a syllable that has the same onset and rhyme as the target syllable but with a different lexical tone.
- The onset distractor (OD), namely a syllable that has the same tone and rhyme as the target syllable but with a different onset.
- The rhyme distractor (RD), namely a syllable that has the same onset and tone as the target syllable but with a different rhyme.
- The filling distractor (FD), namely a syllable that distinguishes from the target by tone and onset/rhyme simultaneously.

In this study, the TDs, ODs, and RDs had two functions: they could be used as distracting opinions in the target trials, and as targets in the filling trials[1]. Thus, besides the targets displayed in Table 4.1, we also provided the recordings of three TDs, one OD, and one RD for each and every target displayed in Table 4.1. It is noted that the target syllables only differentiate with their own ODs by the feature of [±aspiration]. Meanwhile, the simple-rhyme syllables and their RDs have different main vowels, the open-rhyme syllables and their RDs have different main vowels or secondary vowels, and the nasal-rhyme syllables and their RDs have different nasal codas [n] or [ŋ]. Because Table 4.1 could not provide the full list of distractors itself, we had to make up the word list to ensure all words in Table 4.1 and their TDs, ODs, and RDs were included as the stimuli in this experiment. After filling the missing materials, we had 12 rhymes * 4 tones * 3 stimulus types (target, TD, and RD) = 144 monosyllables in total as the recording materials. As for the FDs, since we only included one of them as a distracting opinion in each and every trial but not as a stimulus in any filling trial, we did not need their recordings in the current study.

With the 144 monosyllables, we embedded them in the carrying sentence "wǒ zài shuō X zhè gè zì" (*I am saying the character X*. X represents one of the embedded monosyllables) and invited a male native speaker of Standard Chinese to produce these sentences for three times in a sound-treated booth. Then, we extracted the soundtracks of these monosyllables and scales their amplitude peak to 75dB by Praat 6.1.53 (Boersma and Weenink, 2021). Finally, for each and every monosyllable, we picked the file with the best sound quality as the experiment stimulus.

4.1.2 Experiment procedures

The hardware of the present experiment consisted of a Lenova M428 desktop with a 23# LCD screen, and an EDIFIER R10U loudspeaker. The software was Mousetracker V2.84 (Freeman and Ambady, 2010; Freeman, 2018) running on Windows 10 operating system.

As a popular software package that allows researchers to record and analyze hand movements traveling toward potential responses on the screen via the x, y coordinates of the computer mouse, MouseTracker can track the precise characterizations of mouse trajectories' temporal and spatial dynamics, and thus opens up a single reaction time into

---

[1] The target trials are the trials using the syllables in Table 4.1 as targets, and the filling trials are the trials using TDs, ODs, and RDs as targets.

a continuous stream of rich cognitive output (Freeman, 2018). Up to now, considerable research using this software had been published in a variety of areas such as psychology (Chua and Freeman, 2022), cognitive science (Oh et al., 2021), and linguistics (Tomlinson and Ronderos, 2021).

With Mousetracker, a four-alternative forced-choice task (4AFC) was designed to investigate the participants' performance in perceiving Mandarin lexical tones under the circumstance of the three distractors. For a given target syllable, six trials are designed to check the participants' responses. In each trial, the participants have to choose one from the following four options: a target syllable, a TD, a OD/RD, and a FD. In this study, although all the 144 recorded monosyllables, 864 trials (144 * 6 tone pairs) were involved in the experiment, only the 288 target trials (48 monosyllables * 6 tone pairs) concerning the target syllables in Table 4.1 were included for the following analysis.

As an example, Table 4.2 displayed the six trials of kē. From Table 4.2, one might note that three TDs (ké, kě, and kè) are appearing two times, the OD gē and RD kā are appearing three times, and each of the six FDs (gé, ká, gě, kǎ, gè and kà) is appearing one time among the six trials.

Table 4.2: The six trials of the target kē

Condition	Target	TD	OD/RD	FD
T1T2	kē	ké	gē	gé
T1T2	kē	ké	kā	ká
T1T3	kē	kě	gē	gě
T1T3	kē	kě	kā	kǎ
T1T4	kē	kè	gē	gè
T1T4	kē	kè	kā	kà

Figure 4.1 illustrated the possible arrangements of the four options for a kē trial displayed on the screen.

From Figure 4.1, one might note there are eight possible arrangements in which the 'start' button is at the center and the four options are located at the four corners of the screen. In particular, the target kē could be located at any one of the four quadrants; to make it easier to normalize the trajectory data, the TD and OD/RD are always located at the adjacent quadrants of kē, and the FD is always located at its opposite quadrant. Overall, among the 864 trials, there are 108 trials for each of the eight types of arrangements.

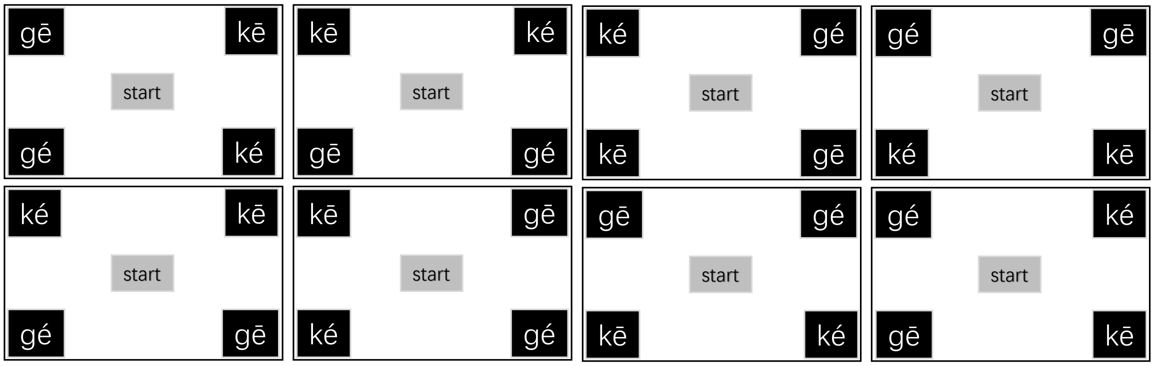


Figure 4.1: The possible arrangements of the four options in a given trial

The experimental procedure is illustrated by Figure 4.2.

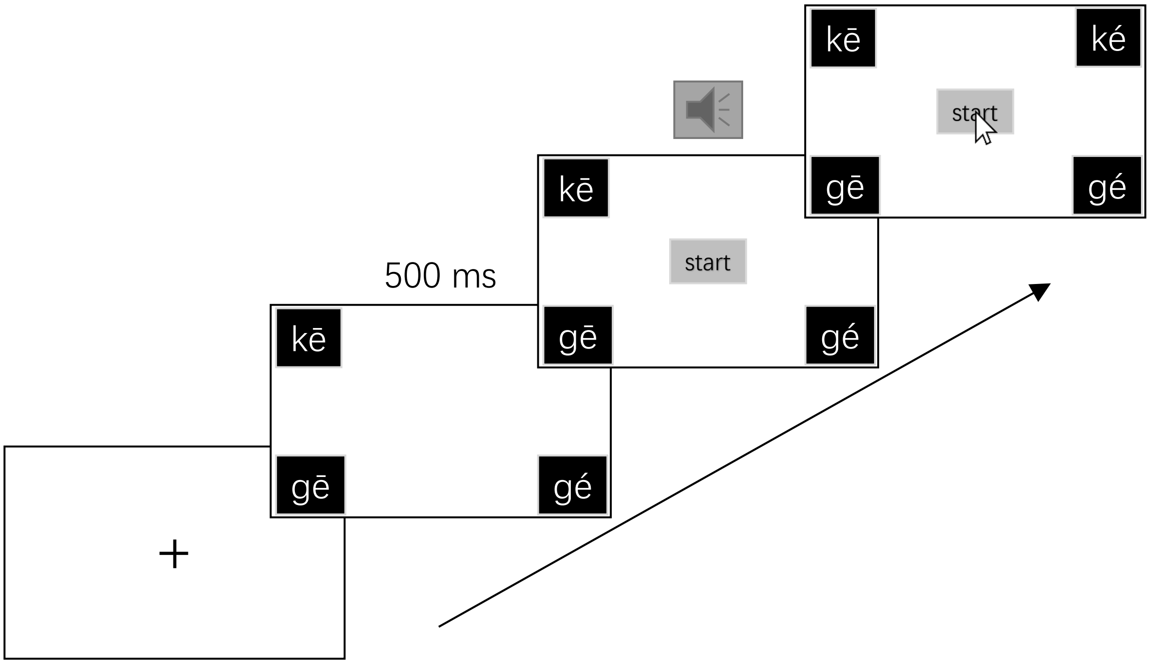


Figure 4.2: The sketch of the Mousetracker experiment for one trial

As sketched in Figure 4.2, the experimental procedure for one trial is as follows: a cross is promoted in the middle of the screen before the participants clicking the left button of the mouse to start the task. When the task starts, the four options but not the ’start’ button appear on the screen for 500 ms. Then, the ‘start’ button appears while the stimulus sound is played. At this period of time, the mouse is locked from moving or clicking. Immediately after the stimulus is played, the mouse arrow appears in the middle of the screen. By now, the participants can move the mouse to one of the corners and click the left button to choose what sound they have just heard. After the choice is made, a cross would be promoted on the screen again. The participants can decide to take a break or move on to the next trial. In this experiment, the participants are encouraged to make the decision precisely and quickly. If the participants failed to move the mouse in 1000 ms, the message "you are too slow, please make your decision a little quicker next time" will appear in the middle of the screen after the choice-making. If the participants do not make any decision in 6000 ms, the current trial would be terminated with an unfinished mark, and a cross mark for the next trial is promoted on the screen.

During the experiment, deaf participants were required to use their own hearing devices with comfortable settings. Before the main experiment, a practice procedure was provided for the participants to familiarize the procedures. In general, it would take 1.5 hours for a participant to finish the whole experiment including the practice, the main experiment, and the possible breaks.[2]

4.1.3 Data processing

(1) Data normalization

Besides the commonly used parameters like the choice and RT, Mousetracker also recorded the trajectory of mouse movement (MT). Since the destinations were usually set at different locations among the trials, and the distances of the MTs among the trials were also different, it would be impossible to compare MT data directly. To overcome this, Mousetracker provided two options 'Remap' and 'Time normalization' to normalize MT data, in which 'Remap' normalized MT data in the spatial domain by stipulating the same destination among the trajectories of all trials, and 'Time normalization' normalized MT data in the temporal domain by fitting the x, y coordinates of the trajectories into the same time-step using linear interpolation. In this study, we set the upper-left corner of quadrant 2 as the destination and used the 'Remap' function to adjust the trajectories of all trials[3]. Meanwhile, we adopted the 'Time normalization' function to fit the x, y coordinates of each trajectory into 101-time steps using linear interpolation, which time-normalized the

---

[2] Although some participants would feel tired after a long-time experiment, all participants' data are reliable because these data have reasonable correctness rates and response times.

[3] In fact, it is appropriate to re-map to any quadrant because the important thing is to adjust the trajectories of all trials to one corner. In this study, we have randomly chosen the upper-left corner of quadrant 2 as the target of data re-mapping.

MT data among all trajectories.

As an example, Figure 4.3 displays a CI participant's MT data in judging the trials of the simple rhyme V monosyllables in Table 4.2, in which the left and right figures display the MT data before and after remapping respectively.

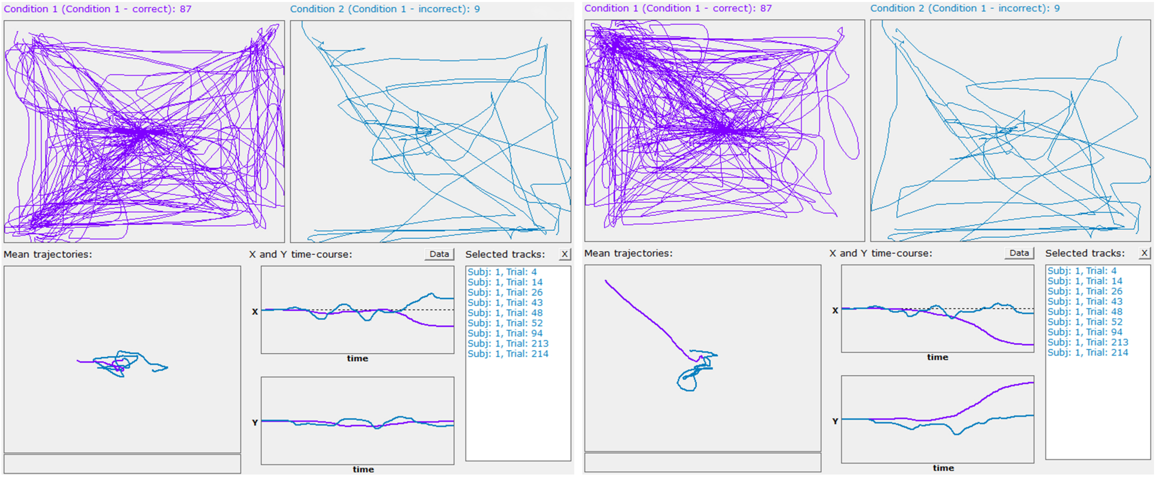


Figure 4.3: The results of remapping by Mousetracker: An example

From Figure 4.3, one can note that, among the 96 trials, 87 are chosen correctly (in purple) and 9 are chosen incorrectly (in blue). Before remapping the correctly chosen data, the destinations of their trajectories distribute at the four corners of the screen (the upper-left subplot of the left figure), and the mean trajectory does not target any particular corner (the bottom-left subplot of the left figure). After remapping the MT data, in contrast, the trajectories of all correct chosen trials share the same destination (the upper-left subplot of the right figure), and their mean trajectory clearly directs to the corner of quadrant 2 (the bottom-left subplot of the right figure). As for the incorrectly chosen trials, Figure 4.3 reveals that their trajectories (the upper-right subplots of the two figures) and mean trajectory (the bottom-left subplots of the two figures) are also changed, but no one and only destination is found for these trials after data remapping.

The example indicated that data remapping could spatially normalize the MT data of the correctly chosen trials but not the incorrectly chosen trials. Thus, this study would process the correctly and incorrectly chosen trials separately in the following analysis: For the correctly chosen trials, besides the choice results, this study would also analyze the RT and MT results; for the incorrectly chosen trials, this study would only analyze the choice results.

(2) Parameter preparation

In this experiment, two CI participants (CI_P6 and CI_P14) did not fulfill the experiment and were thus removed from the analysis[4]. With the remaining participants' data, the incorrectly chosen data of the CI, HA, and NH groups were 31.48% (1632 in 5184 trials among 18 participants), 32.73% (2451 in 7488 trials among 26 participants), and 1.67% (144 in 8640 trials among 30 participants) respectively.

With the incorrectly chosen data, this study analyzed them from two angles. The first angle concerned the distributions of different error types (ErrorType: TD, RD, OD, and FD) among three types of rhymes (RimeType: Simple, Open, and Nasal). Instead of using the ratio values of the four error types in each of the three rhyme types, this study calculated the rationalized arcsine units of the error types (RAU_Incorrect1) using Equation 4.1 (Studebaker, 1985) to make the data suitable for parametric statistical analysis.

$$RAU = 46.47324337[arcsin\sqrt{X/(N+1)} + arcsin\sqrt{(X+1)/(N+1)}] - 23 \tag{4.1}$$

In which the constant N is 288 trials / 3 rhyme types = 96, and the variable X is the number of a given participant's incorrect choices on a given rhyme type.

The other angle concerned the impact of tone pairs and the confusing directions inside each tone pair (TonePair) to the incorrect choices. In this study, we defined two confusing directions of tone pairs (ToneDir), namely ToneDir 1 (T1 to T2, T1 to T3, T1 to T4, T2 to T3, T2 to T4, and T3 to T4) and ToneDir 2 (T2 to T1, T3 to T1, T4 to T1, T3 to T2, T4 to T2, and T4 to T3). Again, this study also transformed the incorrectly chosen data into rationalized arcsine units (RAU_Incorrect2) with Equation 4.1, in which the constant N is 288 trials / 6 tone pairs / 2 directions = 24 here, and X is the number of a given participant's incorrectly chosen trials on each direction of each tone pair.

For both RAU_Incorrect1 and RAU_Incorrect2, the bigger their values are, the more incorrect choices the participants have made, and thus the worse they have performed in the experiment. It is noted that the distribution of the incorrectly chosen data was unbal-

---

[4] As we have said in Section 1.2.4.3 and Section 3.1.3.1, all participants performed the experiment in their own places at their own paces due to COVID-19. So, there are many possibilities some participants did not fulfill all the experiment tasks. In this study, we did receive the two participants' incomplete experiment data. However, we had removed their data from further analysis because there are limited trials left in their experiment data. It is noted that the excluded data would not cause more problems in analysis results because the linear mixed model has a much higher tolerance than traditional ANOVA methods.

anced among the groups, tone pairs, error types, and the participants. Thus, this study calculated the RAU_Incorrect1 and RAU_Incorrect2 values of certain types of errors when the participants did commit that types of errors.

Using Equation 4.1, the rationalized arcsine units (RAU_Correct) were also calculated for the correctly chosen data, where the constant N is 288 trials / 6 tone pairs / 3 rhyme types = 16, and X is the number of a given participant's correct choices on each tone pair under each rhyme type. In this study, the larger the RAU_Correct values are, the more correct choices the participants have made, indicating the better their performance in the experiment.

Besides RAU_Correct, this study extracted two time parameters: the initial time (InitTime) and reaction time (RT). InitTime is the amount of time that starts at the end of the stimulus sound's playing and ends at the beginning of the mouse's moving, reflecting the effort the participants have done to process the stimulus sound of a given trial. In contrast, the conventional parameter RT is the total time to fulfill that trial that reflects the general complexity of the experimental task.

Furthermore, we also extracted two MT parameters: the maximum deviation of trajectory (MD) and MD time (MDTime). MD is the largest perpendicular deviation between the actual trajectory and its idealized trajectory (a straight line between each trajectory's start and endpoints), and MDTime is the time of experiencing that deviation. For a given trial, theoretically speaking, the larger the MD is, the more the trajectory deviates toward the unselected alternatives, and the longer MDTime elapsed to overcome the trajectory deviation.

Among the five parameters for the correctly chosen data, RAU_Correct values reflected the ratios of correct choices under certain RhymeType and TonePair conditions, so they did not need further processing before statistics. With the remaining four parameters, we first cleaned the extreme values using the standard of mean±2sd for each participant. As a result, 5.47% (971 out of 17760), 4.41% (1625 out of 36820), and 4.53% (1957 out of 43200) cases were excluded for CI, HA, and NH groups respectively. After data cleaning, we calculated the average values of each parameter for all participants under all RhymeType and TonePair conditions. Taking the average values of these parameters and the RAU_Correct values together, we conducted a battery of statistical analyses to check the participants' performance on correctly chosen data.

(3) Statistic analysis

Considering the imbalanced distributions of the data among the participant groups, this study adopted a battery of linear mixed models to analyze the incorrectly chosen and correctly chosen data respectively. The same as to Section 3.1.4.3, we also adopted the Linear mixed-effect models (Gałecki and Burzykowski, 2013) using *lme4* package (Bates et al., 2015) and *lmerTest* package (Kuznetsova et al., 2017) to check the impact of these factors. The final models of the incorrectly chosen data were as follows:

- The final model of RAU_Incorrect1 was RAU_Incorrect1 ~ RhymeType $*$ ErrorType + Group + (1 + ErrorType | Subject), in which Group, RhymeType and ErrorType were the independent factors, 1 was the random intercept, ErrorType was the random slope, and Subject was the random factor. It is noted that the simple coding method was used for all three independent factors in this final model, in which CI, nasal, and FD were set as the baselines for Group, RhymeType, and ErrorType respectively.
- The final model of RAU_Incorrect2 was RAU_Incorrect2 ~ Group + TonePair + ToneDir + (1 + ToneDir | Subject), in which Group, TonePair, and ToneDir were independent factors, 1 was the random intercept, ToneDir was the random slope, and Subject was the random factor. It is noted that the simple coding method was used for all three independent factors in this final model, in which CI, T1T2, and ToneDir 1 were set as the baselines for Group, TonePair, and ToneDir respectively.

The final models of the correctly chosen data were as follows:

- Using RAU_Correct as the dependent variable, the final model was RAU_Correct ~ Group $*$ RhymeType + TonePair + (1 | Subject), in which Group, RhymeType, and TonePair were the independent factors, 1 was the random intercept, and Subject was the random factor. It is noted that the simple coding method was used for all three independent factors in this final model, in which CI, nasal, and T1T2 were set as the baselines for Group, RhymeType, and TonePair respectively.
- Using InitTime as the dependent variable, the final model was InitTime ~ Group $*$ RhymeType + (1 | Subject), in which Group and RhymeType were the independent factors, 1 was the random intercept, and Subject was the random factor. It is noted that the simple coding method was used for the independent factors in this final

model, in which CI and nasal were set as the baselines for Group and RhymeType respectively. Moreover, TonePair was not included in this final model because it failed to report any significance or the final model was not converged when containing this factor.

- Using RT as the dependent variable, the final model was RT ~ Group $*$ RhymeType + Group $*$ TonePair + InitTime + (1 | Subject), in which Group, RhymeType, and TonePair were the independent factors, 1 was the random intercept, and Subject was the random factor. For a given trial, since the lengthening InitTime leads to the longer RT but not vice versa, we included InitTime as the covariant in this final model. It is noted that the simple coding method was used for all three independent factors in this final model, in which CI, nasal, and T1T2 were set as the baselines for Group, RhymeType, and TonePair respectively.

- Using MD as the dependent variable, the final model was MD ~ Group + TonePair + RhymeType + (1 + RhymeType | Subject), in which Group, RhymeType, and TonePair were the independent factors, 1 was the random intercept, RhymeType was the random slope, and Subject was the random factor. It is noted that the simple coding method was used for all three independent factors in this final model, in which CI, nasal, and T1T2 were set as the baselines for Group, RhymeType, and TonePair respectively.

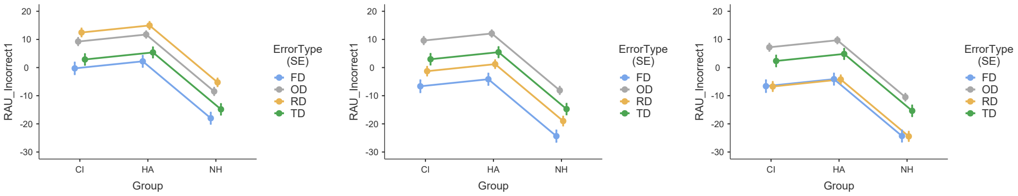
- Using MDTime as the dependent variable, the final model was MDTime ~ Group $*$ RhymeType + Group $*$ TonePair + MD + (1 | Subject), in which Group, RhymeType, and TonePair were the independent factors, 1 was the random intercept, and Subject was the random factor. Considering the MD time of a given trial could be impacted by its MD but not vice versa, we included MD as the covariant in this final model. It is noted that the simple coding method was used for all three independent factors in this final model, in which CI, nasal, and T1T2 were set as the baselines for Group, RhymeType, and TonePair respectively.

4.2 Results

4.2.1 Incorrectly chosen data

(1) RAU_Incorrect1

Figure 4.4 displays the means and standard errors (SEs) of RAU_Incorrect1 values. In Figure 4.4, while the patterns of RAU_Incorrect1 values among the participant groups are remaining constant (HA > CI > NH), they are varied among the four error types under three rhyme types.



(a) nasal (b) open (c) simple

Figure 4.4: The means and standard errors of RAU_Incorrect1 accross rhyme types

Table 4.3 presents the ANOVA results of the final model for RAU_Incorrect1, in which the main effects of Group, RhymeType, and ErrorType, as well as the interaction of RhymeType and ErrorType are significant.

Table 4.3: The ANOVA table of the final model for RAU_Incorrect1

	F	Num df	Den df	*p*
Group	107.4	2	73.1	< 0.001
RhymeType	62.7	2	332.8	< 0.001
ErrorType	34.7	3	54.3	< 0.001
RhymeType × ErrorType	26.2	6	327.4	< 0.001

Note: Satterthwaite method for degrees of freedom.

In the final model, the F value of Group is larger than RhymeType, and the F value of RhymeType is larger than the F value of ErrorType; revealing that Group contributes more than RhymeType, and RhymeType contributes more than ErrorType to the model. Moreover, the interaction of RhymeType and ErrorType indicates more certain types of incorrect choices in certain types of rhymes are made than other situations. The following check of the simple effects reports that, except for the ErrorType TD of RhymeType nasal and ErrorType RD of RhymeType simple, significant simple effects are found in all remaining conditions.

The estimates of the final model for RAU_Incorrect1 are displayed in Table 4.4, in which all effects are significant except for the case of the HA group. The estimate of the HA group is positive and insignificant, revealing HA people had committed a few more mistakes than CI people but the difference between the two groups was not big enough to report any significance; in contrast, the effect of the NH group was negative and highly significant, indicating NH people had made much lesser incorrect choices than CI people in the task. Moreover, both RhymeType simple and RhymeType open were negative and highly significant, revealing that the participants had made more mistakes in processing the trials with nasal rhymes. Furthermore, the three error types OD, RD, and TD are positive and highly significant, confirming that the participants had distracted much lesser by FDs than the other three types of distractors in the experiment. As for the interactions between RhymeType and ErrorType, while the interactions of RhymeType and RD are highly negative, all other effects were negative and significant. That is, the participants had distracted much lesser by RDs, but more by ODs and TDs in processing the target syllables with open or simple rhymes.

Table 4.4: The parameter estimates of the fixed effects for RAU_Incorrect1

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	-2.97	0.898	-4.734	-1.21	63.3	-3.31	0.002
Group1	HA - CI	2.50	1.432	-0.307	5.31	58.1	1.75	0.086
Group2	NH - CI	-17.72	1.571	-20.795	-14.64	80.8	-11.28	< 0.001
RhymeType1	open - nasal	-4.92	0.651	-6.194	-3.64	335.1	-7.55	< 0.001
RhymeType2	simple - nasal	-7.02	0.649	-8.294	-5.75	333.1	-10.81	< 0.001
ErrorType1	OD - FD	13.20	1.517	10.224	16.17	51.8	8.70	< 0.001
ErrorType2	RD - FD	5.99	1.467	3.119	8.87	58.1	4.09	< 0.001
ErrorType3	TD - FD	7.24	1.211	4.868	9.61	47.7	5.98	< 0.001
RhymeType1 × ErrorType1	open - nasal × OD - FD	6.72	1.909	2.980	10.46	328.5	3.52	< 0.001
RhymeType2 × ErrorType1	simple - nasal × OD - FD	4.29	1.860	0.644	7.94	326.5	2.31	0.022
RhymeType1 × ErrorType2	open - nasal × RD - FD	-7.39	2.023	-11.350	-3.42	339.4	-3.65	< 0.001
RhymeType2 × ErrorType2	simple - nasal × RD - FD	-12.87	2.043	-16.874	-8.86	338.1	-6.30	< 0.001
RhymeType1 × ErrorType3	open - nasal × TD - FD	6.45	2.048	2.434	10.46	327.8	3.15	0.002
RhymeType2 × ErrorType3	simple - nasal × TD - FD	5.79	1.984	1.903	9.68	320.9	2.92	0.004

(2) RAU_Incorrect2

Table 4.5 lists the confusion matrices of the three participant groups' lexical tone identification. From Table 4.5, one should see that the two deaf groups performed much worse than the NH group in identifying Mandarin tones. Inside the two deaf groups, although the CI group generally performed worse than the HA group, both of them similarly committed most errors in T2 trials and the lest errors in T4 trails.

Table 4.5: The confusion matrices of the three groups' lexical tone identification

	CI					HA					NH			
	T1	T2	T3	T4		T1	T2	T3	T4		T1	T2	T3	T4
T1	52.08%	20.37%	10.42%	17.13%	T1	70.47%	10.11%	7.34%	12.07%	T1	98.61%	0.83%	0.00%	0.56%
T2	18.29%	46.06%	21.06%	14.58%	T2	14.85%	53.18%	19.25%	12.72%	T2	0.97%	98.61%	0.28%	0.14%
T3	10.19%	15.05%	66.67%	8.10%	T3	7.18%	20.07%	63.13%	9.62%	T3	0.28%	0.28%	99.17%	0.28%
T4	9.26%	10.88%	8.80%	71.06%	T4	6.53%	10.93%	6.36%	76.18%	T4	0.28%	0.28%	0.28%	99.17%

The ANOVA results of the final model for RAU_Incorrect2 reported no significant main effect on ToneDir [$F(1, 50.1) = 0.864$, $p = 0.357$], indicating that no significant difference was found between the two confusion directions. In contrast, the main effects on Group [$F(2, 66.2) = 3.636$, $p = 0.032$] and TonePair [$F(5, 170) = 7.513$, $p < 0.001$] were significant in this final model. The following Post Hoc tests indicated that, among the three participant groups, the HA group was significantly different from the NH group but not from the CI group, and the CI group was different from NH with marginal significance. As for the cases of TonePair, only four were found significant among the 15 comparisons. In detail, T1T2 was significantly different from T3T4 ($t = 3.757$, $p = 0.0029$), and T2T3 was significantly different from T1T3 ($t = -3.964$, $p = 0.0013$), T1T4 ($t = -4.135$, $p = 0.0007$), and T3T4 ($t = 4.971$, $p < 0.001$). That is, compared with the easier tone pairs T1T3, T1T4, and T3T4, T2T3 is the hardest to tell the difference for the participants, and T1T2 and T2T4 were somewhere between the two extremes. It is reminded that, in the synthesized tone experiment, T2T4 and T2T3 were the first and second most difficult tone pairs for the deaf groups, and T2T3 was the most difficult tone pair for the NH group. Thus, the current results were generally consistent with the results of the synthesized tone experiment that T2T3 and T2T4 are hard to categorically perceive for prelingually deaf people.

The estimates of the final model for RAU_Incorrect2 are displayed in Table 4.6. Again, the estimates of the groups reported that the NH group had a negative and highly significant effect, indicating that NH people had made much fewer mistakes than CI people; in contrast, the HA group showed a positive but not significant effect, showing that they had made a little more mistakes than their CI counterparts. Meanwhile, ToneDir 2 decreased RAU_Incorrect2 values but showed no significance with the baseline ToneDir 1. Namely, in a particular tone pair such as T2T3, the number T2 mistaken for T3 was not different from the number of T3 mistaken for T2. Last but the most important, except for the cases of T2T3 and T2T4, all other tone pairs reported negative and significant effects on RAU_Incorrect2, indicating the participants had committed lesser mistakes in processing T1T3, T1T4, and especially T3T4; in contrast, the participants had made more wrong decisions on T2T4, T1T2, and especially T2T3 trials.

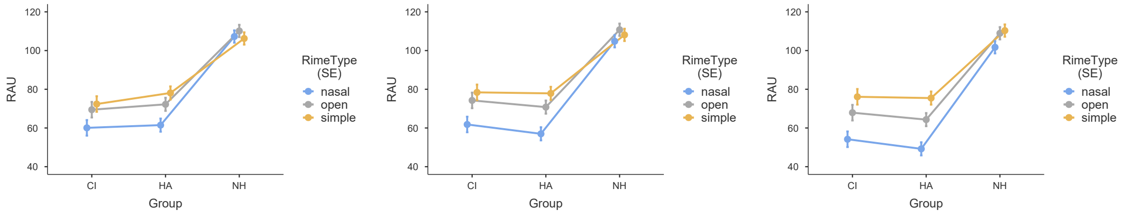
Table 4.6: The parameter estimates of the fixed effects for RAU_Incorrect2

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	10.586	2.10	6.46	14.71	69.9	5.0333	< 0.001
Group1	HA - CI	0.417	4.31	-8.02	8.86	52.5	0.0968	0.923
Group2	NH - CI	-13.053	5.52	-23.87	-2.24	76.5	-2.3652	0.021
TonePair1	T1T3 - T1T2	-6.991	2.51	-11.90	-2.08	258.6	-2.7903	0.006
TonePair2	T1T4 - T1T2	-6.813	2.33	-11.38	-2.25	264.9	-2.9237	0.004
TonePair3	T2T3 - T1T2	2.884	2.20	-1.43	7.20	264.8	1.3104	0.191
TonePair4	T2T4 - T1T2	-2.886	2.37	-7.54	1.76	261.0	-1.2164	0.225
TonePair5	T3T4 - T1T2	-9.363	2.47	-14.20	-4.53	263.2	-3.7948	< 0.001
ToneDir1	2 - 1	-1.754	1.89	-5.45	1.95	50.1	-0.9293	0.357

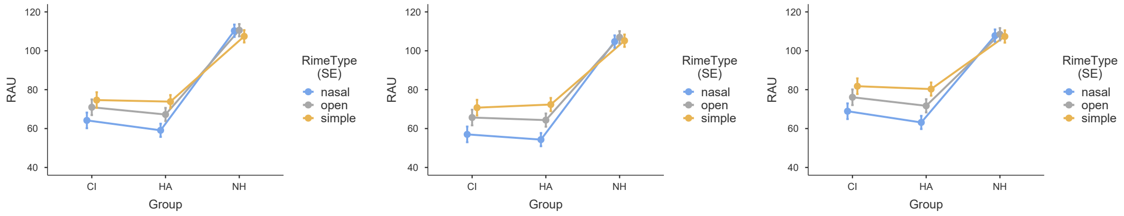
4.2.2 Correctly chosen data

(1) RAU_Correct

Figure 4.5 displays the means and SEs of RAU_Correct values. Comparing the six subplots of Figure 4.5, although the three lines had some differences in distribution, they shared similar patterns in general: from the X direction, the NH group had higher RAU_Correct values than the CI and HA groups; from the Y direction, the RAU_Correct values of the syllables with simple rhymes were generally higher than that of open and nasal rhymes for CI and HA group, but no clear pattern was found NH group. The patterns illustrated in Figure 4.5 were also confirmed by the ANOVA results of the final model for RAU_Correct. In this final model, there were significant effects of all three factors and the interactions between every two factors ($p < 0.001$). Since all the following simple main effects were significant, this study also applied post hoc tests for the three independent factors. The multiple comparisons of RhymeType in different participant groups showed that, except for the cases of nasal - simple and open - simple in the NH group, significant differences were found in all other conditions. The multiple comparisons of Group among the rhyme types reported that, while no difference was found between the CI group and the HA group, all the differences between the NH group and the two deaf groups were significant in all three rhyme types. These results indicated that, compared with the NH group, the two deaf groups had similarly worse performances because of the impact of the rhyme types. Moreover, the post hoc results of TonePair inside each participant group revealed that, while no significance was found among the tone pairs in the NH group, the two deaf groups apparently made more mistakes on T2T3 and T2T4, and fewer mistakes on T3T4.



(a) T1T2 (b) T1T3 (c) T1T4



(d) T2T3 (e) T2T4 (f) T3T4

Figure 4.5: The means and standard errors of RAU_Correct

The estimates of the final model for RAU_Correct are displayed in Table 4.7. To save space, only the effects with significance were included in this table. From Table 4.7, one could see the following significant main effects. First, agreeing with the results of RAU_Incorrect1 and RAU_Incorrect2, the RAU_Correct of the NH group is positive and highly significant, indicating NH people had made much more correct choices in the experiment than the baseline CI people. Moreover, three tone pairs showed significance when compared with the baseline T1T2: the estimate of T3T4 was positively significant, and the estimates of T1T4 and T2T4 were negatively significant. These results, in concordance with the results of RAU_Incorrect2, generally indicated that T3T4 is the easiest tone pair for the participants. Finally, both open and simple rhymes were reported positive and highly significant, confirming again that the participants had experienced much more difficulties in processing the target syllables with nasal rhymes.

The significant interactions between every two factors were as follows. First, between

Table 4.7: The parameter estimates of the fixed effects for RAU_Correct

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	81.379	1.799	77.853	84.904	71.0	45.242	< 0.001
Group2	NH - CI	38.476	4.508	29.640	47.312	71.0	8.535	< 0.001
TonePair2	T1T4 - T1T2	-3.240	1.059	-5.315	-1.164	1227.0	-3.060	0.002
TonePair4	T2T4 - T1T2	-3.997	1.059	-6.072	-1.922	1227.0	-3.775	< 0.001
TonePair5	T3T4 - T1T2	3.145	1.059	1.070	5.220	1227.0	2.971	0.003
RhymeType1	open - nasal	7.984	0.749	6.516	9.451	1227.0	10.663	< 0.001
RhymeType2	simple - nasal	11.646	0.749	10.178	13.113	1227.0	15.555	< 0.001
Group1 × TonePair1	HA - CI × T1T3 - T1T2	-6.233	2.729	-11.582	-0.884	1227.0	-2.284	0.023
Group1 × TonePair2	HA - CI × T1T4 - T1T2	-6.367	2.729	-11.716	-1.018	1227.0	-2.333	0.020
Group1 × TonePair3	HA - CI × T2T3 - T1T2	-6.506	2.729	-11.855	-1.157	1227.0	-2.384	0.017
Group1 × TonePair5	HA - CI × T3T4 - T1T2	-7.184	2.729	-12.533	-1.835	1227.0	-2.632	0.009
Group2 × TonePair5	NH - CI × T3T4 - T1T2	-8.289	2.654	-13.490	-3.088	1227.0	-3.123	0.002
Group2 × RhymeType1	NH - CI × open - nasal	-6.487	1.876	-10.165	-2.810	1227.0	-3.457	< 0.001
Group1 × RhymeType2	HA - CI × simple - nasal	4.303	1.930	0.521	8.086	1227.0	2.230	0.026
Group2 × RhymeType2	NH - CI × simple - nasal	-13.310	1.876	-16.988	-9.632	1227.0	-7.093	< 0.001
TonePair2 × RhymeType2	T1T4 - T1T2 × simple - nasal	9.620	2.534	4.653	14.588	1227.0	3.796	< 0.001

Group and TonePair, the effect estimates of the HA and NH group's T3T4-T1T2, as well as the HA group's T1T4-T1T2 and T2T4-T1T2 were significantly lower than these differences in the CI group, indicating that CI people's accuracies had experienced much more variation across tone pairs than that HA and NH people. Moreover, the Group and RhymeType interactions were negative and highly significant both in simple and open rhymes for the NH group, but only positive and significant in simple rhymes for the HA group. Namely, NH people's accuracies had smaller differences among the three rhyme types, and HA people's accuracies experienced larger differences between simple and nasal rhymes compared with the CI group. Finally, there was only one significant interaction between TonePair and RhymeType, revealing that the participants' accuracies among the tone pairs were not affected by the rhyme types or vice versa.

(2) The initial time (InitTime)

Figure 4.6 displays the means and SEs of InitTime values as a function of group and rhyme type. From Figure 4.6, one could note that the initial times of the CI group were always longer than the HA group, and the HA group were longer than the NH group in all three rhyme types. Meanwhile, the initial times of simple rhymes were always longer than that of open rhymes, and open rhymes were longer than nasal rhymes.[5]

---

[5] One might specuate that the participants use much longer initial time to process simple rhyme tokens is that they were waiting to see if something else was coming after the vowel. But we can't say this because simple rhyme has the same duration as a nasal rhyme or an open rhyme, even though there is only one vowel in the simple rhyme.

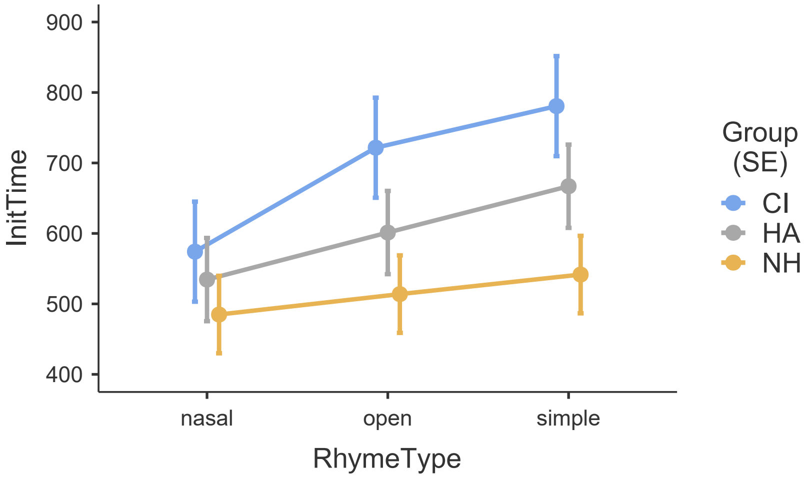


Figure 4.6: The means and standard errors of InitTime

The ANOVA results of the final model for InitTime indicated that, although Group failed to report any significance [$F(2, 71.0) = 2.08$, $p = 0.133$], there were significnat effects of RhymeType [$F(2, 1251.0) = 81.24$, $p < 0.001$] and the interaction between Group and RhymeType was significant [$F(4, 1251.0) = 9.53$, $p < 0.001$].

The estimates of the final model for InitTime are displayed in Table 4.8. From Table 4.8, one could see that, compared with the negative and insignificant effect of the HA group, the effect of the NH group was negative and significant, implying both HA and NH people had spent shorter initial times in processing the experiment trials, but only the initial time shortening of NH people was meaningful in statistics. Meanwhile, both open and simple rhyme reported positive and highly significant effects, indicating that the participants used much longer initial times in processing tokens with open and simple rhyme. Finally, all interactions were negative and significant, showing that HA and NH people had experienced less changes than CI people in InitTime when processing the syllables with different types of rhymes.

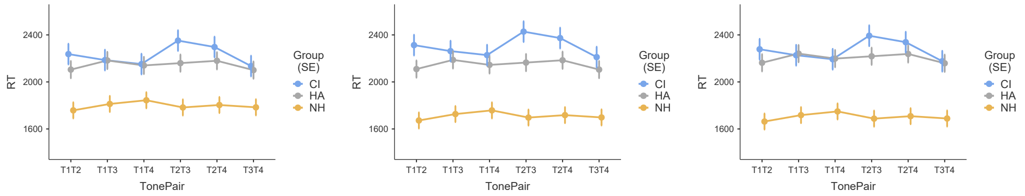
(3) The reaction time (RT)

Figure 4.7 displays the means and standard SEs of RT values as a function of group and rhyme type.

In Figure 4.7, the NH group's RTs were always shorter than that of the HA and CI groups in all tone pairs and rhyme types. As for the differences between HA and CI people,

Table 4.8: The parameter estimates of the fixed effects for InitTime

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	602.2	35.3	532.9	671.39	71.0	17.05	< 0.001
Group1	HA - CI	-91.2	91.0	-269.7	87.21	71.0	-1.00	0.320
Group2	NH - CI	-178.7	88.5	-352.2	-5.19	71.0	-2.02	0.047
RhymeType1	open - nasal	81.1	10.4	60.6	101.55	1251.0	7.76	< 0.001
RhymeType2	simple - nasal	131.9	10.4	111.5	152.37	1251.0	12.64	< 0.001
Group1 × RhymeType1	HA - CI × open - nasal	-80.8	26.9	-133.5	-27.98	1251.0	-3.00	0.003
Group2 × RhymeType1	NH - CI × open - nasal	-118.5	26.2	-169.8	-67.25	1251.0	-4.53	< 0.001
Group1 × RhymeType2	HA - CI × simple - nasal	-74.1	26.9	-126.8	-21.36	1251.0	-2.75	0.006
Group2 × RhymeType2	NH - CI × simple - nasal	-149.8	26.2	-201.1	-98.55	1251.0	-5.73	< 0.001



(a) nasal (b) open (c) simple

Figure 4.7: The means and standard errors of RT

while the HA group had shorter RTs in general, it seemed that the RT differences between the two groups were smaller in some cases. It is noted that, under the simple rhyme condition, HA people's RTs of T1T3 and T1T4 were even a little longer than CI people. Another interesting characteristic displayed in Figure 4.7 is the RT variations among the tone pairs. Compared with the other two groups, CI people had experienced more variations among the six tone pairs in all three rhyme types. Moreover, CI people showed a pattern with longer RTs on T2T3 and shorter RTs on T1T4. In contrast, it seemed that HA and NH people generally displayed a mirror pattern with CI people. That is, HA and NH people had shorter RTs on T2T3 and longer RTs on T1T4 almost in all rhyme types.

Table 4.9 presents the ANOVA results of the final model for RT. Except for the main effect of RhymeType, all main effects and interactions in the model were significant. In particular, the covariate InitTime reported a very large F value, implying it contributed a lot to the final model. The following simple effect test of Group on TonePair reported that the differences between NH and CI/HA were significant on all tone pairs, and the differences between HA and CI were not significant on all tone pairs. The simple effect test of RhymeType on Group were as follows: In CI group, open and simple rhymes were not significant different, but both of them were significant longer than nasal rhymes; In HA group, nasal and open rhymes were boundary significant, and both of them were significant different with simple rhymes; In NH group, there were no significant differences between each two of the three rhyme types.

Table 4.9: The ANOVA table of the final model for RT

	F	Num df	Den df	*p*
Group	18.9680	2	71.0	< 0.001
RhymeType	0.0131	2	1250.6	0.987
TonePair	3.1970	5	1234.0	0.007
InitTime	417.9209	1	1113.3	< 0.001
Group × TonePair	2.2051	10	1233.9	0.015
Group × RhymeType	4.6831	4	1236.1	< 0.001

Note: Satterthwaite method for degrees of freedom.

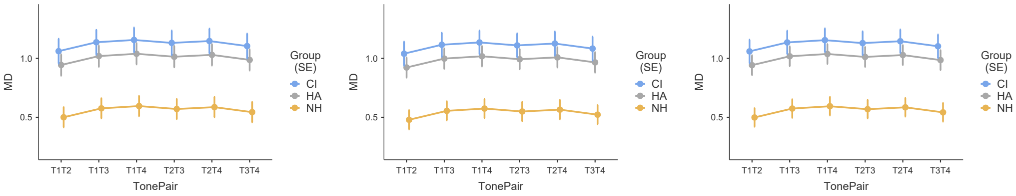
The estimates of the final model for RT are displayed in Table 4.10. To save space, only the effects with significance were included in this table. In Table 4.10, the effect of the NH group was negative and highly significant, indicating NH people used much shorter times to process the experiment trials than CI people. Moreover, both T2T3 and T2T4 reported positive and significant effects, implying the participants had consumed more time in processing the T2T3 and T2T4 trials in the experiment. This result, agreeing with the deaf groups' error pattern that they made more mistakes in these two tone pairs, might reveal that T2T3 and T2T4 are the hardest tone pairs for deaf people. Thirdly, the effect of the covariate InitTime was positive and highly significant, confirming that, if the participants needed longer initial times, they did use longer RT in processing the trials. Fourthly, the interaction between NH and T1T4 was positive and significant, showing that NH had experienced more variances on RT from T1T2 to T1T4 than CI people. Finally, the interaction between NH and open rhyme, as well as the interaction between NH and simple rhyme were negative and significant, revealing that NH had experienced lesser variances on RT among the three rhyme types than CI people.

(4) The maximum deviation of trajectory (MD)

Figure 4.8 displays the means and standard SEs of MD values. Observing Figure 4.8, one could note that the NH group's MD values were apparently lower than the other two groups, and the HA group's MD values were a little lower than the CI group's values in all three subplots. In contrast, although the MD values were also different among the six tone pairs, the variations in each participant group were quite small.

Table 4.10: The parameter estimates of the fixed effects for RT

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	2056.07	38.6474	1980.320	2131.81	70.5	53.2007	< 0.001
Group2	NH - CI	-528.27	97.2528	-718.886	-337.66	71.5	-5.4320	< 0.001
TonePair3	T2T3 - T1T2	64.91	28.0141	10.008	119.82	1234.4	2.3172	0.021
TonePair4	T2T4 - T1T2	60.21	28.0262	5.283	115.14	1234.0	2.1485	0.032
InitTime	InitTime	1.02	0.0499	0.922	1.12	1113.3	20.4431	< 0.001
Group2 × TonePair2	NH - CI × T1T4 - T1T2	170.87	70.1627	33.356	308.39	1233.9	2.4354	0.015
Group2 × RhymeType1	NH - CI × open - nasal	-162.30	49.9541	-260.207	-64.39	1238.8	-3.2490	0.001
Group2 × RhymeType2	NH - CI × simple - nasal	-136.30	50.1629	-234.616	-37.98	1241.7	-2.7171	0.007



(a) nasal

(b) open

(c) simple

Figure 4.8: The means and standard errors of MD

The ANOVA results of the final model for MD indicated that, except for RhymeType [$F(2, 73.0) = 0.677$, $p = 0.511$], significant main effects were found both on Group [$F(2, 71.0) = 13.140$, $p < 0.001$] and TonePair [$F(5, 1103.4) = 4.100$, $p = 0.001$].

The estimates of the final model for MD are displayed in Table 4.11. To save space, only the effects with significance were included in this table. As displayed in Table 4.11, the effect of the NH group was negative and highly significant, confirming that NH people had experienced much less mouse movements in processing the experiment trials. Meanwhile, except for T3T4 which was positive and marginally significant, all other tone pairs were positive and significant, indicating the participants had experienced more mouse movements when processing these tone pairs than the baseline T1T2.[6] Finally, no significant effect was found among the rhyme types, showing that different rhymes did not cause too much change in mouse movement.

(5) The time elapsed to overcome the maximum deviation of trajectory (MDTime)

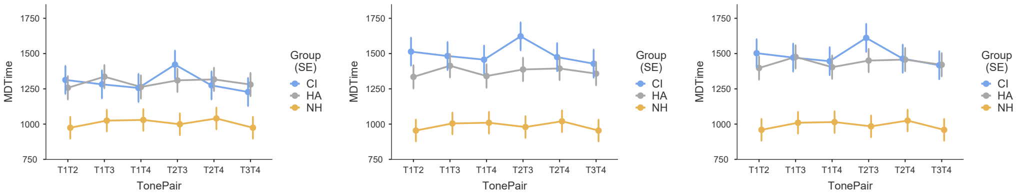
Figure 4.9 displays the means and standard SEs of MDTime values.

---

[6] More mouse movements mean increased cognitive load in processing stimuli, which could lead to more errors and longer reaction times but the link might not be strong. In the current study, beside the two difficult tone pairs T2T3 and T2T4, we can also see the participants experience more mouse movements in processing T1T3 and T1T4 but not commit significantly more mistakes and longer reaction times.

Table 4.11: The parameter estimates of the fixed effects for MD

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	0.88901	0.0536	0.78394	0.9941	72.4	16.5821	< 0.001
Group1	HA - CI	-0.11727	0.1263	-0.36473	0.1302	71.0	-0.9288	0.356
Group2	NH - CI	-0.56080	0.1228	-0.80139	-0.3202	71.0	-4.5685	< 0.001
RhymeType1	open - nasal	-0.02118	0.0211	-0.06258	0.0202	73.2	-1.0027	0.319
RhymeType2	simple - nasal	-0.00160	0.0275	-0.05554	0.0523	73.0	-0.0581	0.954
TonePair1	T1T3 - T1T2	0.07568	0.0244	0.02782	0.1235	1103.1	3.0991	0.002
TonePair2	T1T4 - T1T2	0.09515	0.0244	0.04723	0.1431	1103.7	3.8917	< 0.001
TonePair3	T2T3 - T1T2	0.07018	0.0244	0.02232	0.1180	1103.1	2.8740	0.004
TonePair4	T2T4 - T1T2	0.08596	0.0244	0.03804	0.1339	1103.7	3.5160	< 0.001
TonePair5	T3T4 - T1T2	0.04329	0.0244	-0.00457	0.0911	1103.1	1.7727	0.077



(a) nasal (b) open (c) simple

Figure 4.9: The means and standard errors of MDTime

Observing Figure 4.9, one could note that while the NH group's MDTime values were apparently lower than the other two groups in all three subplots, the HA group's MDTime values were a little lower than the CI group's values in general. Meanwhile, the HA group's MDTime values were even higher than the latter in some particular tone pairs. Additionally, the CI group's values also uniformly displayed a characteristic that T2T3 had the highest MDTime values in all rhyme types. Thus, combined with the results of MD and MDTime, it would not be difficult to arrive at the conclusion that the CI group have to move mouse movements more frequent and longer in processing the T2T3 trials than processing the other tone pairs.

Table 4.12 presents the ANOVA results of the final model for MDTime, in which all main effects or interactions were significant. In particular, the covariate MD reported the largest *F* value in the table, implying it contributed a lot to the final model. The following simple effect test of RhymeType on Group reported that, in the CI and HA group, open and simple rhyme were not significantly different but both of them were significantly different from nasal rhyme; in the NH group, no significance was found among the three rhyme types. As for the simple effect test of TonePair on Group, significance was found only in the CI group's T1T3-T2T3, T1T4-T2T3, T2T3-T2T4, and T2T3-T3T4. These results indicated that, except for the case of T1T2, the MD time of the CI group's T2T3 was much longer than other tone pairs.

Table 4.12: The ANOVA table of the final model for MDTime

	F	Num df	Den df	*p*
Group	9.18	2	69.0	< 0.001
RhymeType	21.63	2	1231.1	< 0.001
TonePair	3.01	5	1231.4	0.010
MD	22.07	1	1303.9	< 0.001
Group × RhymeType	9.76	4	1231.1	< 0.001
Group × TonePair	2.00	10	1231.1	0.031

Note: Satterthwaite method for degrees of freedom.

The estimates of the final model for MDTime are displayed in Table 4.13. To save space, only the effects with significance were included in this table. In Table 4.13, the effect of the NH group was negative and highly significant, indicating NH people had used much shorter MD time to process the experiment trials than CI people. Moreover, both open and simple rhymes reported positive and significant effects, implying the participants had consumed more time in processing the syllables with open and simple rhymes in the experiment. Thirdly, the effect of T2T3 was positive and significant, revealing that more MD time was needed in Processing T2T3 trials. Fourthly, the effect of the covariate MD was positive and highly significant, confirming that the participants would consume longer MD time if they moved the mouse for longer distances. Finally, the interactions between HA and open, NH and open, and NH and simple were negative and significant, implying NH and HA people's MD times were more constant than CI people in processing different rhyme types.

4.2.3 Summary

To sum up, the results of the linear mixed models reported the following findings:

- The NH group performed much better than the CI and HA group, and the two deaf

Table 4.13: The parameter estimates of the fixed effects for MDTime

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	1262.68	46.3	1171.92	1353.4	68.0	27.2685	< 0.001
Group2	NH - CI	-430.53	116.8	-659.44	-201.6	69.7	-3.6861	< 0.001
RhymeType1	open - nasal	86.00	17.0	52.63	119.4	1231.1	5.0511	< 0.001
RhymeType2	simple - nasal	104.95	17.0	71.63	138.3	1231.0	6.1729	< 0.001
TonePair3	T2T3 - T1T2	61.94	24.1	14.67	109.2	1231.9	2.5682	0.010
MD	MD	117.95	25.1	68.74	167.2	1303.9	4.6980	< 0.001
Group1 × RhymeType1	HA - CI × open - nasal	-123.26	43.9	-209.32	-37.2	1231.1	-2.8072	0.005
Group2 × RhymeType1	NH - CI × open - nasal	-220.48	42.6	-304.04	-136.9	1231.1	-5.1716	< 0.001
Group2 × RhymeType2	NH - CI × simple - nasal	-204.61	42.6	-288.13	-121.1	1231.0	-4.8011	< 0.001

groups performed similarly in the lexical tone perception experiment. In this study, the absolute advantage of the NH group was found in all aspects: the highest accuracies, the shortest reaction times, and the least amount of mouse movement. As for the performance between the two deaf groups, although none of the parameters showed significant results, they did have some differences in value: the CI group was a little better than the HA group in accuracy parameters, and the HA group was a little better than the CI group in all other parameters. That is, CI people had higher accuracy rates, but they consumed longer reaction time and made more mouse movement in processing the experiment trials than HA people in the experiment.

- TonePair was an important factor that impacted the participants' performance in the experiment. Except for the case of RAU_Incorrect1 and InitTime, TonePair was included in the final models of all the remaining parameters by reporting significant effects, indicating that the participants had different performances on the trials of these tone pairs. Compared with the baseline T1T2, the processing of T2T3 and T2T4 trials had significantly longer MD time and reaction time, experienced more mouse movement, and still made more incorrect choices. In contrast, although T3T4 was not significantly different from T1T2 on MD and RT parameters, the RAU parameters of T3T4 were significant better than that of T1T2, indicating that T3T4 was much easier for the participants than the baseline T1T2.

- RhymeType also played an important role in the participants' performance in the experiment. Except for RT and MD, all other parameters were reporting significance. In detail, the syllables with the three types of rhymes were not significantly different in the duration of reaction time and the distance of mouse movement. However, compared with the syllables with nasal rhymes, the participants had consumed longer initial time and MD time to process the syllables with open and simple rhymes but acquired higher accuracies in processing these syllables.

- Concerning the incorrectly chosen data, it seemed that the impact of ErrorType was more important than that of ToneDir. In this study, significant effects were found on ErrorType and its interactions with RhymeType. That is, the participants had made more OD and TD mistakes in processing the syllables with open and simple rhymes, and made more RD mistakes in processing the syllables with nasal rhymes. In contrast, no significance was found on ToneDir in the experiment. That is to say, the number of ToneDir 1 was not significantly different from the number of ToneDir 2 in all six tone pairs.

4.3 Discussion

Now, it is time to answer the three questions advanced at the beginning of the chapter.

4.3.1 Do the participant groups in the current task perform better than in the synthesized tone perception tasks?

As for the first question, the answer would be yes. In this study, the accuracy rates of CI, HA, and NH people were 68.52%, 67.28%, and 98.33%. Considering the chance level in a 4AFC task was 25.00%, all three participant groups correctly identified all four Mandarin tones in the current experiment in general. Compared with their performance in the synthesized tone identification tasks[7], it would be safe to say that the deaf groups had performed much better when identifying Mandarin tones in real human-produced syllables.

Taking the synthesized tone discrimination under consideration, we could see that all three participant groups performed much better in the current experiment when compared with the synthesized tone perception experiment. In the experiment of the previous chapter, even NH people could barely identify and discriminate the categories of Mandarin tones in the synthesized tone perception tasks. In contrast, in the current experiment, they hit

---

[7] Since the synthesized tone identification tasks only involve two options, and there is no clear answer to which opinion is the correct one, we could not get the accuracy rates of that task. Thus, to compare with the current results, we adopted the categorical identification data displayed in Table 3.5, which indicated the category identification of the synthesized tone by the three groups were 37.50% for CI (45 in 120), 56.41% for HA (88 in 156), and 95.00% for NH (171 in 180) respectively.

an extremely high accuracy rate in identifying Mandarins tones embedded in real human-produced syllables. Moreover, contrasting with their performance in the experiments of the previous chapter, both CI and HA people performed much better in the current task, indicating that real human-produced stimuli were also easier to process than synthesized stimuli for the deaf participants. Last but more impressively, although CI people had performed worse than their HA counterparts in the synthesized tone perception experiments, they caught up with HA people in the current study with an even higher accuracy rate, revealing that CI people might benefit more from processing Mandarin tones embedded in real human-produced syllables.[8]

Theoretically speaking, the experimental design of the current experiment was more complex than the experiments in the previous chapter for two reasons. For one thing, because the current study involved four-alternative forced-choice task but not the two-alternative forced-choice task involved in the synthesized tone perception experiments, the chance level of one given trial in the current study was 25%, much lower than the chance level of a particular trial 50% in the experiments of the previous chapter. For another, this study manipulated the locations of the four options of each trial to make sure that the participants could not start to respond before visually processing all four options. To fulfill the experimental task, the current experiment continuously and intensively occupied the participants' visual attention to monitoring the mouse movement. Given the extra complexity of experiment design, it is no surprise to expect that the participants would do more poorly on this task than on the synthesized tone experiment. Therefore, the fact that the participants did better on this experiment is even more impressive.

Compared with the results of the synthesized tone perception experiment, the current results indicated that all three participant groups performed much better in the current study, and the CI group had caught up with the HA group. In our opinion, there were three possible reasons to account for this phenomenon. Firstly, the participants were more familiar with the stimuli of the current experiment which might help them to process and hit higher accuracies. In this study, all stimuli were real human-produced syllables that were accessible to the participants at all times. In contrast, although the experimenter tried to ensure the naturalness of the stimuli in the synthesized tone experiments, most of the stimuli were

---

[8] From the current results, one could see that all three participant groups did better in the real-world task than in the lab-task, which might reveal to the importance of ecological validity in experimental research. More important, compared with their performance in the synthesized tone experiment, the CI group got more improvements than the HA group in the current experiment, which might indicate that the CI participants benefit more from their hearing devices than their HA counterparts (see Section **??** for more details).

still quite unfamiliar to the participants because they could not access these stimuli in everyday life. As a result, the difference in familiarity might help the participants to take some advantage and perform better in the current study[9].

Secondly, the acoustic cues for tone perception were much easier to process in the current study. In the current study, the tones in all stimuli were considered as the prototypes of these tones in Standard Chinese that consisted of the typical acoustic cues for these tones. That is, the stimuli in the current study were probably located at the extremes of the continua for all six tone pairs. Thus, it would be quite easy for the participants to decide the tones in the current experiment. On the contrary, most of the synthesized stimuli consisted of the in-between acoustic cues of the prototypes of Mandarin tones, which could make it harder for the participants to decide the belongings of tonal categories for these stimuli. Especially during the discrimination experiment, all stimuli in a given trial were only differentiated by subtle acoustic cues of tones, which made it even harder for the participants to make the right decision. Finally, compared with the synthesized tone perception experiment, the CI group benefited more from their device in the current experiment. Limited by the developing level of contemporary multichannel technology, CI is well known for its deficits in encoding the cues of very-low-frequency such as tonal information. As a result, CI people could not collect enough tonal information from the acoustic cues with subtle nuance when processing the synthesized tone stimuli. In contrast, although they still could not gather enough tonal information directly when processing the real human-produced stimuli, benefiting from the CI device, the CI participants had managed to collect enough tonal information indirectly from other components of the syllables.[10] Thus, the CI participants reached a similar result as their HA counterparts on identifying Mandarin tones.

4.3.2 What impact does the complexity of syllabic rhyme on the participants' Mandarin tone perception?

As the first study focusing on Mandarin tone perception by the prelingually deaf people with severe deafness, the current research is also the first study that concerns the influence of the rhyme complexity on the perception of Mandarin tones. Considering no existing

---

[9] It is common sense that the prototypes of sound categories are formed based on familiar sounds but not unheard sounds. Thus, if the participants are more familiar with the stimuli, they should perform better in categorical identification tasks.

[10] The source of indirect information might be formant structure, tonal duration, etc. However, we can't tell the specific source clearly because more studies are needed to answer this question.

study has discussed this issue, we would like to begin with the impact of rhyme complexity on Mandarin rhyme acquisition.

The complexity of Mandarin rhymes does impact the order of rhyme acquisition for young children who are learning Chinese as the first language. As a tone language that has a quite simple syllabic structure, Standard Chinese only has three types of syllabic rhymes, the simple rhyme, the open rhyme, and the nasal rhyme. Among the three rhyme types, the nasal rhyme is more complex than the open rhyme, and the open rhyme is more complex than the simple rhyme. In the past, considerable research had repeatedly reported that nasal rhymes are harder than open rhymes, and open rhymes are harder to acquire than simple rhymes both for the typically developed children (Li, 2018; Liu, 2007) and the special children such as children with cognitive impairments (Fan and Wei, 2016).

As for deaf children, previous studies had also investigated their production and perception of Mandarin rhymes. In the area of rhyme production, some studies on articulation intelligibility had reported that although deaf children apparently lagged behind normal hearing children, they also followed the same line with the latter that the nasal rhymes were the hardest, followed by the open rhymes, and the simple rhymes were the easiest for deaf children (Xia et al., 2012; Huang et al., 2021). For instance, a recent study investigated the articulation intelligibility of four participant groups aged four to six years old with different deafness conditions: normal hearing, mild deaf, moderate deaf, and severe deaf (Huang et al., 2021). In this study, the authors found that, although the four participant groups performed differently on the articulation intelligibility of Mandarin rhymes, they followed the same order that the simple rhymes had higher intelligibility rates than the open rhymes, and the open rhymes had higher intelligibility rates than the nasal rhymes. Especially, the intelligibility rates of the children with severe deafness were 62.17%, 31.22%, and 23.39% on simple rhymes, open rhymes, nasal rhymes respectively, indicating that the open and nasal rhymes were far more difficult to articulate than the simple rhymes for this participant group. In the area of rhyme perception, although some studies had investigated deaf children's perception of Mandarin rhymes in general (Mao et al., 2017; Yang et al., 2013; Zheng et al., 2016), none of them had assessed deaf children's different performance in the perception of the three rhyme types in detail, not to mention the relations between rhyme and tone perception.

Since the deaf participants in the current study are prelingually deaf adults with severe deafness starting in their early young ages, it would not be an unfounded guess that they have similar difficulties in processing Mandarin rhymes to those deaf children with severe deafness. Namely, the prelingually deaf adults had experienced and might be still experiencing more difficulties in producing nasal rhymes than simple and open rhymes. To a further step, considering that perception and production are closely related to each other, it might be easy to predict that prelingually deaf adults would suffer more difficulties in nasal rhyme's perception than simple and open rhymes. The current results, not surprisingly, report that there are significant effects of RhymeType and its interaction with ErrorType on the parameter RAU_Incorrect1. That is, the deaf participants made more incorrect decisions, and are distracted more by RDs than other distractors when processing tones with nasal rhymes. Thus, the current results prove that, although the prelingually deaf adults might have abundant experience in using spoken Chinese, they still experience more difficulties in identifying nasal rhymes than simple and open rhymes.

More importantly, the results of the correctly chosen data report that the main effects of RhymeType and its interactions with Group are significant on all parameters except for the case of MD. These results, for one thing, might indicate that the identification of Mandarin tones follows the general difficulty order that tones with nasal rhymes are harder to identify than tones with open rhymes, and tones with open rhymes are harder to identify than tones with simple rhymes. These results also show that rhyme complexity has different impacts on the three participant groups: the complexity order of the rhymes has a similar impact on the two deaf groups, in which tones with nasal rhymes are hardest to identify, tones with simple rhymes are easiest to identify, and tones with open rhymes are in the middle somewhere; for NH people, it seems the complexity order of the rhymes has no impact on identifying Mandarin tones.

Now, an interesting question could be advanced here is why rhyme complexity contribute deaf people's lexical tone perception. In our opinion, the acoustic features of the rhymes might be a promising direction of answering this question. In Standard Chinese, the tone is realized across the whole rhyme, including the nasal coda or the vowel (or glide) after the main vowel. For nasal rhymes, part of the tone will be realized on the nasal that is a much quieter sound than the vowel. As a result, compared with the fact that all parts of the tonal contour are relatively loud on CV and CVV syllables, part of the tonal contour (the part on the nasal) on CVN syllables is very quiet. Since deaf people have more problems in perceiving quiet sounds, they would experience more difficulties in processing Mandarin tones carried by nasal rhymes than carried by simple and open rhymes. For open rhymes, the latter vowel or glide is shorter and weaker than the main vowel. As a result, the part of tonal information carried by the glide is not that prominent. As a result, deaf people experience a little more difficulties in processing tones embedded in open rhymes than in simple rhymes because of the missing of part of tonal information.

4.3.3 Whether the participants face more difficulties in identifying the tones in some tone pairs in the current experiment?

In the past, considerable studies had repeatedly reported that T2T3 is harder than other tone pairs both in production and perception for normal hearing children (Jeng, 1979; Yang et al., 2021), deaf children (Han et al., 2009; Zhu et al., 2014b), and adult L2 learners of Standard Chinese (Hao, 2012; Yang, 2015). It is noted that, not in line with this consensus, some researchers reported that T2T4 is also a very difficult tone pair for deaf children (Lee et al., 2002; Shen et al., 2013b; Tao et al., 2015). Moreover, although some studies had noticed the chance of confusion from one tone to the other is higher than the reverse condition in a given tone pair, none of them had considered it seriously in a statistical way.

With these in mind, the current study advances whether participants face more difficulties in identifying tones in certain tone pair such as T2T3. It is noted that, to account this question, we have to answer the two sub-questions. Firstly, we have to check whether certain tone pairs are harder to identify than other tone pairs. On top of that, we also need to check if a tone is harder to identify than the other one in these harder tone pairs.

To answer the first sub-question, the current study tries to include TonePair as an independent factor in the Mixed linear model for RAU_Incorrect2, as well as in the models for all parameters of the correctly chosen data. As a result, the final models are successfully converged except for the case of InitTime. Considering InitTime is a parameter that only reflects the efforts participants have made to process the stimuli sounds, the current results might have two folds of meanings. On one hand, the factor TonePair could be not included in the model for InitTime indicates that at this stage, the participants haven't arrived at the period of judging the tonal targets from two competing tones. Thus, the differences in the initial times among the participants are totally caused by Group and RhymeType. On the other hand, the effects of TonePair on all other parameters are significant, indicating that the tonal information carried by the four options does impact the participants' judgements.

Corresponding to findings of the synthesized tone perception experiment, the current results indicate that identifying T2T3 is very hard for the deaf participants. Taking together with the findings of previous studies on prelingually deaf children, the current study proves that prelingually deaf adults still face a lot of difficulties in identifying T2 and T3 many years after using Standard Chinese as their first spoken language. Moreover, also agreeing with results of the synthesized tone perception experiment, the current experiment reveals that T2T4 is also significantly harder to identify than all remaining tone pairs. Thus, the current study provides more evidence for the argument that distinguishing T4 from T2 is also difficult for deaf people. In contrast, no difference is found among the tone pairs for the normal hearing participants in this study, indicating that the capability of identifying real human-produced tones in Standard Chinese has been developed well among normal hearing adults.

To answer the second sub-question, namely whether one tone is easier to be confused than the other in a given tone pair, the present study includes ToneDir as an independent factor in the final model for RAU_Incorrect2. As reported in 4.2.1, although there are more mistakes in TonDir1 than TonDir2, no significant difference is found between the two directions. Thus, the current results might prove that the hardness of identifying a given tone is not caused by one particular tone in that tone pair, but by some acoustic features shared by the two tones. In this study, there are one rising tone and one falling tone both in T2T3 and T2T4, which might indicate that the deaf participants have difficulties in differing the rising and falling features of the contour tones. In contrast, the participants have the best performance in processing T3T4, which might reveal that they have good capability in differing the tonal registers. In addition, the deaf participants also perform quite well in processing tone pairs containing the high level tone T1, which might suggest they are good at distinguishing a level tone from a contour tone in Standard Chinese.

Overall, the answer for this question is that deaf participants, but not normal hearing people, do face more difficulties in identifying T2T3 and T2T4 than other tone pairs. Inside each tone pair, one tone does not cause more mistakes to deaf people than the other tone.

4.4 Conclusion

The findings of the current chapter confirmed that all three participant groups benefited from the real human-produced stimuli in identifying Mandarin tones. As a result, compared with their performance in the synthesized tone perception experiment, they performed much better in the current experiment task. In detail, the NH group could easily identify the four tones under the distractions of competing tones and segmental distractors. Meanwhile, both the CI and HA group could correctly identify Mandarin tones in the majority of cases, suggesting they had basically established the categories of Mandarin tones. Especially, benefiting from the CI device, CI people had gathered enough tonal information indirectly from the real human-produced speech sound and reached the same level as their HA counterparts in identifying Mandarin tones.

Compared with the NH group, both of the two deaf groups perform much worse in the current task. One manifestation of the gap lies on the impact of rhyme complexity on tone identification. While rhyme complexity plays no role on normal hearing people's tone identification, it does impact the two deaf groups' performance in the current task. Moreover, the specific tones involved also impact the two deaf groups, but not the NH group in the current experiment. Among the six tone pairs, T2T3 and T2T4 are much harder to distinguish from each other than other tone pairs for deaf people. Overall, although both CI and HA adults have abundant experience of hearing device using and spoken Chinese speaking, they still face a lot of difficulties in perceiving Mandarin tones.

Chapter 5

The tone production experiment

After seeing how the deaf participants performed on Mandarin tone perception, it would be interesting to explore how they behave in tone production tasks. In the present chapter, we asked the participants to produce the stimuli that had been used in the lexical tone perception experiment. After that, we analyzed the acoustic characteristics of the lexical tones produced by the participants. We also invited 20 normal hearing judges to subjectively assess the extracted tone data produced by the participants. By fulfilling the two experiments, the current chapter was determined to detect the deaf participants' performance on Mandarin tone production, and took it as the basis for the discussion between the links between PDA's tone production and perception.

In particular, the experiments of this chapter tried to provide answers for the following questions. Firstly, what acoustic characteristics do the deaf participants have in Mandarin tone production? Secondly, can subjective assessment with multiple judges tell the difference of the three groups' production qualities of Mandarin tones? Finally, are the results of acoustic analysis correlated with the results of subjective assessment?

The organization of this chapter is as follows: after introducing the preparation of tone production data, we will introduce the two experiments separately in the methodology section: for the acoustic experiment, we would introduce the procedures of acoustic parameter extraction and data cleansing; for the subjective assessment experiment, we would introduce the judges, the assessment system, the data cleansing, etc. After that, we would introduce the statistical methods for analyzing the data of the two experiments in the methodology section. After that, we will report the findings of the two experiments in the result section. At last, we will try to answer the three questions advanced here and provide some discussion before concluding this chapter.

5.1 Methodology

This section consists of four parts: recording data preparation, acoustic analysis, subjective assessment, and statistical analysis.

5.1.1 Recording data preparation

As we have mentioned in the previous chapter, the word list for recording the tone production data is the same as the one for the lexical tone perception experiment. The word list, as shown in Table 4.1, is displayed here again in Table 5.1.

Table 5.1: The word list for data recording in the tone production experiment

Final	T1	T2	T3	T4
e [ɤ]	kē [kʰɤ^{55}] *section*	ké [kʰɤ^{35}] *shell*	kě [kʰɤ^{214}] *thrist*	kè [kʰɤ^{51}] *special*
i [i]	tī [tʰi^{55}] *kick*	tí [tʰi^{35}] *lift*	tǐ [tʰi^{214}] *body*	tì [tʰi^{51}] *instead*
a [a]	tā [tʰa^{55}] *he*	pá [pʰa^{35}] *climb*	tǎ [tʰa^{214}] *tower*	tà [tʰa^{51}] *tread*
u [u]	tū [tʰu^{55}] *bare*	tú [tʰu^{35}] *map*	tǔ [tʰu^{214}] *soil*	tù [tʰu^{51}] *rabbit*
ai [ai]	tāi [tʰai^{55}] *tire*	tái [tʰai^{35}] *raise*	kǎi [kʰai^{214}] *model*	tài [tʰai^{51}] *too*
ei [ei]	bēi [pei^{55}] *cup*	péi [pʰei^{35}] *accompany*	běi [bei^{214}] *north*	pèi [pʰei^{51}] *match*
ao [au]	tāo [tʰau^{55}] *draw out*	táo [tʰau^{35}] *escape*	tǎo [tʰau^{214}] *beg*	tào [tʰau^{51}] *cover*
ou [əu]	tōu [tʰəu^{55}] *steal*	tóu [tʰəu^{35}] *head*	kǒu [kʰəu^{214}] *mouth*	tòu [tʰəu^{51}] *lucid*
an [an]	tān [tʰan^{55}] *beach*	tán [tʰan^{35}] *talk*	tǎn [tʰan^{214}] *blanket*	tàn [tʰan^{51}] *sigh*
en [ən]	pēn [pʰən^{55}] *puff*	pén [pʰən^{35}] *basin*	kěn [kʰən^{214}] *agree*	bèn [bən^{51}] *stupid*
ang [aŋ]	tāng [tʰaŋ^{55}] *soup*	táng [tʰaŋ^{35}] *sugar*	tǎng [tʰaŋ^{214}] *lie*	tàng [tʰaŋ^{51}] *hot*
eng [əŋ]	pēng [pʰəŋ^{55}] *cook*	péng [pʰəŋ^{35}] *friend*	pěng [pʰəŋ^{214}] *boost*	pèng [pʰəŋ^{51}] *touch*

To avoid any possible distractions by the materials, all syllables in Table 5.1, are legitimate lexical words that can be displayed by frequently-used Chinese characters. Moreover, most of the onsets are aspirated stops [pʰ], [tʰ], and [kʰ], with [tʰ] being the most commonly used onset in Table 5.1. The reasons for this are as follows: firstly, there is a long gap between an aspirated stop's burst and the following vowel's onset, which minimizes the influence of that consonant on the following rhyme; secondly, when there is a longer gap between onset and rhyme, it is much easier to extract the pitch contour of the rhyme. Among the three aspirated stops [pʰ, tʰ, and kʰ], [tʰ] can combine with more rhymes than the other two consonants in Mandarin. In cases where [t^h] cannot be used as the onset, the other two aspirated stops are used to form meaningful syllables; if [p^h] and [k^h] also cannot surface, the unaspirated stops [p], [t], and [k] were substituted. It is noted that, since the onset consonant is not considered as an independent variable in the following statistic models, it is not necessary to control it across either rows or columns in Table 5.1.

With the 48 target words in Table 5.1, we embedded them in the carrying sentence "wǒ zài shuō X zhè gè zì" (*I am saying the character X*. X represents one of the embedded monosyllables), and asked all participants produce these sentences three times with their normal speech rates and loudness. Then, we extracted the soundtracks of the target words by Praat 6.1.53 (Boersma and Weenink, 2021). Since data recording was done by the participants themselves in their own places, two CI participants (CI_P14 and CI_P18) did not fulfill data recording tasks in this study. Among the remaining 74 participants * 48 targets words * 3 repetitions = 10656 tokens, no sound signal was found in 81 tokens (CI: 51; HA: 13; NH: 17). Therefore, there were 10575 tokens that were included for the following acoustic analysis.

5.1.2 The acoustic analysis

(1) Data extraction

Considering the recorded data lacked necessary quality control by the experimenters, the current study applied *Robust Epoch And Pitch EstimatoR* (REAPER) (Talkin, 2015) to extract F0 values. As a new tool for pitch tracking, REAPER has been proven robust in estimating F0 values even in noisy recording conditions or during intervals of creaky voice (Jouvet and Laprie, 2017; Dallaston and Docherty, 2019; Illner et al., 2020). Since REAPER tracks the F0 value every 5 ms, the tone contours with longer durations would have more F0 points than those tokens with shorter durations. Thus, we applied a method of curve fitting to interpolate the F0 data of each token to make sure all of them have 101 F0 values with equal time intervals. To ensure the accuracy of interpolation, the mean F0s before and after interpolation of each token were checked in this study.

(2) Parameter calculation

With the interpolated pitch contour data, we extracted two types of parameters, the global parameters reflecting the general features of the tones, and the dynamic tone parameters reflecting the F0 changes of the tone contours in the time dimension.

In recent years, the parabola parameters, namely the mean F0 (MeanF0), the slope (Slope), and the curve (Curve, one half the second derivative of f in Equation (5.1) if f were a parabola) were tested effective in describing the general features of Mandarin tones (Tang et al., 2021; Tupper et al., 2020). Thus, in line with these studies, we fit a parabola to the tone contour of each sound file by using Equation (5.1).

$$f(t) \approx c_0 + c_1(t - 1/2) + c_2\left[(t - 1/2)^2 - 1/12\right] \tag{5.1}$$

in which $f(t)$ is a particular F0 value, t is the relative time calculated by the timestamp of that F0 value, c_0 is the mean of $f(t)$, c_1 is the slope, and c_2 is the curve of that tone contour.

Conventionally, tone duration (Duration) is also a commonly used parameter for describing the general characteristics of tone contour. Moreover, our past studies had revealed that prelingually deaf adults have longer durations in all four tones than normal hearing adults (Chen et al., 2017f; Hou et al., 2019). Thus, this study also included duration as a parameter for describing the general features of Mandarin tones. Overall, the data size of the global parameters was 10575 tokens * 4 parameters in this study.

Considering the F0 values of tone contours were constantly changing across the time dimension, the global parameters might not reveal this characteristic of tone contours. With this in mind, this study set up a dynamic tone dataset to analyze F0 changes in tone contours along the time dimension. That is, in line with the previous research, this study extracted 11 F0 values with equal time intervals from 0% to 100% time for each token. Since the absolute F0 values were different among the participants, this study transformed F0 into T-value to normalize the pitch data for all participants using Equation (5.2).

$$T = 5 \times \frac{\log x - \log b}{\log a - \log b} \tag{5.2}$$

in which x is the original F0 value in Hertz, a is the maximum F0, and b is minimum F0 of a speaker's pitch range.

After T-value calculation, the size of the dynamic tone dataset was 10575 tokens * 11 time points = 116325 rows in this study.

(3) Data cleaning

In this study, we cleansed the extracted data using two standards. First, if the duration of a token was outside the range of mean±2std (67 ms, 512 ms; mean = 290 ms, std = 111 ms), all data of this token were excluded from further analysis. Second, if the F0 value at a particular time point of a given token was below 25 Hz, only the F0 value at that time point was dropped, and the other data of this token remained for further analysis.

After data cleaning, there were 10142 rows of data for the global parameters, and 116114 rows of data for the dynamic tone parameters.

5.1.3 The subjective assessment

(1) The materials

The materials for the subjective assessment were the extracted tokens from the sound files produced by the participants. Considering it would take too much time to judge all 10575 tokens, for each participant, only the second repetitions of the recordings for the word list in Table 5.1 were selected for subjective assessment. It is noted that, to make the assessment data more dynamic and ecological, we did not delete the tokens with no sound signal from the materials. That is, 19 no-sound-signal tokens (CI: 9; HA: 6; NH: 4) were included in the subjective assessment.[1] Overall, there were 74 * 48 = 3552 monosyllabic trials in the current experiment.

(2) The judges

In this study, 20 judges (9 males and 11 females), aged from 18 to 21 years (Mean = 19.42, Std = 0.99) were recruited for fulfilling the assessing task. All judges were college students who majored in pathologic speech rehabilitation, at Cangzhou Normal University. At the time of the experiment, all of them had no or limited experience in pathologic speech assessment. In addition, all judges spoke Standard Chinese as their first language and reported no history of hearing diseases or speech disorders.

(3) The assessment

---

[1] Because the experiment tasks are fulfilled by the participants in their own places, it is not surprising they might fail to record some stimuli. Using these silent tokens, we could check whether the judges are serious-minded in fulfilling the tasks. Thus, we did not delete those silent tokens or tell the judges there were some silent tokens in the assessment data. In fact, from the assessment results, we could see that all participants performed serious in the task.

Because of the Pandemic, the assessment was fulfilled by the judges at home with their own computers and headphones. Considering the convenience to the judges and the safety of the materials, we developed a program *Speech Subjective Evaluation System Training Model* (SSEST) (Chen et al., 2021), and enclosed the materials for judgment into SSEST.

With SSEST, the judges could listen to each trial multiple times before deciding its tone category. After deciding the tone category, the judges also needed to decide its score following the evaluation scale from one to five (1: poorest; 2: poor; 3: fair; 4: good; 5: finest) that we advanced and applied in the past studies (Chen et al., 2017e, 2021). In this study, the judges were encouraged to decide the target tone and its score for each and every trial. However, if the judges could not make the decision for a particular trial, what they needed to do was to make no decision and move to the next trial. After the subjective assessment was finished, the judges could download the result files and e-mail them to us. In this study, the judges were encouraged to fulfill the assessment in their own comfortable steps. Usually, it took the judges about one week to finish the assessment.

(4) Data processing

After collating all judges' results together, we found that all judges had made no decision on all 19 no-sound-signal trials, indicating that these judges were serious-minded in fulfilling the tasks. Thus, as the first step of data processing, we only excluded these no-sound-signal trials from further analysis.

Then, we processed each judge's remaining data in the following three steps:

- Step 1, for trials that tone categories were correctly judged, we calculated the mean score $\triangle J$ of them by Equation (5.3).

$$\triangle J = \frac{\sum_{i=1}^{n} x_i}{n} \tag{5.3}$$

  in which n is the number of the correctly judged trials, and x_i is the score (from 1 to 5) of the ith correctly judged trial assigned by the judge.

- Step 2, for the trials that were not judged or incorrectly judged, we uniformly assigned their scores by Equation (5.4).

$$x'_j = -1 * \triangle J \quad {}_{(j=1,2,\ldots,3533-n)} \tag{5.4}$$

in which the value of x'_j was remaining the same among all *3533-n* trials.

- Step 3, we took all trials' scores together and built up an adjusted dataset. Based on this adjusted dataset, we transformed all trials' scores into *Z-score* (DeVore, 2017) values using Equation (5.5).

  $$z_k = \frac{x''_k - \triangle J'}{\sigma} \quad {\scriptstyle (k=1,2,...,3533)} \tag{5.5}$$

  in which x''_k is the score of the kth trial, $\triangle J'$ is the mean value, and σ is the standard deviation of the adjusted dataset.

Finally, after the *Z-score* transformation was finished for all judges, we calculated the mean (MeanZ) and standard deviation (SDZ) of the *Z-score* for each and every trial. By now, we built up a dataset of subjective assessment for statistical analysis, in which the two parameters were considered dependent variables in the statistical models.

5.1.4 The statistical analysis

In this study, we would apply three types of statistics: to investigate the impacts of factors such as the participant groups (Group) and the gender of the participants (Gender) on the general tone features and the results of subjective assessment, the linear mixed model was a good option; to check the changes of tone contour across time dimension, the generalized additive mixed model would be a better choice; to test the relationships among the parameters of the general tone features and the subjective assessment, the correlation analysis was an effective method. To make it easier to follow for the readers, we would organize this section based on the research purposes, and introduce the involving statistical methods accordingly.

5.1.4.1 The analysis of the acoustic parameters

(1) The global parameters

Using the same methods we applied in the previous chapters, we fit a battery of linear mixed models for the four global parameters: Duration, MeanF0, Slope, and Curve. The possible independent factors in these models were Group (CI, HA, and NH), RhymeType

(Simple, Open, and Nasal), Tone (T1, T2, T3, and T4), and Gender (Male and Female).[2] Furthermore, the participants (Subject) and the numbers of the 144 tokens produced by each participant (SoundNumber) were considered possible random parameters in these mixed linear models.

The final models of the four global tone parameters were as follows:

- The final model of Duration was Duration ~ Group $*$ RhymeTpye + Group $*$ Tone + RimeTpye $*$ Tone + Group + RhymeType + Tone + (1 | Subject), in which Group, RhymeType, and Tone were the independent factors, 1 was the random intercept, and Subject was the random factor. It is noted that the simple coding method was used for all three independent factors in this final model, in which CI, nasal, and T1 were set as the baselines for Group, RhymeType, and Tone respectively. Moreover, Gender was not included in this final model because it failed to report any significance or the final model was not converged when containing this factor.
- The final model of MeanF0 was MeanF0 ~ Gender $*$ Tone + Gender + RhymeType + Tone + Duration + (1 | SoundNumber) + (1 | Subject), in which Group, RhymeType, and Tone were the independent factors, 1 was the random intercept, Subject and SoundNumber were the random factors. Considering the duration of a given tone contour could impact its F0, we included Duration as the covariant in this final model. It is noted that the simple coding method was used for all three independent factors in this final model, in which female, nasal, and T1 were set as the baselines for Group, RhymeType, and Tone respectively. Moreover, Group was not included in this final model because it failed to report any significance or the final model was not converged when containing this factor.
- The final model of Slope was Slope ~ Gender $*$ Tone + Gender $*$ RhymeType + RimeTpye $*$ Tone + Gender + RhymeType + Tone + (1+ Gender + Tone | Subject), in which Group, RhymeType, and Tone were the independent factors, 1 was the random intercept, and Subject was the random factor. In this model, Gender and Tone were

---

[2] One might note that we include gender as an independent factor in tone production analysis but not in tone perception analyses in the current study. The reason behind this is that gender might play more important role in the analysis of tone production than perception. Since male and female adults have dramatic differences on acoustical features of tones, we have to include gender as an independent factor in these linear mixed models because we use original acoustic parameters in tone production analysis. In contrast, since the impact of gender is not that important in tone perception, we do not need to consider it as an independent factor in tone perception analysis.

also included as the random slopes. It is noted that the simple coding method was used for all three independent factors in this final model, in which CI, nasal, and T1 were set as the baselines for Group, RhymeType, and Tone respectively. Moreover, Group was not included in this final model because it failed to report any significance or the final model was not converged when containing this factor.

- The final model of Curve was Curve ~ RhymeType $*$ Tone + RhymeType + Tone + Duration + (1 | Subject), in which RhymeType and Tone were the independent factors, 1 was the random intercept, Subject was the random factor. In this model, Duration was included as the covariant. It is noted that the simple coding method was used for the two independent factors in this final model, in which nasal and T1 were set as the baselines for RhymeType and Tone respectively. Moreover, Group and Gender were not included in this final model because they failed to report any significance or the final model was not converged when containing the two factors.

(2) The dynamic tone parameters

Considering the T value data are dynamic tone data, for each tone category, the present study adopted the generalized additive mixed model to check the changes in tone contour across the time dimension.

Taking T1 as example, using the *mgcv* package (Wood, 2011) in R 4.0.5 environment (Team, 2021), we created the generalized additive mixed model for T1's Tvalue by the following steps:

- Step 1, we set up the model with separate smooths Tvaule ~ Group + s(TimePoint, by = Group, bs = ”cr”). In this basic model, we used the smoothing class ”cubic regression splines” (bs=”cr”) to check each group's Tvalue changes over the 11 time points. By builting up this model, we could see whether the changes of tone contour were linear or not.

- Step 2, after converting the grouping factor Group into the ordered factor Group.ord, we set up the model with difference smooth Tvaule ~ Group.ord + s(TimePoint, bs = ”cr”) + s(TimePoint, by = Group.ord, bs = ”cr”), in which includes a parameter term for Group.ord, a smooth over TimePoint without any grouping specification, and a smooth over TimePoint with the group specification by=Group.ord. By doing this, we determined to check the differences among the three groups' tone contours.

- Step 3, to investigate the influence of Duration, we introduced this factor by building up the third model Tvaule ~ Group.ord + s(TimePoint, bs = ”cr”) + s(Duration, bs = ”cr”) + t(TimePoint, Duration) + s(TimePoint, by = Group.ord, bs = ”cr”), in which s(Duration, bs = ”cr”) was used to check the main effect of Duration, and t(TimePoint, Duration) was used to determine the interaction between Duration and TimePoint.

- Step 4, based on the third model, we set the model with random smooth Tvaule ~ Group.ord + s(TimePoint, bs = ”cr”) + s(Duration, bs = ”cr”) + t(TimePoint, Duration) + s(TimePoint, by = Group.ord, bs = ”cr”) + s(TimePoint, Subject, bs = ”fs”, xt = ”cr”), in which Group.ord was an ‘ordered factor’ converted from the categorical grouping variable Group, TimePoint and Duration were two time factors.[3] In this final model, we added the random effects of Subject to the smooth over TimePoint using the smoothing class ”factor smooth interactions” (bs=”fs”).

Using the same method, we also fit the generalized additive mixed models the other tones.

5.1.4.2 The analysis of the subjective assessment parameters

In this study, we also fit the linear mixed models for MeanZ and SDZ. The same to the models for the global acoustic parameters, the possible independent factors in these models were Group, RhymeType, Tone, and Gender, the possible random parameters were Subject and SoundNumber in the final models.

The final models of the correctly chosen data were as follows:

- Using MeanZ as the dependent variable, the final model was MeanZ ~ Group $*$ Tone + RhymeType $*$ Tone + Group $*$ RhymeType + Group + RhymeType + Tone + (1 | Subject), in which Group, RhymeType, and Tone were the independent factors, 1 was the random intercept, and Subject was the random factor. It is noted that the simple coding method was used for all three independent factors in this final model, in which CI, nasal, and T1T2 were set as the baselines for Group, RhymeType, and Tone

---

[3] TimePoint is a factor that reflects the normalized time and Duration is a factor that reflects the absolute durations of tone tokens. Using TimePoint, we could check the general pattern of F0 change along time dimension; using Duration, we could see the impact of absolute duration on tone contour.

respectively. Moreover, Gender was not included in this final model because it failed to report any significance or the final model was not converged when containing this factor.

- Using SDZ as the dependent variable, the final model was SDZ ~ Group $*$ Tone + RhymeType $*$ Tone + Group + RhymeType + Tone + (1 | Subject), in which Group, RhymeType, and Tone were the independent factors, 1 was the random intercept, and Subject was the random factor. It is noted that the simple coding method was used for all three independent factors in this final model, in which CI, nasal, and T1T2 were set as the baselines for Group, RhymeType, and Tone respectively. Moreover, Gender was not included in this final model because it failed to report any significance or the final model was not converged when containing this factor.

5.1.4.3 The correlation analysis between the global parameters and the subjective assessment parameters

To check the possible correlation between the results of subjective assessment and the results of acoustic analysis, we also put the subjective assessment parameters and the global acoustic parameters together, and adopted a battery of correlation analyses using the *Hmisc* package (Harrell Jr and with contributions from Charles Dupont and many others, 2021) and the *corrplot* package (Wei and Simko, 2021) in R 4.0.5 environment (Team, 2021).

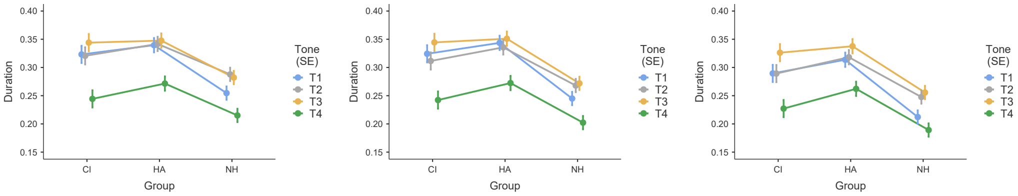
5.2 Results

5.2.1 The acoustic analysis

5.2.1.1 The global parameters

(1) Duration

Figure 5.1 displays the means and standard errors of Duration values. From Figure 5.1, one might note that the three rhyme types display a similar pattern of T3 > T1 ≈ T2 > T4 (“>” means “longer”) among the four tones, and a pattern of HA > CI > NH (“>” means “longer”) inside each type of tone.



(a) nasal (b) open (c) simple

Figure 5.1: The means and standard errors of Duration

Table 5.2 presents the ANOVA results of the final model for Duration, in which the main effects of Group, RhymeType, and Tone, as well as the interactions Group × RhymeType, Group × Tone, and RhymeType × Tone are significant. In the final model, the F value of Tone is larger than RhymeType, and RhymeType is larger than Group; revealing that Tone contributes more than RhymeType, and RhymeType contributes more than Group to the model. Moreover, among the significant interactions, the interaction of Group and Tone plays a more important role in the model than the other two interactions. the following simple effect and post hoc analysis of Group × Tone indicate that, while CI and HA people's tonal durations are not significantly different in all situations, NH people's four tones are significantly shorter than that of the other two groups.

Table 5.2: The ANOVA table of the final model for Duration

	F	Num df	Den df	*p*
Group	10.47	2	70.8	< 0.001
RhymeType	170.76	2	10047.2	< 0.001
Tone	912.51	3	10047.4	< 0.001
Group × RhymeType	7.91	4	10047.2	< 0.001
Group × Tone	28.75	6	10047.4	< 0.001
RhymeType × Tone	7.26	6	10046.9	< 0.001

Note: Satterthwaite method for degrees of freedom.

The estimates of the final model for Duration are displayed in Table 5.3. To save space, only the effects with significance were included in this table. As shown by Table 5.3, all main effects are significant except for the difference between the HA and CI group. In detail, NH people had a negative and significant effect, and HA people had a positive but nonsignificant effect, indicating that the tone durations produced by NH people were much shorter than CI and HA people; meanwhile, Both Open and Simple rhymes had negative and highly significant effects, revealing that the tone durations both in open and simple syllables, especially in simple syllables were much shorter than that in nasal syllables; furthermore, T2 and T3 had positive and highly significant effects, T4 had a negative and highly significant effect, showing the pattern of T3 > T2 > T1 > T4 ("$>$" means "longer") that follows the general rule of tone durations in Standard Chinese.

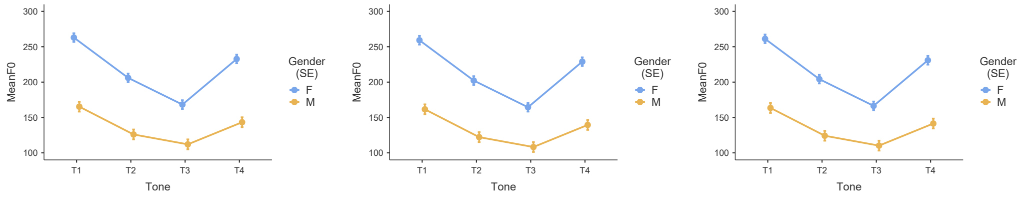
Table 5.3: The parameter estimates of the fixed effects for Duration

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	0.28758	0.00754	0.27279	0.30236	70.8	38.126	< 0.001
Group2	NH - CI	-0.05462	0.01890	-0.09167	-0.01757	70.8	-2.890	0.005
RhymeType1	Open - Nasal	-0.00499	0.00145	-0.00784	-0.00214	10046.9	-3.434	< 0.001
RhymeType2	Simple - Nasal	-0.02531	0.00145	-0.02816	-0.02246	10047.3	-17.418	< 0.001
Tone1	T2 - T1	0.00832	0.00167	0.00506	0.01159	10046.9	4.995	< 0.001
Tone2	T3 - T1	0.02389	0.00170	0.02056	0.02723	10047.6	14.053	< 0.001
Tone3	T4 - T1	-0.05768	0.00166	-0.06093	-0.05443	10047.4	-34.790	< 0.001
Group2 × RhymeType1	NH - CI × Open - Nasal	-0.01071	0.00362	-0.01780	-0.00362	10046.8	-2.961	0.003
Group1 × RhymeType2	HA - CI × Simple - Nasal	0.00785	0.00377	4.61e-4	0.01525	10047.4	2.082	0.037
Group2 × RhymeType2	NH - CI × Simple - Nasal	-0.00873	0.00363	-0.01583	-0.00162	10046.9	-2.406	0.016
Group2 × Tone1	NH - CI × T2 - T1	0.03572	0.00415	0.02760	0.04385	10046.9	8.617	< 0.001
Group1 × Tone2	HA - CI × T3 - T1	-0.01277	0.00443	-0.02145	-0.00410	10047.9	-2.885	0.004
Group1 × Tone3	HA - CI × T4 - T1	0.01078	0.00430	0.00235	0.01921	10047.5	2.506	0.012
Group2 × Tone3	NH - CI × T4 - T1	0.03924	0.00414	0.03114	0.04735	10046.9	9.487	< 0.001
RhymeType1 × Tone1	Open - Nasal × T2 - T1	-0.01019	0.00398	-0.01799	-0.00239	10046.9	-2.559	0.011
RhymeType2 × Tone2	Simple - Nasal × T3 - T1	0.01586	0.00404	0.00793	0.02378	10046.9	3.922	< 0.001
RhymeType2 × Tone3	Simple - Nasal × T4 - T1	0.01649	0.00395	0.00874	0.02423	10046.9	4.173	< 0.001

The significant interactions between every two factors were as follows. First, the interactions between Group and RhymeType indicated that the duration differences among the three rhyme types of NH people were much smaller than that of CI people, and the difference between simple and nasal syllables of HA people was even larger than that of CI people. Second, the interactions between Group and Tone showed that NH people had a much more distinguishable tone pattern in tonal duration than CI and HA people. In detail, for NH people, the duration differences between T2 and T1, T4 and T1 were larger than that of CI people, and the duration difference between T3 and T1 of NH people was similar to CI people; for HA people, the duration difference between T4 and T1 was larger than that of CI people, and the duration difference between T2 and T1 was similar to that of CI people, and the difference between T3 and T1 of HA people was even smaller than their CI counterparts. Finally, the interactions between RhymeType and Tone revealed that simple syllables had a much more distinguishable tone pattern in tonal duration than open and nasal syllables.

(2) MeanF0

Figure 5.2 displays the means and standard errors of MeanF0 values. From Figure 5.2, one might note that both male and Female participants display a similar MeanF0 pattern T1 > T4 > T2 > T3 (“>” means “higher”) on all three syllable types.



(a) nasal (b) open (c) simple

Figure 5.2: The means and standard errors of MeanF0

Table 5.4 presents the ANOVA results of the final model for MeanF0, in which the main effects of Gender, RhymeType, Tone, and Duration, as well as the interactions Gender × Tone are significant.[4] In the final model, the F value of Tone is much larger than the other factors, indicating MeanF0 might be a useful parameter to differentiate the tones in Standard Chinese. Moreover, the covariant Duration has the second largest F value, revealing that the mean F0 value of a given token is impacted by its duration. Furthermore, the main effect of Gender and its interaction with Tone also have larger F values, showing that Gender is also an important factor that impacts the mean F0 value. Finally, the F value of RhymeType is small, revealing that it has a limited contribution to the final model.

Table 5.4: The ANOVA table of the final model for MeanF0

	F	Num df	Den df	*p*
Gender	77.8	1	71.8	< 0.001
RhymeType	11.7	2	141.6	< 0.001
Tone	2302.0	3	155.3	< 0.001
Duration	174.8	1	10120.5	< 0.001
Gender × Tone	305.8	3	9928.2	< 0.001

Note: Satterthwaite method for degrees of freedom.

The estimates of the final model for MeanF0 are displayed in Table 5.5. From Table 5.5, one might note that the mean F0 values of the male participants were much lower than that of the female participants, and the mean F0 values of the four tones displayed the pattern

---

[4] Group is one of the main factors of interest in the models for all parameters. However, we did not find any significant effect of Group in the final model for MeanF0 and thus removed it from the final model. In the following sections, Group was also removed from the final models for Slope and Curve, because it failed to report any significance.

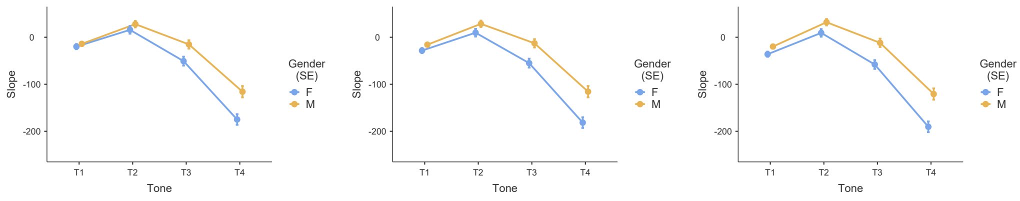
T1 > T4 > T2 > T3 (">" means "higher"). Meanwhile, the interactions between Gender and Tone reported that the male participants' MeanF0 differences among the four tones were smaller than that of the female participants, indicating that the male participants have a narrow pitch range than the female participants. Furthermore, although both the mean F0 values of open and simple syllables were significantly lower than that of nasal syllables, the effects of RhymeType were much smaller than that of the other factors.

Table 5.5: The parameter estimates of the fixed effects for MeanF0

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	175.19	4.589	166.19	184.180	72.1	38.18	< 0.001
Gender1	M - F	-80.87	9.169	-98.84	-62.895	71.8	-8.82	< 0.001
RhymeType1	Open - Nasal	-3.78	0.783	-5.32	-2.249	139.2	-4.83	< 0.001
RhymeType2	Simple - Nasal	-1.81	0.790	-3.36	-0.265	144.4	-2.29	0.023
Tone1	T2 - T1	-48.15	0.907	-49.93	-46.371	140.9	-53.10	< 0.001
Tone2	T3 - T1	-74.03	0.921	-75.84	-72.229	149.7	-80.41	< 0.001
Tone3	T4 - T1	-26.18	0.934	-28.01	-24.350	158.1	-28.04	< 0.001
Duration	Duration	-56.87	4.301	-65.30	-48.440	10120.5	-13.22	< 0.001
Gender1 × Tone1	M - F × T2 - T1	17.64	1.423	14.85	20.432	9926.9	12.40	< 0.001
Gender1 × Tone2	M - F × T3 - T1	41.44	1.445	38.61	44.272	9929.9	28.67	< 0.001
Gender1 × Tone3	M - F × T4 - T1	8.23	1.417	5.45	11.004	9926.8	5.81	< 0.001

(3) Slope

Figure 5.3 displays the means and standard errors of Slope values. From Figure 5.3, one can see that all three types of syllables show the same pattern T2 > T1 > T3 > T4 (">" means "higher"), in which only T2 has a positive slope among the four tones. From Figure 5.3, we also could see two more facts: first, the slopes of male participants are higher than female participants in all conditions except for T1 in nasal rhymes; second, the female participants' downward slopes are more exaggerated than that of the male participants.



(a) nasal (b) open (c) simple

Figure 5.3: The means and standard errors of Slope

Table 5.6 presents the ANOVA results of the final model for Slope, in which all main effects and the interactions between each two factors are significant. It is noted that the F value of Tone is much larger than the other factors, indicating Slope is also an important parameter to differentiate the tones in Standard Chinese. Moreover, Gender has the second largest F value, revealing that the slopes of tone contours are different between male and female participants. Finally, the F values of RhymeType and the interactions between each two factors are quite small, revealing that they have limited contribution to the final model.

Table 5.6: The ANOVA table of the final model for Slope

	F	Num df	Den df	*p*
Gender	33.65	1	59.7	< 0.001
Tone	78.45	3	70.6	< 0.001
RhymeType	11.15	2	9838.1	< 0.001
Gender × Tone	6.45	3	70.6	< 0.001
Gender × RhymeType	8.53	2	9838.0	< 0.001
Tone × RhymeType	2.37	6	9840.6	0.027

Note: Satterthwaite method for degrees of freedom.

The estimates of the final model for Slope are displayed in Table 5.7.

Table 5.7: The parameter estimates of the fixed effects for Slope

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	-46.375	2.92	-52.10	-40.647	59.7	-15.869	< 0.001
Gender1	M - F	33.902	5.84	22.45	45.357	59.7	5.801	< 0.001
Tone1	T2 - T1	42.850	4.67	33.70	51.995	69.5	9.183	< 0.001
Tone2	T3 - T1	-11.632	7.28	-25.91	2.645	72.1	-1.597	0.115
Tone3	T4 - T1	-127.500	8.37	-143.91	-111.093	71.8	-15.231	< 0.001
RhymeType1	Open - Nasal	-3.013	1.29	-5.54	-0.482	9837.9	-2.333	0.020
RhymeType2	Simple - Nasal	-6.094	1.29	-8.62	-3.565	9839.3	-4.723	< 0.001
Gender1 × Tone1	M - F × T2 - T1	6.334	9.33	-11.96	24.625	69.5	0.679	0.500
Gender1 × Tone2	M - F × T3 - T1	29.726	14.57	1.17	58.280	72.1	2.040	0.045
Gender1 × Tone3	M - F × T4 - T1	53.342	16.74	20.53	86.157	71.8	3.186	0.002
Gender1 × RhymeType1	M - F × Open - Nasal	6.689	2.58	1.63	11.748	9837.9	2.592	0.010
Gender1 × RhymeType2	M - F × Simple - Nasal	10.531	2.58	5.48	15.586	9839.0	4.083	< 0.001
Tone1 × RhymeType1	T2 - T1 × Open - Nasal	2.427	3.60	-4.64	9.493	9840.2	0.673	0.501
Tone2 × RhymeType1	T3 - T1 × Open - Nasal	4.166	3.67	-3.03	11.357	9839.3	1.135	0.256
Tone3 × RhymeType1	T4 - T1 × Open - Nasal	1.847	3.58	-5.17	8.867	9837.5	0.516	0.606
Tone1 × RhymeType2	T2 - T1 × Simple - Nasal	9.783	3.60	2.72	16.842	9852.8	2.716	0.007
Tone2 × RhymeType2	T3 - T1 × Simple - Nasal	8.962	3.67	1.77	16.151	9846.8	2.443	0.015
Tone3 × RhymeType2	T4 - T1 × Simple - Nasal	0.695	3.58	-6.32	7.709	9845.1	0.194	0.846

In Table 5.7, the most important fact is the main effect of Tone. Compared with the flat tone T1, T2 had a positive and significant effect, indicating that T2 is a rising tone; meanwhile, T4 had a negative and significant effect, indicating that T4 is a falling tone; in contrast, T3 had a zero effect, indicating that T3 might be a flat tone or a dipping tone that falls down first and rise up again afterward which makes the overall slope is flat. Another interesting fact in Table 5.7 is the interaction between Gender and Tone. That is, while the difference between male and female participants' T2 - T1 was non-significant, both the other two conditions were positive and significant, revealing that the male participants have flatter T3 and T4, and thus a much flatter tone pattern than female participants. Finally, the interaction between Gender and RhymeType reported that both the differences between male and female participants' Open - Nasal and Simple - Nasal were positive and significant, showing that the female participants' tone slopes were more stable among the three types of syllables.

(4) Curve

Figure 5.4 displays the means and standard errors of Curve values. From Figure 5.4, one can see that all three types of syllables show the same pattern T3 > T2 > T4 > T1 (">" means "larger"), revealing that T3 is the most curved tone and T1 is the flattest tone among the four tones. Inside each tone, Figure 5.4 also shows that the differences of curve values are small among the three rhyme types.

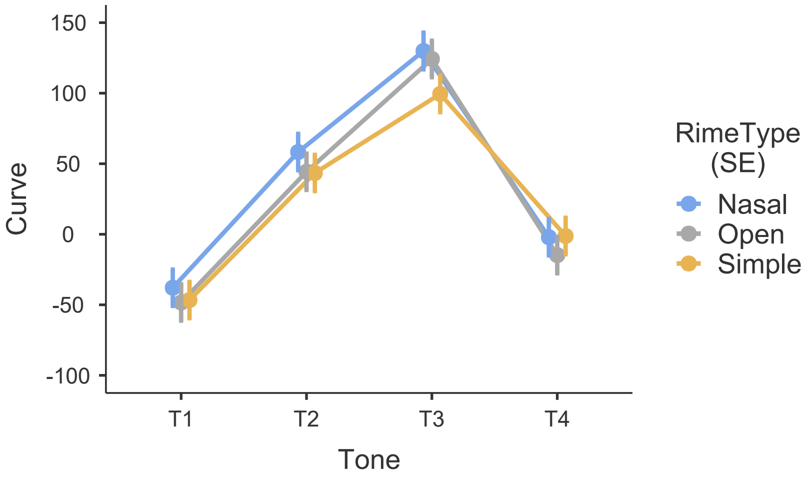


Figure 5.4: The means and standard errors of Curve

Table 5.8 presents the ANOVA results of the final model for Curve, in which the main effects of RhymeType and Tone, the main effect of the covariant Duration, as well as the interaction RhymeType × Tone is significant. In the final model, the F value of Tone is much larger than the other F values, revealing that Curve is an eligible parameter for differentiating tones. Moreover, Duration also has a large F value, indicating the curves of tone contours are also affected by their durations. Finally, the F values of the main effect of RhymeType and its interaction with Tone are quite small, showing that RhymeType contributes little to the final model.

Table 5.8: The ANOVA table of the final model for Curve

	F	Num df	Den df	*p*
RhymeType	6.63	2	10065	0.001
Tone	464.28	3	10079	< 0.001
Duration	60.08	1	9700	< 0.001
RhymeType × Tone	2.52	6	10055	0.020

Note: Satterthwaite method for degrees of freedom.

The estimates of the final model for Curve are displayed in Table 5.9.

Table 5.9: The parameter estimates of the fixed effects for Curve

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	29.06	11.77	5.98	52.135	72.2	2.468	0.016
RhymeType1	Open - Nasal	-10.69	3.83	-18.19	-3.193	10055.3	-2.795	0.005
RhymeType2	Simple - Nasal	-13.28	3.88	-20.89	-5.675	10073.2	-3.422	< 0.001
Tone1	T2 - T1	92.96	4.40	84.34	101.584	10056.4	21.138	< 0.001
Tone2	T3 - T1	162.17	4.51	153.33	171.005	10067.2	35.968	< 0.001
Tone3	T4 - T1	38.24	4.59	29.24	47.247	10100.9	8.325	< 0.001
Duration	Duration	204.35	26.36	152.68	256.028	9700.1	7.751	< 0.001
RhymeType1 × Tone1	Open - Nasal × T2 - T1	-3.55	10.76	-24.64	17.546	10054.6	-0.329	0.742
RhymeType2 × Tone1	Simple - Nasal × T2 - T1	-6.11	10.74	-27.17	14.946	10054.3	-0.569	0.570
RhymeType1 × Tone2	Open - Nasal × T3 - T1	4.80	10.94	-16.65	26.247	10054.4	0.439	0.661
RhymeType2 × Tone2	Simple - Nasal × T3 - T1	-21.73	10.94	-43.17	-0.294	10055.2	-1.987	0.047
RhymeType1 × Tone3	Open - Nasal × T4 - T1	-2.22	10.69	-23.17	18.732	10054.4	-0.208	0.835
RhymeType2 × Tone3	Simple - Nasal × T4 - T1	9.60	10.69	-11.35	30.546	10055.4	0.898	0.369

In line with the ANOVA results, Table 5.9 reveals that tone contours are affected by their durations, and the estimates of the tones indicate that the parameter Curve could tell the differences of the four tones in Standard Chinese. Moreover, as shown by Table 5.9, the effects of both the open and simple syllables are negative and significant, indicating that their tone contours experience fewer curve changes than the nasal syllables. As for the interaction between RhymeType and Tone, only the difference of T3 - T1 between Simple and Nasal is negative and significant, indicating that the T3 contours of the nasal syllables are more curved than the T3 contours of the simple syllables.

5.2.1.2 Dynamic tone parameters

Figure 5.5 displays the general tone patterns of the three groups building up by the T-values of tone tokens. From Figure 5.5, one can note that the tone patterns of the three groups, as well as the tone patterns of the male and female participants in each participant group, are quite similar. That is, T1 is a level tone occupying the middle to upper part of the tone register; T2 is a rising tone starting from the lower part and ending at the middle part of the tone register; T3 is a dipping-rising tone that dips from the middle of the tone register, reaches the bottom somewhere and then rises again to the middle of the tone register; and T4 is falling tone starting from the upper part and ending at the lower part of the tone register.

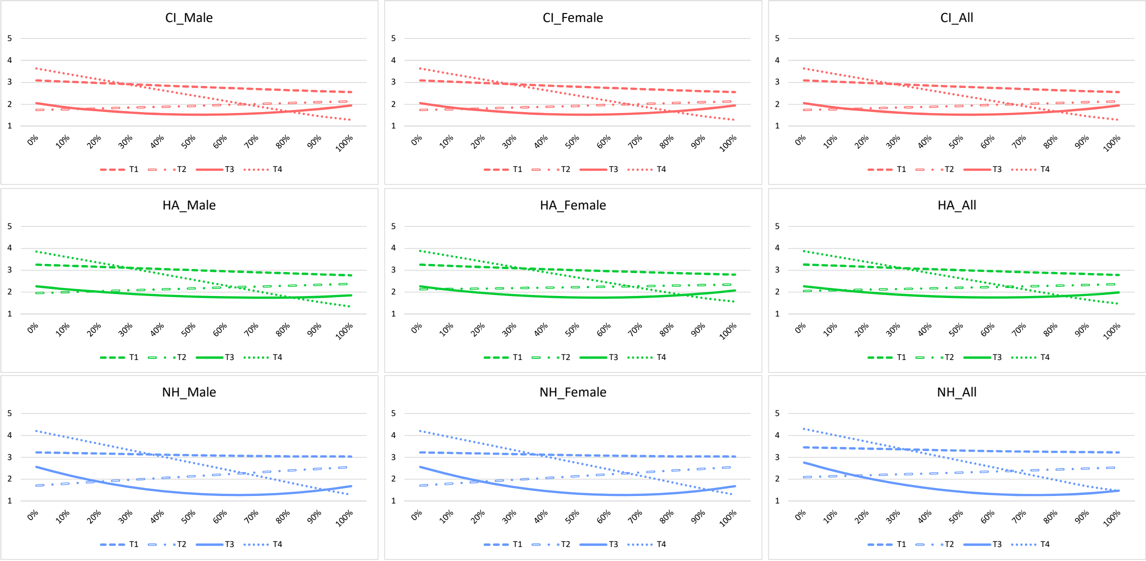


Figure 5.5: The general tone patterns of the three participant groups

Nevertheless, Figure 5.5 also reveals the differences of tone patterns between deaf and normal hearing people. In Figure 5.5, one striking difference between deaf and normal hearing people is that the tone patterns of the deaf groups are more compressed than that of the NH group. For one thing, the highest points of deaf people's tone registers, the ceilings of tone registers, are lower than that of NH people. In Figure 5.5, while the highest points of NH people are higher than 4 (NH_Male: 4.20 ± 0.42; NH_Female: 4.34 ± 0.34; NH_All: 4.29 ± 0.38), the ceilings of the two deaf groups are below 4 (CI_Male: 3.63 ± 0.64; CI_Female: 3.81 ± 0.58; CI_All: 3.71 ± 0.63; HA_Male: 3.85 ± 0.55; HA_Female: 3.88 ± 0.67; HA_All: 3.87 ± 0.62). For another thing, the lowest points of deaf people's tone registers, the floors of tone registers, are higher than that of NH people. In Figure 5.5, the lowest points of deaf people's tone registers are located at the ending points of T4 (CI_Male: 1.29 ± 0.72; CI_Female: 1.85 ± 1.13; CI_All: 1.63 ± 0.87; HA_Male: 1.34 ± 0.88; HA_Female: 1.57 ± 0.84; HA_All: 1.47 ± 0.87), are higher than NH people's ceilings of tone registers that locate at somewhere of T3 (NH_Male: 1.29 ± 0.47; NH_Female: 1.26 ± 0.71; NH_All: 1.28 ± 0.63).

Since the deaf groups' tone registers are more compressed, the differences among the four tones of deaf people are more subtle than that of normal hearing people. As a result, not only deaf people's four tones are acoustically harder to differentiate from each other, but also these tones might be more divergent from their counterparts of NH people. Thus, using the generalized additive mixed model, we analyzed the divergences of each tone among the three groups. The results of the generalized additive mixed model are displayed as follows.

(1) T1

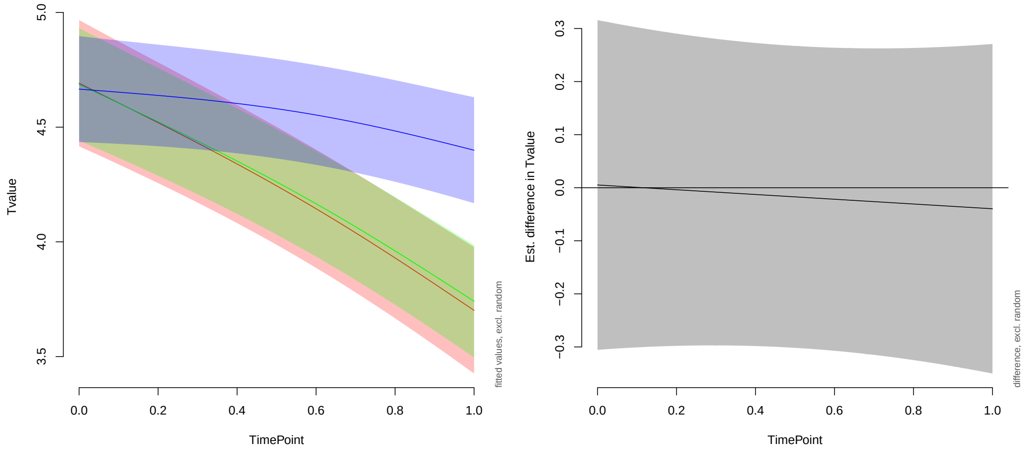
The model with separate smooths for T1 (Step 1) reported that the effective degrees of freedoms (*edf*s)[5] of all three groups were approaching 1 (s(TimePoint):GroupCI, 1.287; s(TimePoint):GroupHA, 1.479; s(TimePoint):GroupNH, 1.003), indicating that all three groups' estimated T1 contours (especially the HA group's T1 contour) were approximating straight lines.

The final model with random smooths for T1 (Step 4) reported that, compared with the baseline CI, NH's parametric coefficient ($t = 2.369$, $p = 0.018$) and approximate significance of smooth term ($t = 31.877$, $p < 0.001$) reported significance and high significance respectively. In other words, the estimated T1 contour of the NH group is significantly different from that of the CI group. On the contrary, the parametric coefficient ($t = 0.017$, $p = 0.906$) and approximate significance of smooth term ($t = 0.124$, $p = 0.725$) of the HA group reported no significance and no approximate significance, respectively. In other words, the estimated T1 contour of the HA group is not significantly different from that of the CI group.

---

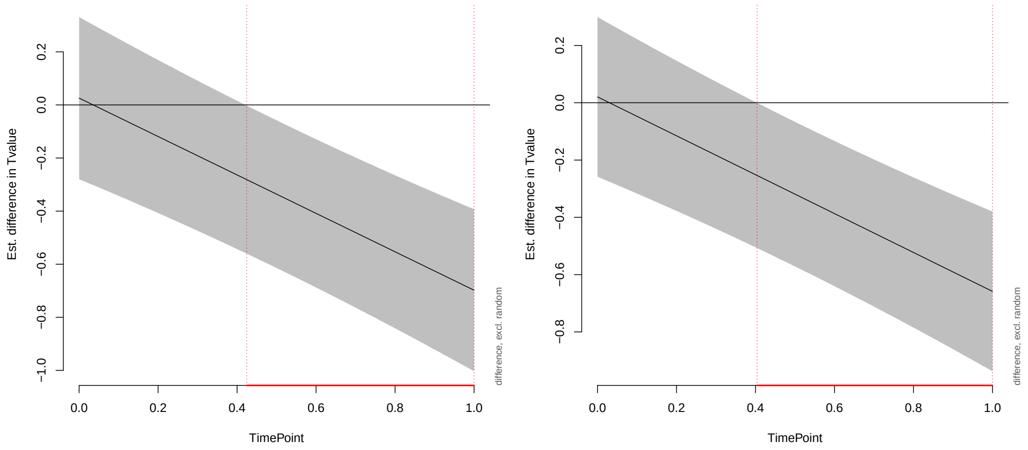
[5] The *edf* estimated from generalized additive models were used as a proxy for the degree of non-linearity in stressor-response relationships. (a) An *edf* of 1 is equivalent to a linear relationship, (b) an *edf* > 1 and ≤ 2 is a weakly non-linear relationship, and (c) an *edf* > 2 indicates a highly non-linear relationship. Highly non-linear relationships are those that are most likely to have inflection points and exhibit threshold responses. One could see Hunsicker et al. (2016) for more details.

Figure 5.6 shows the three participant groups' fitted T1 contours and the differences in T1 contours between each two participant groups in the final model.



(a) smooth results (CI: red HA: green NH: blue)

(b) Difference CI - HA



(c) Difference CI - NH

(d) Difference HA - NH

Figure 5.6: The fitted T1 results of the model with random smooths

From Figure 5.6a, one could note that the fitted T1 contours of the two deaf groups were quite similar, and they were apparently steeper than the NH group's T1 contour. The remaining subplots of Figure 5.6 revealed the details of the differences in tone contours among the three participant groups. From Figure 5.6b, one could see that there was no significant difference between the CI and HA groups' T1 contours from 0% to 100% of the time. In contrast, Figure 5.6c displayed that the NH group's T1 contour was significantly different from that of the CI group from 42.42% to 100% of the time, Figure 5.6d showed that the NH group's T1 contour was significantly different from that of the HA group from 40.40% to 100% of the time. Compared with the CI group, the HA group has a wider range of differences from the NH group in the T1 tone contour.

Besides Group, the final model of T1 reported that the main effect of TimePoint ($F = 38.205$, $p < 0.001$) and Duration ($F = 15.068$, $p < 0.001$), as well as the interaction between TimePoint and Duration ($F = 8.249$, $p = 0.002$) were significant, indicating that both TimePoint and Duration played important roles on the shapes of the fitted T1 contours. Moreover, the random effect of Subject ($F = 10.773$, $p < 0.001$) was significant, revealing that the variances caused by the participants also impacted the shapes of the fitted T1 contours.

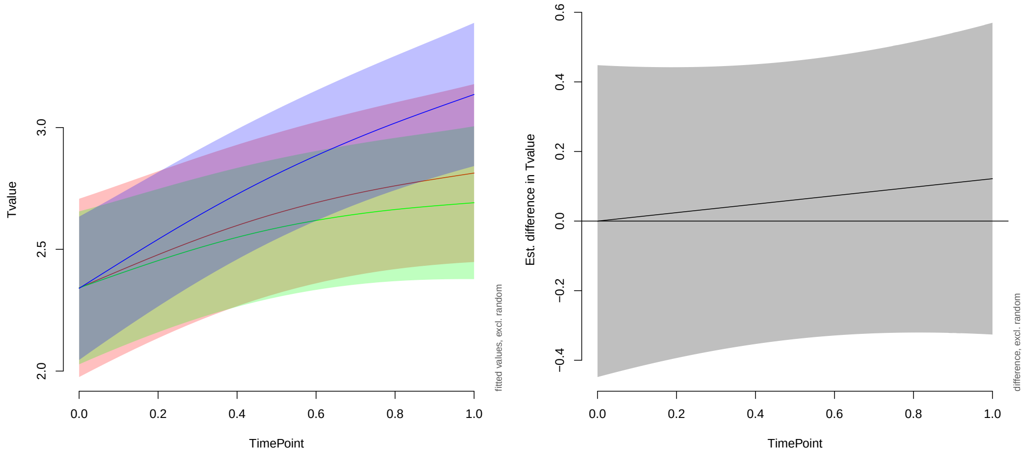
Overall, the results of the generalized additive mixed model for T1 revealed that, while all three participant groups' T1 contours were straight or approximately straight lines, the T1 contours of the NH group were different from the deaf groups. Moreover, all other factors TimePoint, Duration, and Subject significantly impacted the shapes of the fitted T1 contours.

(2) T2

The model with separate smooths for T2 (Step 1) reported that the *edf*s of all three groups were approaching 1 (CI: 1.598; HA: 1.170; NH: 1.005), indicating that the estimated T2 contours of all three groups (especially the contour of the NH group) were approximately straight lines.

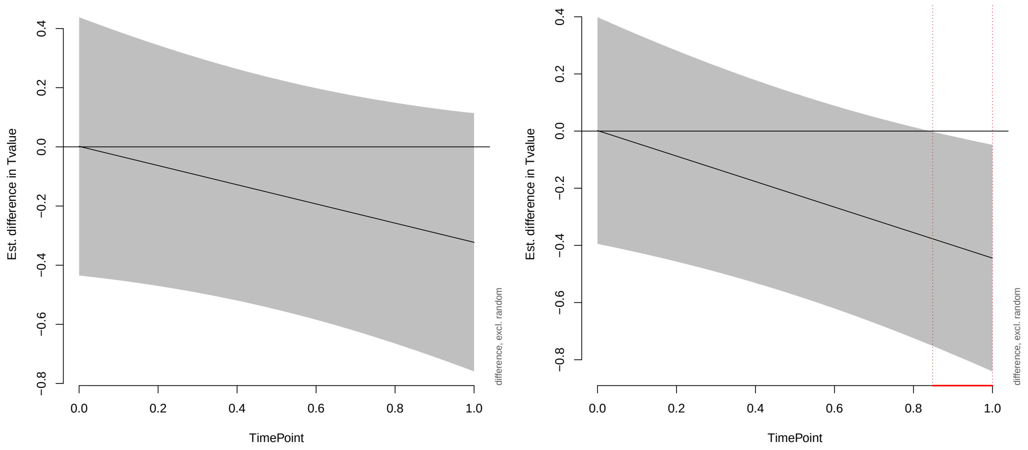
The final model with random smooths for T2 (Step 4) reported that, compared with the baseline CI, both NH and HA's parametric coefficient and approximate significance of smooth term failed to report any significance, indicating that the three groups' T2 contours were similar to each other. These results were further confirmed by Figure 5.7 that shows the three participant groups' fitted T2 contours and the differences in T2 contours between each two participant groups in the final model. From Figure 5.7a, one could see that all three groups' fitted T2 contours were quite similar. From the remaining subplots of Figure 5.7, only Figure 5.7d showed that the NH group's T2 contour was significantly different from that of the HA group from 84.85% to 100% of the time.

Besides Group, the final model of T2 reported that the main effect of TimePoint ($F = 5.115$, $p = 0.003$), the main effect of Duration ($F = 15.875$, $p < 0.001$), and the interaction between TimePoint and Duration ($F = 13.166$, $p < 0.001$) were significant, indicating that



(a) smooth results (CI: red HA: green NH: blue)

(b) Difference CI - HA



(c) Difference CI - NH

(d) Difference HA - NH

Figure 5.7: The fitted T2 results of the model with random smooths

both Duration and TimePoint played important roles on the shapes of the fitted T2 contours. Moreover, the random effect of Subject ($F = 22.520$, $p < 0.001$) was significant, revealing that the variances caused by the participants also impacted the shapes of the fitted T2 contours.

Overall, the results of the generalized additive mixed model for T2 revealed that all three participant groups' T2 contours were straight or approximately straight lines, and there was no apparent difference in T2 contours among the three participant groups. Moreover, all other factors TimePoint, Duration, and Subject significantly impacted the shapes of the fitted T2 contours.

(3) T3

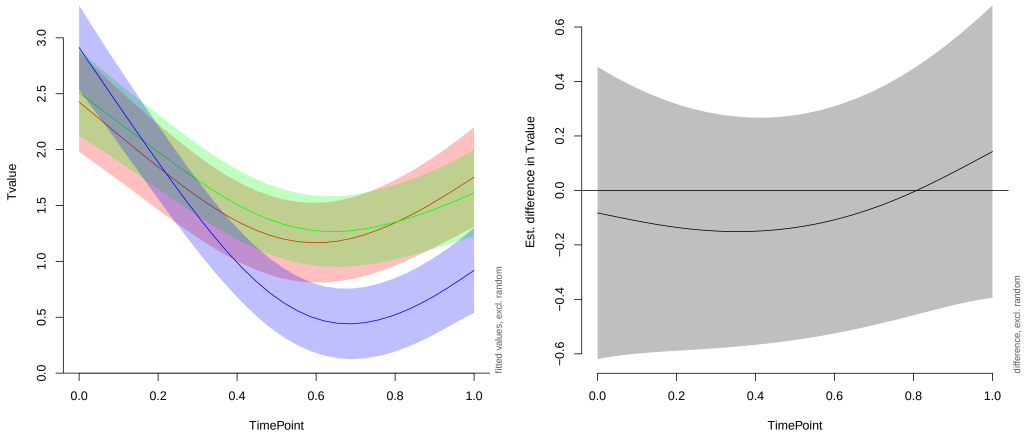
The model with separate smooths for T3 (Step 1) reported that the *edf*s of all three groups were much larger than 1 (CI: 3.434; HA: 3.069; NH: 4.029), indicating that the estimated T3 contours of all three groups were curved lines.

The final model with random smooths for T3 (Step 4) reported that, compared with the baseline CI, NH's parametric coefficient (t = -1.918, p = 0.055) and approximate significance of smooth term (F = 37.154, $p < 0.001$) reported approaching significance and high significance, respectively. Namely, the estimated T3 contour of the NH group is significantly different from that of the CI group. In contrast, both the parametric coefficient and the approximate significance of smooth term of the HA group reported no significance. That is, the estimated T3 contour of the HA group is not significantly different from that of the CI group.

Figure 5.8 shows the three participant groups' fitted T3 contours and the differences in T3 contours between each two participant groups in the final model. From Figure 5.8a, one could note that the fitted T3 contours of the two deaf groups were quite similar, and the two curves were less curved than the NH group's T3 contour. The remaining subplots of Figure 5.8 revealed the details of the differences in tone contours among the three participant groups. From Figure 5.8b, one could see that there was no significant difference between the CI and HA groups' T3 contours from 0% to 100% of the time. In contrast, Figure 5.8c displayed that the NH group's T3 contour was significantly different from that of the CI group from 43.43% to 100% of the time, Figure 5.8d showed that the NH group's T3 contour was significantly different from that of the HA group from 34.34% to 100% of the time. Compared with the CI group, the HA group has a wider range of differences from the NH group in the T3 tone contour.

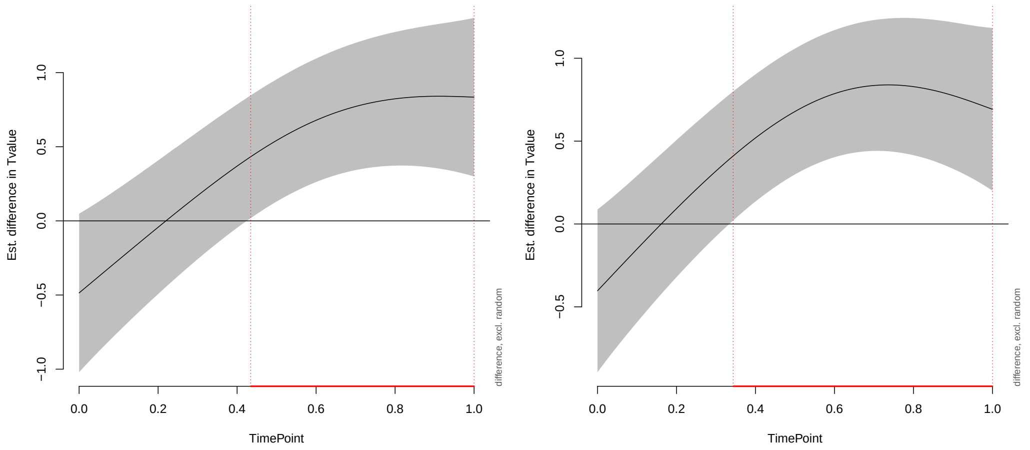
Besides Group, the final model of T3 also reported that the main effect of TimePoint (F = 10.713, $p < 0.001$) and Duration (F = 12.375, $p < 0.001$), as well as the interaction between TimePoint and Duration (F = 9.198, $p < 0.001$) were significant, indicating that both TimePoint and Duration played important roles on the shapes of the fitted T3 contours. Moreover, the random effect of Subject (F = 16.343, $p < 0.001$) was significant, revealing that the variances caused by the participants also impacted the shapes of the fitted T3 contours.

Overall, the results of the generalized additive mixed model for T3 revealed that all three participant groups' T3 contours were curves, and the NH group's T3 contour was more significantly curved that the deaf groups' T3 contours. Moreover, all other factors TimePoint, Duration, and Subject significantly impacted the shapes of the fitted T1 contours.



(a) smooth results (CI: red HA: green NH: blue)

(b) Difference CI - HA



(c) Difference CI - NH

(d) Difference HA - NH

Figure 5.8: The fitted T3 results of the model with random smooths

(4) T4

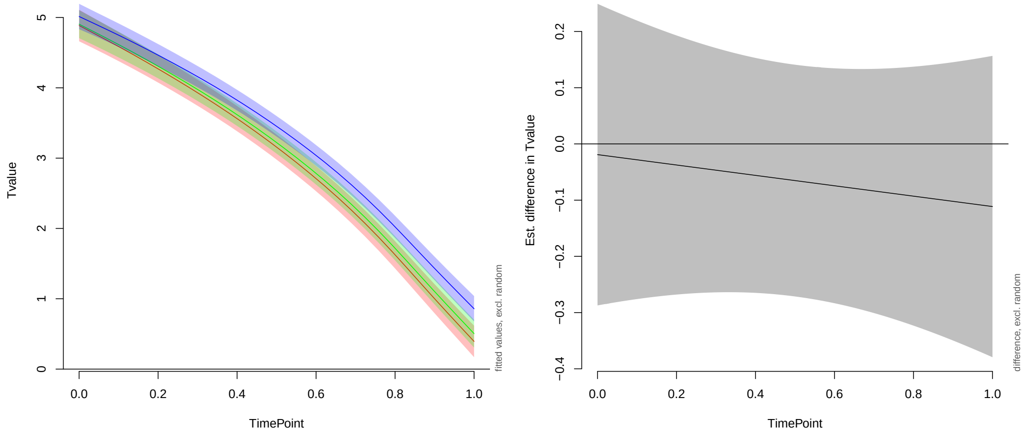
The model with separate smooths for T4 (Step 1) reported that the *edf*s of all three groups were much larger than 1 (CI: 3.473; HA: 3.908; NH: 4.042), indicating that the estimated T4 contours of all three groups were curved lines.

The final model with random smooths for T4 (Step 4) reported that, compared with the baseline CI, NH's parametric coefficient (t = 2.891, p = 0.004) and approximate significance of smooth term (F = 3.863, p = 0.050) reported significance and high significance, respectively. That is, the estimated T4 contour of the NH group is significantly different from that of the CI group. On the contrary, both the parametric coefficient (t = 0.620, p = 0.535) and the approximate significance of smooth term (F = 0.279, p = 0.598) of the HA group reported no significance. In other words, the estimated T4 contour of the HA group is not significantly different from that of the CI group.

Figure 5.9 shows the three participant groups' fitted T4 contours and the differences in T4 contours between each two participant groups in the final model. From Figure 5.9a, one could note that the fitted T4 contours of the two deaf groups were quite similar, and they were apparently steeper than the NH group's T4 contour. The remaining subplots of Figure 5.9 revealed the details of the differences in tone contours among the three participant groups. From Figure 5.9b, one could see that there was no significant difference between the CI and HA groups' T4 contours from 0% to 100% of the time. In contrast, Figure 5.9c displayed that the NH group's T4 contour was significantly different from that of the CI group from 26.26% to 100% of the time, Figure 5.9d showed that the NH group's T4 contour was significantly different from that of the HA group from 34.34% to 100% of the time. Compared with the CI group, the HA group has a narrower range of differences from the NH group in the T4 tone contour.

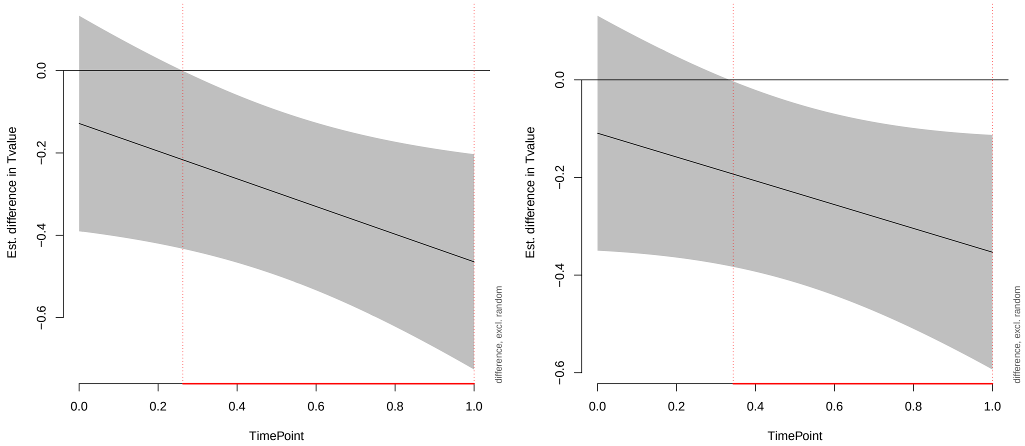
Besides Group, the final model of T4 also reported that the main effect of TimePoint (F = 200.603, $p < 0.001$) and Duration (F = 6.853, $p < 0.001$), as well as the interaction between TimePoint and Duration (F = 6.503, $p < 0.001$) were significant, indicating that both TimePoint and Duration played important roles on the shapes of the fitted T4 contours. Moreover, the random effect of Subject (F = 11.452, $p < 0.001$) was significant, revealing that the variances caused by the participants also impacted the shapes of the fitted T4 contours.

Overall, the results of the generalized additive mixed model for T4 revealed that all three participant groups' T4 contours were placed somewhere between straights and curves, and there was no statistically significant difference in T4 contours among the three participant groups. Meanwhile, the factors TimePoint, Duration, and Subject all impacted the shapes of the fitted T4 contours. Moreover, all other factors TimePoint, Duration, and Subject significantly impacted the shapes of the fitted T4 contours.



(a) smooth results (CI: red HA: green NH: blue)

(b) Difference CI - HA



(c) Difference CI - NH

(d) Difference HA - NH

Figure 5.9: The fitted T4 results of the model with random smooths

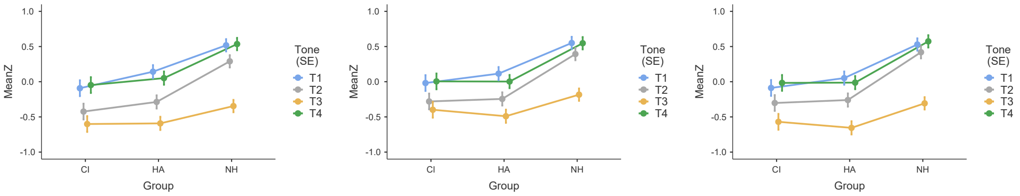
5.2.1.3 Summary

To sum up, the results of acoustical analysis reported that the three participant groups were not greatly different from one another in all global parameters except for Duration, indicating that the three groups had similar tone patterns in general. However, because the NH group has wider pitch range but shorter tonal durations than the deaf groups, the NH group's tone shapes were significantly different from the deaf groups when analyzing the dynamic tone parameters. Taken together, the two deaf groups performed quite similarly acoustically in producing Mandarin tones: both of them could differentiate all four Mandarin tones in production, but the tones produced by the deaf participants were worse than that of the NH participants from the statistical results of acoustical parameters. Based on these results, we predicted that the CI group and the HA group should perform worse than the NH group, but the two deaf groups would perform similar in the subjective assessment.

5.2.2 The subjective assessment

(1) The mean *Z-score* (MeanZ)

Figure 5.10 displays the means and standard errors of MeanZ values. From Figure 5.10, one might note that three types of syllables uniformly display the pattern NH > HA ≈ CI (">" means "higher") among the three participant groups, and the pattern T1 ≈ T4 > T2 > T3 (">" means "higher") among the four tones.



(a) nasal (b) open (c) simple

Figure 5.10: The means and standard errors of MeanZ

Table 5.10 presents the ANOVA results of the final model for MeanZ.

Table 5.10: The ANOVA table of the final model for MeanZ

	F	Num df	Den df	p
Group	11.04	2	71.0	< 0.001
RhymeType	7.94	2	3438.1	< 0.001
Tone	408.09	3	3438.0	< 0.001
Group × Tone	17.25	6	3438.0	< 0.001
RhymeType × Tone	3.15	6	3438.0	0.004
Group × RhymeType	2.19	4	3438.1	0.068

Note: Satterthwaite method for degrees of freedom.

From Table 5.10, one can see that the main effects of Group, RhymeType, and Tone, the interactions Group × Tone and RhymeType × Tone are significant, and Group × RhymeType is marginally significant. In the final model, the F value of Tone was much larger than Group, and Group was a little larger than RhymeType; revealing that Tone had contributed more to the model than the other factors. Since Group × Tone had the large F value among the interactions, this study also applied the following simple effect and post hoc analysis for it. And the results indicated that, while CI and HA people's MeanZ values were not significantly different in all situations, NH people's MeanZ were significantly different from the other two groups in all tone conditions except for T3.

The estimates of the final model for MeanZ are displayed in Table 5.11. To save space, only the effects with significance were included in this table.

Table 5.11: The parameter estimates of the fixed effects for MeanZ

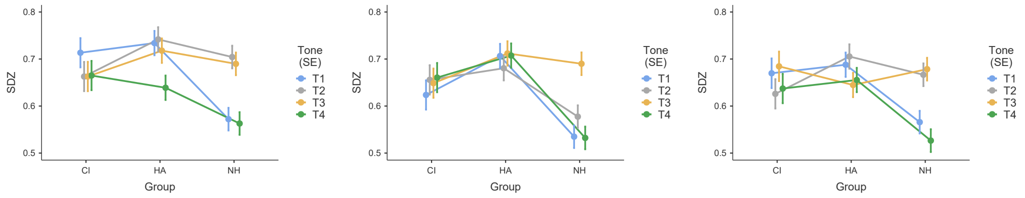
				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
Group2	NH - CI	0.52956	0.1338	0.26738	0.79173	71.0	3.959	< 0.001
RhymeType1	Open - Nasal	0.07082	0.0185	0.03459	0.10705	3438.0	3.831	< 0.001
Tone1	T2 - T1	-0.26707	0.0214	-0.30897	-0.22517	3438.0	-12.493	< 0.001
Tone2	T3 - T1	-0.65021	0.0214	-0.69214	-0.60827	3438.0	-30.389	< 0.001
Group1 × Tone2	HA - CI × T3 - T1	-0.22558	0.0552	-0.33380	-0.11735	3438.0	-4.085	< 0.001
Group2 × Tone2	NH - CI × T3 - T1	-0.35478	0.0537	-0.45995	-0.24960	3438.0	-6.611	< 0.001
Group1 × Tone3	HA - CI × T4 - T1	-0.13595	0.0551	-0.24395	-0.02796	3438.0	-2.467	0.014
RhymeType2 × Tone1	Simple - Nasal × T2 - T1	0.11793	0.0512	0.01764	0.21821	3438.0	2.305	0.021
RhymeType1 × Tone2	Open - Nasal × T3 - T1	0.12929	0.0511	0.02914	0.22943	3438.0	2.530	0.011
Group1 × RhymeType1	HA - CI × Open - Nasal	-0.10095	0.0477	-0.19440	-0.00750	3438.0	-2.117	0.034
Group1 × RhymeType2	HA - CI × Simple - Nasal	-0.09482	0.0478	-0.18860	-0.00103	3438.1	-1.982	0.048

As shown by Table 5.11, the estimates of the three participant groups were that, while no significant difference was found between the HA group and the CI group, the NH group had a negative and significant effect which indicated that NH people received higher scores in the subjective assessment. Moreover, the estimates of the types of syllables revealed that the tone tokens embedded in open syllables received higher scores in the assessment. Furthermore, while T4 reported no significant effect, both T2 and T3 reported negative and highly significant effects, revealing that T1 and T4 tokens received much higher scores than the tokens of T2 and T3. Linking the current results to the results of lexical tone perception, one could see that T2 and T3 were much harder than T1 and T4 both in production and perception.

The estimates of the interaction between every two factors were as follows. First, the difference between T3 and T1 for both HA and NH people, as well as the difference between T4 and T1 for HA people reported negative and significant effects, indicating that HA and NH people (especially HA people) experienced more score variations among the four tones than CI people in the assessment. Second, the difference between T2 and T1 between simple and nasal syllables, and the difference between T3 and T1 between open and nasal syllables were positive and significant, revealing that, compared with the baseline T1 tokens, T2 tokens in simple syllables and T3 tokens in open syllables apparently experienced more score variations than their counterparts in nasal syllables. Finally, the difference between open and nasal syllables, as well as the difference between simple and nasal syllables for HA people reported negative and significant effects, showing that HA people experienced fewer score variations than CI people among the three types of syllables.

(2) The standard deviation of *Z-score* (SDZ)

Figure 5.11 displays the means and standard errors of SDZ values.



(a) nasal (b) open (c) simple

Figure 5.11: The means and standard errors of SDZ

From Figure 5.11, one might note that, although there were apparent differences among the three types of syllables, the three subplots displayed a similar pattern that the NH group's SDZ values were impacted more by the tones than the deaf groups. In detail, while deaf people, especially CI people remained high SDZ values in all conditions, NH people had higher SDZ values in T2 and T3 conditions except for T2 in open rhyme, and lower SDZ values in T1 and T4 conditions. That is, compared with T1 and T4 tokens, the assessment scores of NH people's T2 and T3 tokens were more unstable, even meeting or exceeding the levels of the deaf groups in some conditions.

Table 5.12 presents the ANOVA results of the final model for SDZ, in which the main effects of Group, RhymeType, and Tone are significant. Meanwhile, the interactions Group × Tone and RhymeType × Tone are significant, Group × RhymeType and Group × RhymeType × Tone are marginally significant. In this final model, the F value of Tone is much larger than Group and RhymeType; revealing that Tone had more contribution to the model than the other factors. Since Group × Tone also had a large F value, this study also applied the following simple effect and post hoc analysis for it. And the results indicated that, while CI and HA people's SDZ values were not significantly different in all situations, NH people's SDZ values were significantly different from the other two groups in T1 and T4 tokens embedded in all three types of syllables, and T2 tokens in open syllables.

Table 5.12: The ANOVA table of the final model for SDZ

	F	Num df	Den df	*p*
Group	6.83	2	71.0	0.002
RhymeType	6.51	2	3426.4	0.001
Tone	14.04	3	3426.1	< 0.001
Group × RhymeType	2.12	4	3426.3	0.076
Group × Tone	12.09	6	3426.1	< 0.001
RhymeType × Tone	2.31	6	3426.1	0.032
Group × RhymeType × Tone	1.61	12	3426.1	0.081

Note: Satterthwaite method for degrees of freedom.

The estimates of the final model for SDZ are displayed in Table 5.13. To save space, only the effects with significance or marginal significance were included in this table.

Table 5.13: The parameter estimates of the fixed effects for SDZ

				95% Confidence Interval				
Names	Effect	Estimate	SE	Lower	Upper	df	*t*	*p*
(Intercept)	(Intercept)	0.65396	0.01040	0.63358	0.67435	71.0	62.8735	< 0.001
Group2	NH - CI	-0.05073	0.02607	-0.10183	3.64e-4	71.0	-1.9460	0.056
RhymeType1	Open - Nasal	-0.02802	0.00871	-0.04509	-0.01096	3426.0	-3.2190	0.001
RhymeType2	Simple - Nasal	-0.02643	0.00873	-0.04355	-0.00931	3426.6	-3.0262	0.002
Tone1	T2 - T1	0.02361	0.01007	0.00387	0.04334	3426.0	2.3445	0.019
Tone2	T3 - T1	0.03564	0.01008	0.01589	0.05539	3426.2	3.5366	< 0.001
Tone3	T4 - T1	-0.02465	0.01006	-0.04438	-0.00493	3426.1	-2.4499	0.014
Group2 × Tone1	NH - CI × T2 - T1	0.11256	0.02525	0.06306	0.16205	3426.0	4.4570	< 0.001
Group2 × Tone2	NH - CI × T3 - T1	0.13203	0.02527	0.08249	0.18157	3426.2	5.2239	< 0.001
RhymeType1 × Tone2	Open - Nasal × T3 - T1	0.04491	0.02463	-0.00336	0.09318	3426.0	1.8235	0.068
RhymeType1 × Tone3	Open - Nasal × T4 - T1	0.06264	0.02459	0.01444	0.11084	3426.0	2.5473	0.011
Group1 × RhymeType1 × Tone1	HA - CI × Open - Nasal × T2 - T1	-0.11657	0.06346	-0.24095	0.00782	3426.0	-1.8368	0.066
Group2 × RhymeType1 × Tone1	NH - CI × Open - Nasal × T2 - T1	-0.17226	0.06168	-0.29315	-0.05137	3426.0	-2.7927	0.005

As shown in Table 5.13, the estimates of the participant groups reported that no significant difference was found between the HA group and the CI group, and a marginally significant difference was found between the NH group and the CI group, indicating that the assessment scores of NH people's data were a little more stable than that of deaf people's data. Moreover, both open and simple rhymes reported negative and significant effects, revealing that the assessment scores of tokens embedded in open and simple syllables were more stable than tokens embedded in nasal syllables. Furthermore, while T4 reported a positive and significant effect, both T2 and T3 reported negative and significant effects, revealing that the scores of T1 and T4 tokens were more stable than that of T2 and T3 tokens.

Again, these results were corresponding to the results of lexical tone perception, indicating that T2 and T3 were much harder than T1 and T4 both in perception and production.

The estimates of significant interactions were as follows. First, the NH group had positive and highly significant effects on T2 - T1 and T3 - T1, indicating that NH people's T2 and T3 tokens had much larger variations than T1 tokens compared with the cases of the CI group. Second, T4 - T1 of the open rhymes reported positive and significant effects, revealing that the SDZ differences of T1 and T4 tokens embedded in open syllables were larger than nasal syllables. Finally, the NH group's T2 -T1 also reported a negative and significant effect in the open rhyme conditions, indicating that NH people's T2 tokens experienced apparently fewer variations when they were embedded in open syllables.

5.2.3 The correlation analysis

The results of correlation analysis for each group and all data are displayed in Table 5.14. From Table 5.14, three types of relations could be found: the correlations between every two global parameters, the correlations between the two assessment parameters, and the correlations between an assessment parameter and a global parameter.

Table 5.14: The results of correlation analysis

	Duration	MeanF0	Slope	Curve	MeanZ	SDZ
Duration	—					
MeanF0	0.104^{***}	—				
Slope	0.210^{***}	$-.158^{***}$	—			
Curve	0.018	$-.307^{***}$	0.270^{***}	—		
MeanZ	$-.098^{***}$	0.033^{*}	$-.066^{***}$	0.106^{***}	—	
SDZ	$0.030^{\cdot}$	-0.126^{***}	0.114^{***}	0.137^{***}	$-.105^{***}$	—

Note: $^{\cdot}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

A detailed explanation of the results displayed in Table 5.14 was as follows:

Firstly, except for the case of Curve vs Duration which reported no significant correlation, all the other global parameters were significantly correlated with each other. In particular, Curve and MeanF0 were mediumly correlated with each other (Pearson's r = -.307), indicating that they were more correlated than the other pairs. Furthermore, while Duration only had a higher Pearson's r value with Slope, every two parabola parameters reported comparably higher Pearson's r values. That is, the parabola parameters might be closer to each other than with Duration.

Secondly, the two subjective assessment parameters were negatively and weakly correlated with high significance, showing that the higher the mean score a given token had received, the larger variations the scores were among the judges. That is, the judges have more similar opinions over low-score tokens than high-score tokens.

Finally, as for the correlations between the assessment parameters and the global parameters, all pairs were weakly correlated with high significance except for the case of SDZ vs Duration. It is noted that, compared with the correlations between MeanZ and the global parameters, the correlations between SDZ and the global parameters reported higher Pearson's r values, indicating that SDZ might have closer relations with the global parameters than MeanZ. More important, among the correlations between SDZ and the global parameters, SDZ vs Curve reported the highest coefficient of association, indicating Curve was most influential in subjective assessment.

5.3 Discussion

Based on the analysis to the results of the two experiments, we might answer the three questions advanced at the beginning of the chapter as follows.

5.3.1 What acoustic characteristics do the deaf participants have in Mandarin tone production?

In order to analyze the acoustic characteristics of deaf people's Mandarin tone production, the current chapter extracted two groups of parameters—the global parameters and dynamic tone parameters—from the data, and applied statistical analysis respectively.

The statistic results for the global parameters reported the following findings. Firstly, Tone reported high significance with the largest F values among the factors, proving that all four global parameters are good indicators for distinguishing Mandarin tones. Secondly, syllable types also played an important role in the production of Mandarin tones, in which simple and open syllables had shorter tone durations, lower mean F0 values, more gentle slopes, and smaller curves than nasal syllables. Thirdly, besides the result that female participants had higher F0 values, this study also found that the female participants had steeper slopes, especially in T3 and T4, than their male counterparts. Finally, compared with other factors, it seemed that the Group had no significant impact on the other global parameters except for the case of tone durations. That is, the global parameters such as mean F0, slope, and curve of the tones were not affected by whether the participants were deaf or not. As for the case of Duration, the statistical results indicated that tone tokens produced by deaf people were significantly longer than those produced by normal hearing people.

The analysis for dynamic tone parameters revealed the following facts. On one side, both CI people and HA people's tone patterns were quite similar to their NH counterparts, indicating that the deaf participants, regardless of their genders and what device they use, could produce Mandarin citation tones that reach or approach the level of normal hearing people. On the other side, the tone registers of the deaf groups were much more compressed when compared with that of the NH group, indicating that it would be harder to distinguish the tones produced by the deaf participants because of the subtle register differences among the four tones. More importantly, the generalized additive mixed model analysis for each tone reported the following results. First, as further evidence for the two above-mentioned points, the generalized additive mixed model analysis revealed that the two deaf groups' four tone shapes were generally similar to the NH group's four tones respectively, but the tone shapes of the two deaf groups changed more slowly along the time dimension than that of their normal hearing counterparts. Second, while no significant difference was found in T2 among the three groups, significance or marginal significance was found in the other tones especially T3 and T4, indicating that the deaf groups and the NH group were more different in the tone shapes of T3 and T4. Compared with the deaf groups' T3 and T4, the NH group's T3 had a higher starting point and a lower lowest point, and T4 had a higher starting point. Finally, except for the main effect of TimePoint on T2, significant main effects and interactions of TimePoint and Duration were found on all four tones, indicating the two factors are important impacting factors to the shapes of Mandarin tones. It is noted that the insignificant main effect of TonePoint on T2 might reveal that the slopes of T2 tokens were too small which led to too small F0 changes among the time dimension.

Overall, the current results supported the viewpoint that, similar to the NH group, both the CI and HA group had set up the citation tone patterns that could differentiate one tone from another one. However, compared with the NH group, the two deaf groups had more compressed tone registers that could be manifested by a lower starting point of T4 and a higher lowest-point of T3. Moreover, the current experiment also proved that the tone tokens produced by the two deaf groups had much longer tonal durations than that of the NH group. This characteristic, intertwined with the compressed tone register, made the tone tokens produced by the deaf participants have gentler slopes and smaller curves than the tokens produced by normal hearing people. Because of the above characteristics, although the listeners could distinguish the tones produced by the deaf participants, they found the deaf adults' tone tokens were harder to process than that of normal hearing people. From these results, we could predict that the tone tokens produced by the deaf adults would receive lower scores in the subjective evaluation of the tones.

5.3.2 Can subjective assessment with multiple judges tell the difference of the three groups' production qualities of Mandarin tones?

In the past, the method of subjective assessment was frequently used to evaluate the performance of the participants in the production of Mandarin tones. In these studies, researchers often invited one to three judges who were professional phoneticians in Standard Chinese to fulfill the assessment task. To make the judges' performance more consistent, some of these experiments would adopt some coordination mechanisms such as the pre-evaluation training for the judge, the consistency test of assessment result, and the third party arbitration to the controversy. Nevertheless, limited by the number of judges, the results of these experiments could not minimize the effects of the judge's subjective consciousness to the lowest level.[6]

With this in mind, the current study developed a subjective assessment method based on multiple judges. By adopting this method, we could limit the effects of subjective consciousness in two ways. First, this study did not try to interfere with or correct the judges' judgments but relied on their own intuitions of the Mandarin sound system to fulfill the assessment task. Thus, we could avoid the effects of subjective consciousness caused by the coordination mechanisms. Second, this study transformed the original scores into Z-score values for every judge's data and then calculated the mean Z-score value for each token. This way, the assigned scores among the judges could be more consistent and the

---

[6] Since we had no original assessment data of these studies, we could not say whether inter-rater reliability rates were low or not. However, considering there were only one to three judges involved in these studies, it would be hard to remove the effect of judge's subjective consciousness.

effects of the judges' subjective consciousness could be minimized. Therefore, theoretically speaking, the subjective assessment method developed in the current study could tell the difference in Mandarin tone production among the three participant groups in a more reliable and accurate way.

The subjective assessment developed in the current study revealed that the NH group received significantly higher assessment scores than the two deaf groups, revealing that tone tokens produced by the NH group have much better qualities than that of the deaf groups. Among the four tones, T3 hit the lowest, T2 came in second lowest, and T1 and T4 had similarly the highest average Z-scores among the four tones for all three participant groups. In particular, the results of subjective assessment reported that T2 and T3 produced by deaf adults are harder to perceive than T1 and T4. That is, in line with our previous studies (Chen et al., 2016a, 2017f; Hou et al., 2019), this study provided more evidence that T3 and T2 are harder to produce than T1 and T4 for prelingually deaf adults. Interestingly, this study also revealed that T3 tokens produced by normal hearing people received the lowest evaluation scores among the scores, which might indicate that the T3 was harder to produce than other tones even for normal hearing people, or the production qualities of T3 tokens were more unstable than other tones for this participant group. In contrast, although the NH group's T2 also received a lower score than T1 and T4, the difference for this group was much smaller than the two deaf groups, which might indicate that T2 was not that hard for the normal hearing people but much harder for the deaf people.

It is noted that, different from the subjective assessment method adopted by the previous studies, the subjective assessment developed in the current study had advanced another parameter SDZ. With this parameter, we could see the stabilities of the assigned scores among the judges to the tone tokens. In this study, we found the NH group's T1 and T4 SDZ values were much lower than their T2 and T3 values, proving again that T2 and T3 are harder to produce than T1 and T4, or the production qualities of the former two tones were more unstable than the latter two tones for normal hearing adults. For the two deaf groups, the statistic results of SDZ values revealed that their four tones had higher SDZ values in all three types of rhymes than the NH group's four tones, which might indicate that the production qualities of all four tones were not stable for the deaf participants. Moreover, while no significant effect was found, it seemed that the CI group had lower MeanZ values but stabler SDZ values than the HA group, which might reveal that the CI group's production qualities of the four tones were a little but consistently poorer than that of the NH group.

In general, the answer to this second question, namely whether the subjective assessment by multiple judges can tell the difference of the three groups' production qualities of Mandarin tones, would be yes. Overall, the NH group's tone tokens received significantly higher scores than the two deaf groups, indicating that the NH group's tone production quality was better than the deaf groups. On the basis of the results of the current study, we provided more evidence that T3 and T2 were harder tones both for normal hearing and prelingually deaf adults. Moreover, we could also discover some subtle differences among the three groups in the production of Mandarin tones. Thus, with this method, we were not only able to reduce the effects of subjective consciousness that were unavoidable in the past studies, but also have a better understanding of the difference in Mandarin tone production among the three participant groups.

5.3.3 Whether the results of acoustic analysis are correlated with the results of subjective assessment?

To better understand the differences between prelingually deaf adults and normal hearing adults in the production of Mandarin tones, this study carried out a three-step analysis. The first step was to analyze the acoustic characteristics of Mandarin tones produced by the three participant groups. The second step was to subjectively assess the Mandarin tones produced by the three participant groups. The third step was to check the correlation between the acoustic characteristics and the subjective assessment results.[7]

In the first step, we adopted two types of parameters, the global parameters including the commonly used parameter tonal duration and the parabola parameters, and the dynamic tone parameters containing 11 F0 values equally distributed along the time dimension for each token. The results of the first step analysis revealed that, although the deaf groups and the NH group had some differences in tonal durations and tone shapes, the parabola parameters of the Mandarin tones produced by the three groups were almost the same. In particular, no significant difference was found in all parameters between the CI group and the HA group, indicating that the two deaf groups had the same tone production qualities.

---

[7] In the current study, we did not just focus on the correlations between the acoustic parameter and each of the two subjective ones but studied the correlations for all pairs of parameters. By checking the correlations between the acoustic parameters, we could see the relations among the global acoustical parameters are not very close. That is, from another angle, these results showed that these global parameters might not be good parameters for checking the deaf participant's tone production.

In the second step, we developed a new method of subjective assessment method to minimize the effects of subjective consciousness. In line with the results of the first step analysis, the results of the second step analysis indicated that, while no significant difference was found between the CI group and the HA group, the two deaf groups' data similarly received lower scores than that of the NH group in the subjective assessment of the tones. Meanwhile, the results of the second step analysis revealed that T2 and T3 (especially T3) produced by the deaf groups' are harder to perceive than T1 and T4 produced by them. These results, again, were consistent with the results of the first step analysis. Namely, the deaf groups' tone production qualities were not as good as that of the NH group; among the four tones produced by the deaf groups, the production qualities of T2 and T3 were not as good as those of T1 and T4.

In the third step, we checked the correlation between the acoustic characteristics and the subjective assessment results. If we just checked the correlation analysis among the global parameters and the subjective assessment parameters, the answer to this third question would be like this: yes, tone production qualities were correlated with the subjective assessment results, but the correlations were very weak. However, if we directly compared the results of the first two steps, both of them were revealing that the deaf groups' tone production qualities were similarly not as good as that of the NH group. That is, the results of the acoustic analysis and the results of the subjective assessment could be closely related to each other.

Thus, an interesting question could be advanced here: what caused the weak agreement between the direct impression and the correlation analysis results? In our opinion, the possible answer could be that the global parameters might not have the capability to acoustically represent the complete picture of tone production qualities. Considering a tone token was realized by consistently changing F0 values along the time direction, the global parameters that omitted the detailed tonal changes might fail to completely capture the tone production qualities. Furthermore, while both the analysis for the dynamic tone parameters and the subjective assessment results were revealing significant differences between the deaf groups and the NH group, the analysis of the global parameters failed to report any significant difference among the participant groups except for the case of Duration. Taken together, the above facts might indicate that the global parameters are not good at telling the quality of tone production. As a result, when putting the poorly performed and well-performed parameters together, it would not be surprising that the correlations between the global parameters and the subjective assessment parameters were weak or even nonsignificant.

5.4 Conclusion

The findings of the current chapter confirmed that all three participant groups could produce Mandarin citation tones quite well. In particular, the acoustic analysis revealed that the two deaf groups' data had similar performance in the global parameters to their normal hearing counterparts. The only differences between deaf people and normal hearing people were the former had longer tonal durations and more compressed tone registers. As a result, compared with the NH participants, the deaf participants' four tones had slower changes along the time dimension. Perhaps because of this, deaf people's four tones were easier to confuse with each other than that of normal hearing people.

To objectively assess the differences in tone production qualities among the three participant groups, we developed a new method of subjective assessment based on multiple judges. The results, being consistent with the acoustic analysis, showed that the four tones produced by the three groups could be distinguished by the judges who were lay people in pathologic speech evaluation. Moreover, the results of subjective assessment confirmed the correctness of the results of the acoustic analysis from two angles: from the angle of participant groups, the tone tokens produced by the two deaf groups received lower assessment scores than the NH group, indicating that the two deaf groups similarly performed worse than the NH group in the production of Mandarin tones; from the angle of Mandarin tones, T1 and T4 tokens produced by all three groups received higher assessment scores than T2 and T3 tokens, indicating that T2 and T3 were harder to produce than T1 and T4 for all participants.

Generally speaking, the current chapter confirmed the viewpoint that the prelingually deaf adults, a special group of second language learners of Standard Chinese, had similar difficulties in Mandarin tone production as the native speakers and other second language learners targeted in the Standard Chinese: T2 and T3 were harder to produce than T1 and T4. Moreover, the current study displayed the importance of dynamic tone parameters in describing the production quality of Mandarin tones. In this study, although the global parameters, especially the parabola parameters proved to be valuable in differentiating the four Mandarin tones, they failed to report any difference among the three participant groups. In contrast, the dynamic tone parameters were not only good at differentiating the four

Mandarin tones but also had better performance in telling the difference among the participant groups. Thus, the current study showed that the dynamic tone parameters were a good method to study the detailed difference in the production quality of Mandarin tones. Finally, the current chapter might indicate that the subjective assessment method was not only able to reduce the effects of subjective consciousness but also had a better understanding of the differences in tone production qualities among the three participant groups.

It is noted that this study failed to find any strong correlation between the global parameters and the subjective assessment parameters. We suspected that the reason might lay in the global parameters not being able to represent the complete picture of tone production qualities because so much of tone's characteristics have to do with more micro-level pitch fluctuations.[8] In the forthcoming research, we will try to find some more acoustic parameters that can better represent the production quality of Mandarin tones, and continue to study the correlation between the global parameters and the subjective assessment parameters. After fulfilling that task, we might be able to develop a robust automated assessment system to evaluate the speech quality of Standard Chinese.

---

[8] It is noted that the existing literature on L2 tone perception that L2 learners like English listeners focus on tone global cues (e.g. mean F0) rather than dynamic ones (Gandour, 1983; Yang, 2015; Tupper et al., 2020)). That is, the results of English listeners might report a good correlation with global parameters and a bad correlation with dynamic parameters. In the future, we could use the current methodology to look into this more, and see whether English listeners' perception correlates to global parameters rather than dynamic ones.

Chapter 6

General discussion and conclusion

After summarizing the main findings of the dissertation, this chapter will provide some general discussions to answer the general questions advanced in Table 1.2.3, namely the PDA's performance in perceiving and producing Mandarin tones, the impact of different factors on PDA's production and perception of Mandarin tones, and the underlying mechanism of PDA's Mandarin tone production and perception. Based on these discussions, we will explore the implications of the current findings to prelingually deaf people's development of Mandarin tones. Finally, before concluding with a summary of the dissertation, the chapter will also discuss the limitations of the study and suggest directions for future research.

6.1 The main findings of the dissertation

By adopting three experiments—the synthesized tone perception experiment focusing on the impacts of pure acoustic cues on synthesized tone perception, the lexical tone perception experiment checking the participants' performance in identifying real human-produced Mandarin tones, and the tone production experiment investigating the participants' productions of citation tones, this dissertation investigated prelingually deaf adult's performance in perceiving and producing Mandarin citation tones. The main findings of the experiments were as follows.

- The results of the synthesized tone perception experiment showed that normal hearing people performed much better than deaf people in tone categorical perception of Standard Chinese. That is, while about half of the NH participants categorically perceived all four synthesized tones, only a few deaf participants could categorically perceive the six tone pairs in Standard Chinese. In particular, the results also showed that the CI participants experienced significantly more difficulties than the HA participants in categorically perceiving the synthesized tones. Moreover, while both frequency and duration played important roles in the NH group's categorical perception of synthesized Mandarin tones, the deaf participants' performance in perceiving synthesized tones was more affected by frequency than duration. Finally, the results of the synthesized tone perception experiment revealed that T2T4 and T2T3 were harder to categorically identify and discriminate than other tone pairs for the deaf participants.

- The results of the lexical tone perception experiment reported that all participant groups could categorically identify Mandarin lexical tones. Inside the three participant groups, the NH group performed better than the deaf groups, and the two deaf groups performed similarly in this experiment. Moreover, the current experiment also showed that the deaf groups' lexical tone perception was affected by rhyme complexity because the tones are embedded in syllable rhymes. That is, compared with syllables with nasal rhymes, the deaf participants had higher accuracies in processing syllables with open and simple rhymes. Finally, the current results also showed that T2T3 and T2T4 were significantly harder to categorically identify than other tone pairs for deaf people.

- The results of the tone production experiment revealed a complex pattern of tone production performance among the three participant groups. In general, all three groups could produce all four Mandarin tones that differentiate from each other. However, the NH group performed much better than the two deaf groups revealed by the results of statistical analysis on tone dynamic parameters and subjective assessment parameters. Moreover, the current results also showed that the deaf participants' performance in lexical tone production was affected by syllable complexity, in which syllables with nasal rhymes were much harder to produce than syllables with open and simple rhymes. Finally, the current results also showed that T2 and T3 (especially T3) were harder to produce than the other two tones for the deaf participants.

Taken together, three points could be summarized from these results. Firstly, NH people performed better than deaf people in all experiments; inside the deaf group, while the CI participants performed worse than the HA participants in the synthesized tone perception experiment, they had similar or even better performance in the remaining experiments when compared with their HA counterparts. Secondly, the results of the three experiments also showed that T2 and T3 (especially T3) were harder to perceive and produce than the other two tones for the deaf participants. Finally, the results of the three experiments showed that the deaf participants' performance in perceiving and producing Mandarin tones was affected by syllable complexity. That is, the deaf participants had higher accuracies in processing syllables with open and simple rhymes than syllables with nasal rhymes.

6.2 General discussion

Now, it is the time to answer the general questions advanced in Section 1.2.3, namely the PDA's performance in perceiving and producing Mandarin tones, the impact of different factors (tones themselves, rhyme types, and hearing devices) on PDA's production and perception of Mandarin tones, and the underlying mechanism responsible for the challenges to PDA's Mandarin tone production and perception.

6.2.1 The PDA's performance in perceiving and producing Mandarin tones

When focusing on the participants' performance in the three experiments, one could see the three groups have two main points in common:

First, all three groups perform much better in perceiving the lexical tones than the synthesized tones. In the synthesized tone perception tasks, the NH participants perform quite well in the synthesized tone identification tasks but much worse in the synthesized tone discrimination tasks; meanwhile, some deaf participants could categorically identify tones and only a few of them could categorically discriminate tones. That is to say, not too many participants, including the NH participants, could categorically perceive synthesized Mandarin tones. In the lexical tone perception tasks, as displayed in Table 4.5, the NH participants only judge 1.67% trials incorrectly in the lexical tone perception tasks, indicating that normal hearing people have no problems in categorically identifying the lexical tones no matter what syllabic context they are in; meanwhile, both of the two deaf groups identify more than 66.7% trials correctly in general, showing that they could identify the lexical tones in general, no matter what syllabic context these tones are in. Taken together, all deaf participants perform much worse in the synthesized tone identification tasks than in the lexical tone identification tasks, which might indicate that the deaf participants could also take advantage of the syllabic context to categorically identify Mandarin lexical tones. Especially, although only a few participants in the CI group could categorically perceive the synthesized tones, the CI group almost reach the same level as the HA group in the lexical tone perception task, revealing that CI people are good at using the syllabic context to identify the lexical tones.

Second, although the two deaf groups perform worse than the NH group in all three experiments, the differences among the participant groups in lexical tone perception and production are much smaller than in synthesized tone perception. Besides the above-mentioned differences in the performance of the three groups in synthesized tone perception and lexical tone perception, the current study also report the differences among the three groups in lexical tone production tasks. That is, although the deaf groups' tone ranges are more compressed than that of the NH group, all three groups have set up Mandarin tone patterns that distinguish the four lexical tones from each other, and the deaf groups have built up Mandarin tone patterns that are similar to that of the NH group. As a result, although the deaf groups are significantly different from the NH group in three of the four tone shapes in a fine-grained way, the three groups' tone shapes report no significant difference almost in all global acoustic parameters. Compared with the results of synthesized and lexical tone perception, one might see that the participants (especially the deaf participants) perform better in lexical tone production than in lexical tone perception, and much better than in synthesized tone perception.

From the above results, one could notice that there are loose links between synthesized tone perception and lexical tone perception and production. In the synthesized tone perception experiment, T2T3 and T2T4 are the hardest tone pairs for the two deaf groups. Similarly, we find that the two pairs are also harder for the two deaf groups in the lexical tone perception experiment. Furthermore, we also find that T2 and T3 trials produced by the two deaf groups have received lower scores than T1 and T4 trials in the lexical tone production experiment. From these facts, we could infer that there are some connections between synthesized tone perception and lexical tone perception and production: if the pure and fine-grained acoustic features of some tones are harder to process, these tones are also harder to perceive and produce in the lexical environment.

Compared with their links to the synthesized tone perception, lexical tone perception and production are much more stronger linked with each other. In the current study, we find that the participants in all three groups perform much better in lexical tone perception and production than in the synthesized tone perception. In particular, although the deaf participants still perform much worse than the NH participants, both deaf groups have developed the tone patterns that are similar to that of the NH group. Thus, these results might suggest that, no matter what hearing device the deaf participants use, they could perceive and produce Mandarin lexical tones well, indicating is a strong connection between lexical tone perception and production.

To go a step further, the above results lead to two inferences. First, the perception of Mandarin tones is at most quasi-categorical for Mandarin speakers. As we briefly reviewed in Section 2.1.2, many existing studies have investigated different populations' categorical perceptions of Mandarin tones. These studies, as we mentioned in 3.3, often do not overtly discuss the interactions between identification and discrimination. However, from the results of these studies, one could see that the participants perform much better in identifying Mandarin tones than in discriminating the tones. For example, in a study using the Categorical-Perception Index (CP index) advanced by Van Hessen and Schouten (1999) to investigate the categorical perception of T1 and T4 in Standard Chinese (Liu et al., 2017), Liu et al. (2017) reported that the mean CP index score of T1T4 was 48.075 out of 100, indicating that the perception of T1T4 are approaching quasi-categorical for normal hearing people who speak Mandarin as their first language. Since limited cases following our standards of tone categorical perception are found in all six tone pairs, we do not involve the calculation of CP index score in the current study. Although almost all tone pairs could be categorically perceived by some participants in all three groups, only four tone pairs (T1T2, T1T3, T1T4, and T2T4) are categorically perceived by half or more of the NH participants. Therefore, our results disclose an apparent fact that the perception of Mandarin tones is at most quasi-categorical in some tone pairs for normal hearing people, and might be non-categorical for prelingually deaf adults.

Intertwined with the first inference, a more important inference that could be drawn from the current results is that pure and fine-grained acoustic features of tones might not be necessary for deaf and normal hearing people to perceive and produce Mandarin lexical tones. In this study, all three groups perform not quite well in the synthesized tone perception tasks. Taking T2T3 as an example, only six participants in the NH group could categorically perceive the synthesized T2T3, not to mention the participants in the CI group and HA group, showing that it is hard for our participants (including the normal hearing participants) to categorically perceive this tone pair only relying on fine-grained acoustic features. Nevertheless, all three groups perform much better in lexical tone perception and production. It is noted that, among the three groups, the CI group's performance has improved the most in the lexical tone perception and production experiments. Considering that the CI device is not good at processing tonal information directly (See Section 6.2.2 in details), it is reasonable to infer that the CI participants' improved performance in lexical tone perception is mainly because they are good at taking advantage of the syllabic context. More important, the CI group perform even better in lexical tone production, revealing that they could produce a tone well under the situation that they could not perceive the acoustic features of the tone directly. Taken together, we assume that the participants don't need to rely on fine-grained acoustic features of tones to categorically perceive and produce Mandarin lexical tones.

6.2.2 The impact of different factors on PDA's perception and production

6.2.2.1 Tones

Similar to the populations—NHC, NNS, and PDC—that we reviewed in Chapter 2, one might notice that the deaf groups (and the NH group) experience more difficulties in perceiving and producing T2 and T3 than the other two tones. That is, the two deaf groups perform worse in T2 and T3 trials than in T1 and T4 trials in all three experiments. In the synthesized tone perception tasks, the two groups' tone identification and discrimination in T2T3 and T2T4 are much harder than the other tone pairs. In the lexical tone perception tasks, T2T3 and T2T4 are much harder to process: T2T3 trials took the deaf participants' longest times to make the choices, and T2T4 trials hit the highest incorrect rate among the six tone pairs. Considering T4 is an easy tone to process, one could infer that the reason of the difficulties in processing T2T4 were mainly from T2. From these facts, we could see that the two deaf groups faced more difficulties in perceiving T2 and T3. In the lexical tone production tasks, while T2 and T3 trials received much lower assessment scores than T1 and T4 trials for all three groups, the two deaf groups' T2 and T3 trials received much lower scores than that of the NH group, indicating that T2 and T3 are especially difficult to realize for the two deaf groups. Overall, similar to the populations—NHC, NNS, and PDC—that we reviewed in Chapter 2, the deaf groups (and the NH group) experience more difficulties in perceiving and producing T2 and T3 than the other two tones.

In our opinion, the differences in the acoustic features of T2 and T3 make them harder to perceive and produce than the other two tones. For one thing, the differences in acoustic salience among the four tones might decide how difficult they are in these populations' tone perception. In Mandarin Chinese, T1 is a high tone (Chao system's 55 or 44) that makes it a very salient tone. Moreover, T4 starts with the highest pitch values (Chao's 5) and ends with a mid-pitch value (Chao system's 3) in most situations, which makes it also a salient tone. In contrast, although T3 has the longest duration, it has the lowest pitch values. Considering duration may not play an important role in tone perception for the deaf participants[1], the insensitivity of their hearing devices to low tone pitch might make T3 less salient in perception. As for the case of T2, since people often realize it as mid-tones (Chao system's 23 or 24)(Xu and Shi, 2022), it may not be quite salient in perception. For another, the differences in acoustic salience also make T2 and T3 harder to produce than the other two tones. Compared with realizing a level or falling tone, these populations have more problems in producing a rising tone or a dipping tone, which could lead to more problems in producing T2 and T3 than T1 and T4.

6.2.2.2 Rhyme types

From the experiment results, one could see that the two deaf groups' performance in lexical tone perception and production are impacted by the complexity of Mandarin rhymes. That is, tones embedded in nasal rhymes are much harder to perceive and produce than tones embedded in simple and open rhymes.

In our opinion, there are two intertwined reasons for this phenomenon. For one thing, the acoustic characteristics of rhymes might make the tones embedded in some rhymes harder to perceive and produce than the tones embedded in other rhymes. In a syllable with a nasal rhyme, the loudness in the low-frequency zone is low because of the anti-

---

[1] At present, although limited studies have reported that duration impacts Mandarin tone perception by PDC (Chen and Liu, 2010), and elder native Chinese listeners (Wang et al., 2017b), there is no research directly investigating the role of duration in NNS's lexical tone learning (Riestenberg, 2017). In the current study, we also find that tonal duration has no effect on deaf participants' tone perception. From these facts, we infer that duration may not play an important role in lexical tone perception.

resonance from the nasal sound. Moreover, because of co-articulation, the loss of loudness spreads over the whole rhyme that carries the tone. Closely related to the differences in acoustic prominence, the second possible reason is that the syllables with nasal rhymes have weaker mental representations than syllables with open and simple rhymes. Because deaf people have more problems in perceiving nasal rhymes, they should be more unfamiliar with them and therefore have less cognitive resources to process tones embedded in nasal rhymes. As a result, the deaf participants would find it hard to identify the tones embedded in syllables with nasal rhymes. Although the impact of rhyme type on tone production is not as big as it is on tone perception, rhyme type also has a big impact on the deaf participants' lexical tone production. From the current results, producing a tone embedded in a nasal syllable is also harder than it is in a simple or open rhyme for the deaf participants, indicating the complexity of Mandarin rhymes also influences the deaf participants' lexical tone production.

6.2.2.3 Hearing devices

The experiment results reveal a complex scenario of the impact of hearing devices on PDA's production and perception of Mandarin tones.

On one hand, the results of the synthesized tone perception experiment might reveal that the HA device has some advantages in processing fine-grained tone information over the CI device. Confined by the experimental design in the synthesized tone perception experiment, the participants have no other clue to rely on to categorically identify and discriminate the stimuli except for the acoustic information. More importantly, all the stimuli in this experiment are well controlled to ensure that the participants fulfill the experimental tasks only depending on the differences in fundamental frequencies and/or durations among the involved stimuli. Therefore, when the two deaf groups perform differently in this experiment, it is probably due to the different performances of the hearing devices they are using. In the synthesized tone perception experiment, since the results show that the CI participants have a harder time than the HA participants in categorically perceiving the synthesized Mandarin tones, it is reasonable to assume that this is because of the differences in pitch processing between the devices.

On the other hand, the results of the lexical tone experiments might indicate that the CI participants have benefited more from the hearing devices than the HA participants.

Considering the HA group are better at original hearing abilities than the CI group (see 3.1.1 for details), it would not be surprising to predict that the HA participants would perform better than the CI participants in perceiving and producing Mandarin tones across the board. However, we fail to find this phenomenon in the two lexical tone experiments. Especially, among the 20 CI participants in the current study, 12 of them are original HA users who received cochlear implantation a few years later because of the inefficiency of the original hearing devices.[2] Fortunately, benefiting from their new devices, their performance in perceiving and producing Mandarin tones have reached a similar level to the early and constant HA participants. Thus, it would be reasonable to speculate that the CI group might benefit more from their hearing devices than the HA group.

Considering the HA device prevails over the CI device in processing tone information (See Section 2.2.3 for more details), it would be interesting to advance an interesting question: why do the CI participants benefit more than the HA participants from their hearing devices, especially given that they have more severe hearing loss and that their hearing devices have less capability in processing tone information (both frequency and duration information) than that of the HA participants. In our opinion, this phenomenon might be explained by the fact that the CI device has apparent technical advantages over the HA device.

As a relatively cheaper device for reconstructing hearing, HA is a common tool for solving mild to moderate sensorineural hearing loss and conductive hearing loss. For those who suffer severe-to-profound sensorineural hearing loss, HA is ineffective because this kind of hearing loss often means serious dysfunction in the inner ear. Even for those who suffer mild to moderate sensorineural hearing loss, clear hearing is still a difficult target after fitting HA. As summed up by Kim and Barrs (2006), sensorineural hearing loss has three characteristics that restrict the effectiveness of HA in compensating for hearing loss. First, sensorineural hearing loss is often characterized by a downsloping audiogram and the loudness of speech comes mainly from the low frequencies, so wearing a HA device doesn't ensure deaf people understand a loud-enough speech sound with a "blank" high-frequency-zone. Second, sensorineural hearing loss makes a specific frequency not sharply defined by the human cochlear tuning curves, which results in the inability of the human cochlea to differentiate speech sound from noise close in frequency. Third, the decreased dynamic

---

[2] It is noted that a combination of HA and CI might give deaf people extra abilities and make them perform better in tone perception and production (Zhang et al., 2020a,b). Thus, we could not totally reject the idea these CI participants might benefit from their early experience of HA using to perform better in the lexical tone tasks.

range caused by sensorineural hearing loss is also a difficult issue for a HA fitting. With a decreased dynamic range, amplification cannot be linear: when soft sounds are amplified to the hearing threshold, medium and loud sounds may be too loud. With these shortcomings, many people with hearing loss refuse to use the HA device. For instance, in the United States, 314 in 1000 individuals who are older than 65 years of age are hearing impaired, and only 1 in 5 individuals who stand to benefit from hearing aids acquires one; among those who do not wear a HA device, a quarter of them refuse to use it due to everyday background noise (Kochkin, 2001). In mainland China, many old people also refuse to use HA for similar reasons such as negative impressions of or misconceptions about the HA device (Zheng et al., 2022).

CI is a relatively expensive device for reconstructing hearing. However, because it is a more effective device than HA, cochlear implantation has become the standard treatment for children and adults with severe-to-profound sensorineural hearing loss nowadays (van Schoonhoven et al., 2013). Especially, recent years witnessed the rapid development of CI technology including electric-acoustic stimulation and/or bi-modal stimulation (O'Connor and O'Connor, 2010; Pillsbury III et al., 2018; Guo et al., 2019; Tejani et al., 2021), real-time noise suppression (Srinivasan et al., 2013; Dingemanse and Goedegebure, 2018; Michels et al., 2022), and tripolar electrode stimulation (Bierer, 2007; Bierer and Faulkner, 2010; Bierer et al., 2015; Arenberg et al., 2018). With the development of these technologies, the current CI devices have better performance in enhancing speech sounds and counteracting noise. For example, the application of electric-acoustic stimulation and/or bi-modal stimulation can reduce the surgical damage to the cochlea, so it can make up for the loss of high-frequency hearing and preserve the original low-frequency hearing simultaneously (O'Connor and O'Connor, 2010). Because deaf people who received electric-acoustic stimulation and/or bi-modal stimulation could perceive the low-frequency signal using their residual hearings, they have better performance in perceiving speech from a noisy environment than conventional CI users (Rader et al., 2013; Pillsbury III et al., 2018; Park et al., 2019; Guo et al., 2019; Tejani et al., 2021).

Besides the fact that CI has more technical advantages than HA in providing hearing compensation for people who suffered severe-to-profound sensorineural hearing loss, other factors like a more positive attitude towards CI (Mäki-Torkko et al., 2015; Harris et al., 2016; Kobosko et al., 2018), better family and social support (Wang, 2016; Majorano et al., 2020), and even comfortable and cosmetic considerations (Mitchell-Innes et al., 2017; Rapport et al., 2022) might also contribute to the superiority of CI over HA. As a result, although CI children at the age of 3 to 5 years old might perform poorer in speech recognition tasks than their HA counterparts (Liu et al., 2011), according to Wu (2007) and Bond et al. (2009)'s results, it would be reasonable to predict that these CI children would perform better or at least not worse than their HA counterparts in speech production and perception tasks a few years later.

In line with Wu (2007) and Bond et al. (2009), our results show that the CI participants with more severe hearing loss have similar performance to the HA participants with less severe hearing loss in Mandarin lexical tone production and perception tasks. That is, the CI participants have benefited more from their hearing devices than the HA participants in real speech situations. Considering the CI participants still perform worse than the HA participants in directly processing tone information, we speculate they might benefit more from their hearing device by hearing higher-frequency information more clearly, better speech perception in noise, and higher device compliance. As time elapsed, the CI participants should gain enough experience in Mandarin speech production and perception, which help them to catch up with the HA participants in processing Mandarin tones. Nevertheless, as we said in 4.3.1, we could not tell the specific source of indirect information of tone at present. However, from the existing study results on perceptual cues of whispered Mandarin tones, the possible cues people should rely on including duration, intensity, formant, and spectral tilt when pitch is absent (Gao, 2002; Liu and Samuel, 2004; Kong and Zeng, 2006; Jiao and Xu, 2019; Xu et al., 2022). Thus, we suspected that the CI participants might collect extra tonal information from these clues to identify Lexical Mandarin tones. In the future, we would carry out more experiments to explore the possibly indirect cues on PDA's tone perception.

6.2.3 The underlying mechanism of PDA's Mandarin tone production and perception

Denes et al. (1993) proposed the speech chain theory that explains the procedures of speech processing as follows: first, speech production starts from the speaker's mind by extracting the linguistic forms (mental representations) of what the speaker want to say; next, the brain encodes the linguistic forms into instructions in the form of nervous impulses that are sent to the articulators; thus, the articulators generate speech sound waves (acoustic signals) that travel through the air between the speaker and the listener; after the speech sound waves activate the listener's hearing mechanism and produce nerve impulses that travel along the auditory nerve to the listener's brain, the acoustic signals are transformed into linguistic forms which should be decoded as the speaker's messages. Under the framework of the speech chain, the course of speech production and perception involves three levels (acoustic, physiological, and linguistic) and follows the order of "linguistic level→physiological level→acoustic level→physiological level→linguistic level". It is noted that the speech chain theory assumes "the relationship between a word and a particular sound wave, or between a word and a particular muscle movement or pattern of nerve impulse, is not unique" (Denes et al., 1993, p. 7). That is, in the same vein of many speech perception and production theories such as Motor theory (Liberman and Mattingly, 1985), Perception for action control theory (Schwartz et al., 2007), and Directions into velocities of articulators model (Guenther and Vladusich, 2012),[3] the speech chain theory implicitly suggests an active cognitive process that involves bottom-up processes, top-down processes, and their interactions in speech perception and production.

Based on the framework of the speech chain theory, we propose that the difficulties in speech perception and/or production are caused by deficits at any level along the speech processing path. For instance, many factors, including background noise and deteriorated sounds at the acoustic level, the disorders of speech organs and damages of motor and sensory nerves at the physiological level, and weak mental representation (due to scarce cognitive resource or unfamiliarity with the linguistic form) and L1-L2 interference at the linguistic level, could cause errors in speech production and perception. According to the speech chain theory, the deaf participants' challenges in Mandarin tone perception and production are caused by their deficits mainly at the physiological level which gives rise to deficits at the acoustic and linguistic levels directly or indirectly. Because of hearing loss (the deficits at the physiological level), deaf people have to resort to hearing devices to modulate and transmit speech sound signals. Compared with natural speech sound signals, sound signals modulated by hearing devices are much more scarce and more compressed, which not only makes deaf people harder to distinguish different categories and establish their mental representations (the deficits at the linguistic level), but also makes the speech sounds they produced are also deteriorated and harder to understand (the deficits at the acoustic level).

---

[3] Motor theory is a well-known theory which needs no thorough introduction; for more detailed introductions about the other two theories, please see my review chapter. In addition, for more introductions about other active cognitive process theories, please see Heald and Nusbaum (2014)'s review paper.

Using the speech chain theory, we can explain the deaf participants' challenges in tone perception and production as: because the deaf participants could not clearly hear the acoustic signals associated with the tones (the acoustic level), they showed weaker mental representations for these tones (the linguistic level), and thus experienced more difficulties in realizing them both in perception and production (the physiological level). More important, this framework can also explain the differences of the deaf participants' performance in the three experiments. The synthesized tone perception experiment is the hardest for the deaf participants. In this experiment, it is hard for the deaf participants to perceive to the nuances between the stimuli created by manipulating fine-grained acoustic correlates; meanwhile, also because of the subtle differences between the synthesized stimuli, the deaf participants cannot access extra cues or cognitive resources to resort to in discrimination and identification tasks. In the lexical tone perception experiment, the stimuli are natural speech sounds that contain abundant indirect cues for tone perception (duration, amplitude, overtones of rhymes, etc.), and they should be closer to the linguistic forms stored in the deaf participants' minds. Thus, the deaf participants' performance in this experiment is better than in the synthesized tone perception experiment. In the lexical tone production experiment, the deaf participants have even better performance than the lexical tone perception because of two reasons: first, from the angle of the speakers (the deaf participants), what they have to do is to extract the linguistic forms of the targets and realize the targets in their own ways, in which process they have more control with the tasks; from the angle of the listeners (the normal-hearing judges), these adult native speakers have abundant experience in processing Mandarin tones, which should make them more tolerant when making the judgments.

As for the other populations, namely NHC, MMS, NNS, and PDC, their difficulties in Mandarin tone production and perception could be accounted as follows: NHCs and PDCs are immature Mandarin users who have problems in all three levels, while PDCs have more serious problems due to hearing loss and device using; MSSs are mature Mandarin users who probably have no fault on the linguistic level, but have faults on physiological level (and acoustic level) due to age-related degeneration; NASs are unbalanced bilinguals who have deficits on the linguistic level because their L2 Tone mental representations are impeded their first languages (Bialystok et al., 2009; de Leeuw, 2014).[4] Overall, all pop-

---

[4] It is noted that, Different from NNSs who are typical L2 users, the PDAs in the current study are sign-spoken bi-modal L2 users (Emmorey et al., 2008). As a result, although PDA's L1 should interfere their L2 linguistic forms in the aspects of syntax and semantics, there should be no competition between L1 and L2's phonetic and phonological representions in PDAs' minds. That is, the PDAs are more like Mandarin

ulations' difficulties in Mandarin tone perception and production should be attributed to deficits in acoustic, physiological, and linguistic levels in speech perception and production.

6.3 The implications for PDA's spoken language development

The current results provide some valuable implications for PDA's spoken language development.

First, spoken language development is an extremely long and difficult journey for prelingually deaf people. In the current study, all the deaf participants had received hearing intervention and speech rehabilitation at the early years of their life. Benefited by their hearing devices, speech rehabilitation, and abundant experience of using spoken language, prelingually deaf people have developed hearing and speech ability to a certain extent. However, many years after using Mandarin as their first spoken language, their ability in lexical tone perception and production is still worse than that of normal hearing people. Considering lexical tone is a basic component of Mandarin Chinese, the current results might indicate that the development of spoken language is still a long and difficult journey for prelingually deaf people who take Mandarin as their first spoken language.

Second, after receiving proper hearing intervention and speech rehabilitation, prelingually deaf people with severe-to-profound sensorineural hearing loss can develop the ability of speech production and perception in some extent. Limited by the shortcomings of their hearing devices, the deaf participants in the current study still faced tremendous difficulties on categorical perception of synthesized tones. In particular, most of the CI participants in the current study did not develop the ability to categorically perceive synthesized tones, which fully exposed the limitation of the CI technology. Happily, all the deaf participants performed much better in lexical tone production and perception than in synthesized tone perception. That is, the deaf participants, especially the CI participants have developed the ability of speech production and perception with the help of their hearing devices. Benefited from long time of spoken language using, they could develop the

---

L1 speakers in Phonetics and Phonology but with delay in development which could impair their speech processing in linguistic level.

ability of speech production and perception to a certain extent that ensure that they can speak the language properly.

Last but not the least, cochlear implantation might be a good option for prelingually deaf people who suffer severe-to-profound sensorineural hearing loss. The previous studies have repeatedly revealed that the CI device is a useful tool of rebuilding hearing and developing spoken language for people who suffer severe-to-profound sensorineural hearing loss. At present, there is a consensus that early implantation leads to better output (Dettman and Dowell, 2010; Wie et al., 2020; Karltorp et al., 2020). Interestingly, since limited attention has been paid on prelingually deaf adults, whether a delayed implantation is beneficial for prelingually deaf adult's spoken language development is still an open question. In the current study, we found that the CI participants have reached the same level to the HA participants in Mandarin lexical tone perception and production after using spoken language for more than 10 years.[5] Considering the CI device has worse performance in processing tone information than the HA device, and the CI participants have worse original hearing loss than their HA counterparts, the current results might indicate that the CI device is still a useful tool for rebuilding hearing and developing spoken language for prelingually deaf adults with delayed implantation. Namely, it is never too late to receive cochlear implantation for prelingually deaf adults who suffer severe-to-profound sensorineural hearing loss.

6.4 Conclusion

In conclusion, the current study found some promising results on prelingually deaf adult's ability to perceive and produce Mandarin lexical tones. From the results of the current tasks and analytic technique, the deaf participants in the current study did not develop the ability to categorically perceive synthesized tones. However, benefiting from abundant experience using spoken language, as well as the syllabic context of the lexical tones, these deaf participants performed much better in the lexical tone perception experiment. More important, the deaf participants, especially the CI participants performed better in the lexical tone production tasks than in the lexical tone perception tasks. Thus, the current results might provide some new insights into relevant fields such as L2 tone acquisition and

---

[5] Using the CA and AUD information displayed in Table 3.1 and Table 3.2, we could see that all deaf participants have used spoken language more than 10 years.

production-perception link of Mandarin tones. That is, for L2 tone acquisition, the current results imply that L2 learners might not need to hear all fine-grained cues clearly to perceive Mandarin lexical tones correctly, and might not need to perceive lexical tones well before producing them correctly; for the production-perception link of Mandarin tones, the current results suggest that there must be some critical perceptual cues linking tone production and perception, and tone perception might not be the precondition for tone production.

It is noted that the current study only involved a few prelingually deaf participants, especially the CI participants. In the future, it is necessary to recruit more prelingually deaf adults that could be divided into an early-implanted group and a delayed-implanted group, and check their performance in Mandarin tone perception and production. Moreover, some production and perception studies on other basic components of Mandarin Chinese, such as vowel, consonant, and/or syllable, should be conducted to explore the relation between the development of L2 speech production and perception. In addition, it is also necessary to conduct some studies to compare prelingually deaf adult's performance in Chinese sign language (L1) and Mandarin (L2) to explore the impact of PDA's L1 on their L2 development. By doing these, we would be able to get a more comprehensive understanding of prelingually deaf adult's spoken language development, and provide a more scientific basis for prelingually deaf people's hearing reconstruction and speech rehabilitation.

Bibliography

Adunka, O., Roush, P., Grose, J., Macpherson, C., and Buchman, C. A. (2006). Monitoring of cochlear function during cochlear implantation. *The Laryngoscope*, 116(6):1017–1020. 30

Arenberg, J. G., Parkinson, W. S., Litvak, L., Chen, C., Kreft, H. A., and Oxenham, A. J. (2018). A dynamically focusing cochlear implant strategy can improve vowel identification in noise. *Ear and hearing*, 39(6):1136. 164

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, 67(1). 53, 92

Bent, T. (2005). *Perception and Production of Non-Native Prosodic Categories*. Dr. theis, Northwestern University. 24

Best, C. T. and Tyler, MD. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. Lang Exp. *Second Lang. Speech Learn*, 1334:1–47. 7

Bi, J., Wang, L., Dong, B., and Zhang, L. (2008). An experimental study on the ability of the tone recognition of speech by hearing-impaired children with cochlear implants and hearing aids. *Chinese Scientific Journal of Hearing and Speech Rehabilitation*, (02):40–42 (in Chinese). 26

Bialystok, E., Craik, F. I., Green, D. W., and Gollan, T. H. (2009). Bilingual minds. *Psychological science in the public interest*, 10(3):89–129. 167

Bierer, J. A. (2007). Threshold and channel interaction in cochlear implant users: Evaluation of the tripolar electrode configuration. *The Journal of the Acoustical Society of America*, 121(3):1642–1653. 164

Bierer, J. A., Deeks, J. M., Billig, A. J., and Carlyon, R. P. (2015). Comparison of signal and gap-detection thresholds for focused and broad cochlear implant electrode configurations. *Journal of the Association for Research in Otolaryngology*, 16(2):273–284. 164

Bierer, J. A. and Faulkner, K. F. (2010). Identifying cochlear implant channels with poor electrode-neuron interface: Partial tripolar, single-channel thresholds and psychophysical tuning curves. *Ear and hearing*, 31(2):247. 164

Blicher, D. L., Diehl, R. L., and Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, 18(1):37–49. 16

Boersma, P. and Weenink, D. (2021). Praat: Doing phonetics by computer. 42, 85, 117

Bond, M., Mealing, S., Anderson, R., Elston, J., Weiner, G., Taylor, R. S., Hoyle, M., Liu, Z., Price, A., and Stein, K. (2009). The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: A systematic review and economic model. 165

Burnham, D. and Jones, C. (2002). Categorical perception of lexical tone by tonal and non-tonal language speakers. In *Proceedings of the 9th International Conference on Speech Science and Technology*, pages 2–5. 24

Cao, W. (2010). Mandarin tone perception:A report on level F0 contours. *Studies of the Chinese Language*, (06):536–543+576 (in Chinese). 15

Cao, W., Yao, K., and Zhang, J. (2022). Effect of auditory deprivation time on hearing and speech rehabilitation of post-cochlear implant patients. *Journal of Hunan Normal University(Medical Sciences)*, 19(01):102–104 (in Chinese). 26

Caposecco, A., Hickson, L., and Pedley, K. (2012). Cochlear Implant Outcomes in Adults and Adolescents With Early-Onset Hearing Loss. *Ear & Hearing*, 33(2):209–220. 26

Casas-Tost, H. and Rovira-Esteva, S. (2015). Mapping Chinese language pedagogy from 1966 to 2013. A bibliometric study of the Journal of Chinese Language Teachers Association. *Journal of the Chinese Language Teachers Association*, 50(2):31–58. 23

Chang, F. (2021). An Experimental Study on Honghe Longwei Hani Tone. Master's thesis, Yunnan Normal University. 15

Chang, Y.-p., Chang, R. Y., Lin, C.-Y., and Luo, X. (2016). Mandarin tone and vowel recognition in cochlear implant users: Effects of talker variability and bimodal hearing. *Ear and Hearing*, 37(3):271. 17

Chang Liu, Can Xu, Yuxia Wang, Lilong Xu, Hui Zhang, and Xiaohu Yang (2021). Aging Effect on Mandarin Chinese Vowel and Tone Identification in Six-Talker Babble. *American Journal of Audiology*, 30:616–630. 22, 29

Chao, Y. (1933). Tone and Intonation in Chinese. *Bulletin of the Institute of History and Philology*, 4(3):121–134. 42

Chao, Y.-R. (1956). Tone, intonation, singsong, chanting, recitatives, tonal composition, and atonal composition in Chinese. In Halle, M., editor, *Mouton, The Hague*, pages 52–59. Mouton, The Hague. 13, 14

Chao, Y.-R. (1968). *A Grammar of Spoken Chinese*. University of California Press, Berkeley and Los Angeles. 13, 14, 42

Chen, F., Peng, G., Yan, N., and Wang, L. (2017a). The development of categorical perception of Mandarin tones in four- to seven-year-old children. *Journal of Child Language*, 44(6):1413–1434. 16, 21, 47, 49, 52, 80

Chen, L., Cai, Y., Wang, A., and Gu, C. (2017b). The Study of Speech Intelligibility for Deaf Adults and Older Children with Hearing Impairment. *Chinese Scientific Journal of Hearing and Speech Rehabilitation*, 15(05):397–400 (in Chinese). 15

Chen, Q. (1997). Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers Association*, 32(1):21–39. 24

Chen, S. (2020). Analysis of intonation and interrogative intonation in children aged 4-11: A comparative study of Miao and Chinese. Master's thesis, Minzu University of China. 21

Chen, S., Zhang, C., McCollum, A. G., and Wayland, R. (2017c). Statistical modelling of phonetic and phonologised perturbation effects in tonal and non-tonal languages. *Speech Communication*, 88:17–38. 14

Chen, S., Zhu, Y., and Wayland, R. (2017d). Effects of stimulus duration and vowel quality in cross-linguistic categorical perception of pitch directions. *PloS one*, 12(7):e0180656. 16

Chen, X. and Liu, H. (2010). Study on Tone Recognition Ability after Cochlear Implant Implantation in Patients with Prelingua Deafness. *Journal of Audiology and Speech Pathology*, 18(1):55–56 (in Chinese). 5, 25, 161

Chen, Y. (2006). An experimental research of the Scottish students' Manddarin tonal pitches in monosyllable. *Chinese Teaching in the World*, 47(02):99–109+3–4 (in Chinese). 23

Chen, Y., Chen, Y., Lin, H., Hou, J., Xing, Y., and Dang, J. (2017e). A study of high level tone in standard chinese produced by prelingually deaf adults. In *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 626–629. 27, 120

Chen, Y., Chen, Y., Zhang, J., Zhang, J., Lin, H., Wei, J., and Dang, J. (2016a). Mandarin citation tone patterns of prelingual Chinese deaf adults. In *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pages 1–4. 27, 150

Chen, Y., Hou, J., Xing, Y., Chen, Y., Lin, H., and Dang, J. (2017f). The acoustic characteristics of tone 3 in standard chinese produced by prelingually deaf adults. In *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 647–650. 27, 118, 150

Chen, Y., Huang, J., and Chen, Y. (2021). The subjective assessment system of pathological voice and its effect detection. *Journal of Tianjin University of Technology*, 37(6):22–26 (in Chinese). 120

Chen, Y. and Wong, L. L. (2017). Speech perception in Mandarin-speaking children with cochlear implants: A systematic review. *International Journal of Audiology*, 56(sup2):S7–S16. 25

Chen, Y., Wong, L. L., Zhu, S., and Xi, X. (2016b). Early speech perception in Mandarin-speaking children at one-year post cochlear implantation. *Research in Developmental Disabilities*, 49–50:1–12. 5

Cheng, Y.-Y. and Lee, C.-Y. (2018). The Development of Mismatch Responses to Mandarin Lexical Tone in 12- to 24-Month-Old Infants. *Frontiers in Psychology*, 9:448. 20

Chernecky, C. C. and Berger, B. J. (2013). *Laboratory Tests and Diagnostic Procedures*. Elsevier Health Sciences. 17

Chua, K.-W. and Freeman, J. B. (2022). Learning to judge a book by its cover: Rapid acquisition of facial stereotypes. *Journal of Experimental Social Psychology*, 98:104225. 86

Craddock, L., Cooper, H., Riley, A., and Wright, T. (2016). Cochlear implants for prelingually profoundly deaf adults. *Cochlear Implants International*, 17(sup1):26–30. 26, 27

Cui, L. (2011). *A Study on Tone Perception and Production in Mandarin-speaking Children with Cochlear Impants*. PhD thesis, East China Normal University, Shanghai. 25

Dai, P., Liu, X., Yu, F., Zhu, Q., Yuan, Y., Yang, S., Sun, Q., Yuan, H., WY, H. D., and Han, D. (2006). Molecular etiology of patients with nonsyndromic hearing loss from deaf-mute schools in 18 provinces of China. *Chin J Otol*, 4(1):1–5. 5

Dallaston, K. and Docherty, G. (2019). Estimating the prevalence of creaky voice: A fundamental frequency-based approach. In *Proceedings of the 19th International Congress of Phonetic Sciences*, pages 532–536. Australasian Speech Science and Technology Association Inc. 117

de Leeuw, E. (2014). Maturational constraints in bilingual speech. In Thomas, E. M. and Mennen, I., editors, *Advances in the Study of Bilingualism*, pages 25–37. Multilingual Matters, Clevedon. 167

de Sousa, A. F., Couto, M. I. V., and Martinho-Carvalho, A. C. (2018). Quality of life and cochlear implant: Results in adults with postlingual hearing loss. *Brazilian journal of otorhinolaryngology*, 84:494–499. 27

Denes, P. B., Denes, P., and Pinson, E. (1993). *The Speech Chain*. Macmillan. 32, 165, 166

Deroche, M. L. D., Lu, H.-P., Lin, Y.-S., Chatterjee, M., and Peng, S.-C. (2019). Processing of Acoustic Information in Lexical Tone Production and Perception by Pediatric Cochlear Implant Recipients. *Frontiers in Neuroscience*, 13:639. 2, 25

Dettman, S. and Dowell, R. (2010). Language acquisition and critical periods for children using cochlear implants. *The Oxford handbook of deaf studies, language, and education*, 2:331–342. 169

DeVore, G. R. (2017). Computing the ZScore and Centiles for Cross-sectional Analysis: A Practical Approach. *Journal of Ultrasound in Medicine*, 36(3):459–473. 121

Dhanasingh, A. and Jolly, C. (2017). An overview of cochlear implant electrode array designs. *Hearing Research*, 356. 30

Dillon, H. (2008). Introductory concepts. In Dillon, H., editor, *Hearing Aids*, pages 1–17. Boomerang Press, Turramurra (Australia), second edition. 30

Ding, H. (2020). Categorical Perception of Mandarin Tones 2 and 3 by L2 Learners with Different L1 Backgrounds. Master's thesis, Jiangsu University of Science and Technology. 23

Ding, H., Hoffmann, R., and Jokisch, O. (2011). An Investigation of Tone Perception and Production in German Learners of Mandarin. *Archives of Acoustics*, 36(3):509–518. 24

Ding, H. and Ji, X. (2020). A review of studies on Chinese tone acquisition by Chinese Learners in the last 30 years. *Overseas English*, (11):227–228+230 (in Chinese). 23, 24

Dingemanse, J. G. and Goedegebure, A. (2018). Optimising the effect of noise reduction algorithm ClearVoice in cochlear implant users by increasing the maximum comfort levels. *International Journal of Audiology*, 57(3):230–235. 164

Dong, X., Liu, B., Chen, X., and Gong, S. (2010). Quality of Life for Prelingually Deafened Adolescents with Cochlear Implants. *Journal of Audiology and Speech Pathology*, 18(3):275–277 (in Chinese). 26

Duchesne, L., Millette, I., Bhérer, M., and Gobeil, S. (2017). Auditory performance and subjective benefits in adults with congenital or prelinguistic deafness who receive cochlear implants during adulthood. *Cochlear Implants International*, 18(3):143–152. 26

Elliot, CE. (1991). The relationship between the perception and production of Mandarin tones: An exploratory study. *University of Hawai'i Working Papers in ESL*, 10(2):177–204. 23

Elosua, P. and De Boeck, P. (2020). Educational assessment issues in linguistically diverse contexts: A case study using a generalised linear mixed model. *Language, Culture and Curriculum*, 33(3):305–318. 53

Emmorey, K., Luk, G., Pyers, J. E., and Bialystok, E. (2008). The Source of Enhanced Cognitive Control in Bilinguals: Evidence from Bimodal Bilinguals. *Psychological Science*, 19(12):1201–1206. 167

Fan, J. and Wei, T. (2016). An Analysis on the Characteristics of Vowel Pronunciation of the Mental Retardation Children. *Journal of Suihua University*, 36(7):61–64 (in Chinese). 110

Feng, J. and Wu, X. (2020). Review of the Perception and Production of Lexical Tones by Chinese Cochlear Implant Users. *Chinese Journal of Phonetics*, (02):86–95 (in Chinese). 30

Feng, L. and Wu, D. (2017). Effect of cochlear implantation in combination with hearing aids on tone recognition. *Journal of Otolaryngology and Ophthalmology of Shandong University*, 31(05):27–30+37 (in Chinese). 26

Feng, Y. (2022). *Categorical Speech Perception across the Lifespan*. PhD thesis, The Hong Kong Polytechnic University, Hong Kong. 22

Feng, Y., Peng, G., and Wang, W. S.-Y. (2020). Age-Related Differences of Tone Perception in Mandarin-Speaking Seniors. In *INTERSPEECH*, pages 1629–1633. 22

Flege, J. E. and Bohn, O.-S. (2021). The revised speech learning model (SLM-r). *Second language speech learning: Theoretical and empirical progress*, pages 3–83. 7, 24

Flege, J. E., Schirru, C., and MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech communication*, 40(4):467–491. 7, 24

Flemming, E. and Cho, H. (2017). The phonetic specification of contour tones: Evidence from the Mandarin rising tone. *Phonology*, 34(1):1–40. 13

Fon, J. and Chiang, W.-Y. (1999). What does Chao have to say about tones?-a case study of Taiwan Mandarin. *Journal of Chinese Linguistics*, 27(1):13–37. 14

Freeman, J. B. (2018). Doing Psychological Science by Hand. *Current Directions in Psychological Science*, 27(5):315–323. 9, 85, 86

Freeman, J. B. and Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42(1):226–241. 85

Fu, W. and Wu, W. (2014). Status of speech audiometry materials in China. *Beijing Biomedical Engineering*, 33(03):318–321+326 (in Chinese). 17

Gałecki, A. and Burzykowski, T. (2013). *Linear Mixed-Effects Models Using R*. Springer Texts in Statistics. Springer New York, New York, NY. 53, 92

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11(2):149–175. 15, 154

Gao, M. (2002). Tones in whispered Chinese: Articulatory features and perceptual cues. 165

Gao, Q., Wong, L. L., and Chen, F. (2021). A Review of Speech Perception of Mandarin-Speaking Cochlear Implant Children. *Frontiers in neuroscience*, page 1670. 25

Gao, Y. (2021). A Study on Tone Pattern of Gulang Dialect in Gansu Province. *Journal of Gansu Normal Colleges*, 26(03):14–21 (in Chinese). 15

Gerrits, E. and Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, 66(3):363–376. 47, 78

Ghosh, P. K. and Narayanan, S. S. (2009). Pitch contour stylization using an optimal piecewise polynomial approximation. *IEEE signal processing letters*, 16(9):810–813. 14

Gleason, J. B. and Ratner, N. B. (2017). *The Development of Language*. 9th edition edition. 3

Guenther, F. H. and Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, 25(5):408–422. 166

Guo, R., Ma, X., Liao, M., Liu, Y., Hu, Y., Qian, X., Tang, Q., Guo, X., Chai, R., and Gao, X. (2019). Development and application of cochlear implant-based electric-acoustic stimulation of spiral ganglion neurons. *ACS Biomaterials Science & Engineering*, 5(12):6735–6741. 164

Gussenhoven, C. and Van de ven, M. (2020). Categorical perception of lexical tone contrasts and gradient perception of the statement–question intonation contrast in Zhumadian Mandarin. *Language and Cognition*, 12(4):614–648. 78

Hallé, P. A., Chang, Y.-C., and Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32(3):395–421. 16, 23

Han, D., Liu, B., Zhou, N., Chen, X., Kong, Y., Liu, H., Zheng, Y., and Xu, L. (2009). Lexical Tone Perception with HiResolution and HiResolution 120 Sound-Processing Strategies in Pediatric Mandarin-Speaking Cochlear Implant Users. *Ear & Hearing*, 30(2):169–177. 25, 112

Han, D., Zhou, N., Li, Y., Chen, X., Zhao, X., and Xu, L. (2007). Tone production of Mandarin Chinese speaking children with cochlear implants. *International journal of pediatric otorhinolaryngology*, 71(6):875–880. 25, 26

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 40(2):269–279. 23, 112

Harrell Jr, F. E. and with contributions from Charles Dupont and many others (2021). Hmisc: Harrell Miscellaneous. 125

Harris, M. S., Capretta, N. R., Henning, S. C., Feeney, L., Pitt, M. A., and Moberly, A. C. (2016). Postoperative rehabilitation strategies used by adults with cochlear implants: A pilot study. *Laryngoscope Investigative Otolaryngology*, 1(3):42–48. 164

Häußler, SM., Knopke, S., Dudka, S., Gräbel, S., Ketterer, MC., Battmer, R.-D., Ernst, A., and Olze, H. (2020). Improvement in tinnitus distress, health-related quality of life and psychological comorbidities by cochlear implantation in single-sided deaf patients. *Hno*, 68(1):1–10. 27

Heald, S. L. M. and Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, 8. 166

Hersh, M. A. and Johnson, M. A., editors (2003). *Assistive Technology for the Hearing-impaired, Deaf and Deafblind*. Springer-Verlag, London. 3

Hou, J., Chen, y., Xing, Y., and Dang, J. (2019). Acoustic Attributes of Citation Tones in Standard Chinese Produced by Prelingually Deaf Adults. pages 1786–1790. 13, 27, 28, 118, 150

Huang, TS., Wang, NM., and Liu, SY. (1995). Tone perception of Mandarin-speaking postlingually deaf implantees using the Nucleus 22-Channel Cochlear Mini System. *The Annals of otology, rhinology & laryngology. Supplement*, 166:294–298. 6, 25

Huang, W.-T., Liu, C., Dong, Q., and Nan, Y. (2015a). Categorical perception of lexical tones in mandarin-speaking congenital amusics. *Frontiers in Psychology*, 6. 49, 80

Huang, W.-T., Nan, Y., Dong, Q., and Liu, C. (2015b). Just-noticeable difference of tone pitch contour change for Mandarin congenital amusics. *The Journal of the Acoustical Society of America*, 138(1):EL99–EL104. 16

Huang, X., Wang, M., Shen, Z., and Zhang, J. (2021). The effect of different degrees of hearing impairment on children's articulation ability. *Modern Practical Medicine*, 33(12):1587–1590 (in Chinese). 26, 29, 110

Hunsicker, M. E., Kappel, C. V., Selkoe, K. A., Halpern, B. S., Scarborough, C., Mease, L., and Amrhein, A. (2016). Characterizing driver-response relationships in marine pelagic ecosystems for improved ocean management. *Ecological Applications*, 26(3):651–663. 134

Illner, V., Sovka, P., and Rusz, J. (2020). Validation of freely-available pitch detection algorithms across various noise levels in assessing speech captured by smartphone in Parkinson's disease. *Biomedical Signal Processing and Control*, 58:101831. 117

Ito, K., Suzuki, Y., Toma, M., Shiroma, M., and Kaga, K. (2002). Postlingual collapse of language and its recovery after cochlear implantation. *International journal of pediatric otorhinolaryngology*, 62(3):261–265. 27

Jeng, H.-H. (1979). The acquisition of Chinese phonology in relation to Jakobson's laws of irreversible solidarity. In *Proceedings of the 9th International Congress of Phonetic Sciences*, volume 2, pages 155–161. University of Copenhagen Copenhagen, Denmark. 20, 112

Jeng, J.-Y., Weismer, G., and Kent, R. D. (2006). Production and perception of mandarin tone in adults with cerebral palsy. *Clinical linguistics & phonetics*, 20(1):67–87. 13

Ji, Y., Zhu, X., and Yang, Y. (2015). The history and present situation of Chinese speech audiometry materials. *Journal of Audiology and Speech Pathology*, 23(02):206–211 (in Chinese). 17

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., and Yang, Y. (2012). Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. *Memory & Cognition*, 40(7):1109–1121. 47, 78, 79, 80

Jiao, L. and Xu, Y. (2019). Whispered Mandarin has no production-enhanced cues for tone and intonation. *Lingua*, 218:24–37. 165

Jouvet, D. and Laprie, Y. (2017). Performance analysis of several pitch detection algorithms on simulated and real noisy speech data. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 1614–1618, Kos, Greece. IEEE. 117

Kang, X. and Wu, H. (2020). Acoustic Study on the Perception and Output of Chinese Monosyllabic Tone of Dungan Learners. *Overseas Chinese Education*, 21(01):58–68 (in Chinese). 15, 24

Karltorp, E., Eklöf, M., Östlund, E., Asp, F., Tideholm, B., and Löfkvist, U. (2020). Cochlear implants before 9 months of age led to more natural spoken language development without increased surgical risks. *Acta Paediatrica*, 109(2):332–341. 169

Kauffman, J., Hallahan, D., and Pullen, P. (2013). *Exceptional Learners: Pearson New International Edition PDF eBook: An Introduction to Special Education*. Pearson Education. 3

Ke, C. (2012). Research in second language acquisition of Chinese: Where we are, where we are going. *Journal of the Chinese Language Teachers Association*, 47(3):43–113. 23

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., and Scherbaum, S. (2020). Design factors in mouse-tracking: What makes a difference? *Behavior Research Methods*, 52(1):317–341. 9

Kim, H. H. and Barrs, D. M. (2006). Hearing Aids: A Review of What's New. *Otolaryngology–Head and Neck Surgery*, 134(6):1043–1050. 163

Kiriloff, C. (1969). On the Auditory Perception of Tones in Mandarin. *Phonetica*, 20(2-4):63–67. 23

Kobosko, J., Jedrzejczak, W. W., Gos, E., Geremek-Samsonowicz, A., Ludwikowski, M., and Skarzynski, H. (2018). Self-esteem in the deaf who have become cochlear implant users as adults. *PLoS One*, 13(9):e0203680. 164

Kochkin, S. (2001). MarkeTrak VI: The VA and direct mail sales spark growth in hearing aid market. *Hearing Review*, 8(12):16–24. 164

Kong, Y.-Y. and Zeng, F.-G. (2006). Temporal and spectral cues in Mandarin tone recognition. *The Journal of the Acoustical Society of America*, 120(5):2830–2840. 17, 165

Kral, A., Kronenberger, W. G., Pisoni, D. B., and O'Donoghue, G. M. (2016). Neurocognitive factors in sensory restoration of early deafness: A connectome model. *The Lancet Neurology*, 15(6):610–621. 30

Kral, A. and Sharma, A. (2012). Developmental neuroplasticity after cochlear implantation. *Trends in Neurosciences*, 35(2):111–122. 26

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13). 53, 92

Lan, L., Li, Q., Xie, L., Xiong, F., Han, B., and Wang, Q. (2019). Clinical Application of Speech Recognition in Noise Test in Patients with Auditory Neuropathy. *Chinese Journal of Otology*, 17(01):9–16 (In Chnese). 17

Lazard, D. S., Lee, H.-J., Gaebler, M., Kell, C. A., Truy, E., and Giraud, A.-L. (2010). Phonological processing in post-lingual deafness and cochlear implant outcome. *Neuroimage*, 49(4):3443–3451. 27

Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In Leather, J. and James, A., editors, *Perceptual and Productive Learning of Chinese Lexical Tone by Dutch and English Speakers*, pages 72–89, Amsterdam: University of Amsterdam. 24

Leather, J. (2011). Interrelation of perceptual and productive learning in the initial acquisition of second-language tone. In James, A. and Leather, J., editors, *Second-Language Speech: Structure and Process*, pages 75–102. De Gruyter Mouton, Berlin, New York. 23

Lee, C.-Y. and Cheng, Y.-Y. (2020). Neurophysiological Studies of Mandarin Lexical Tone Acquisition in Early Childhood. In Liu, H.-M., Tsao, F.-M., and Li, P., editors, *Speech Perception, Production and Acquisition*, pages 101–116. Springer Singapore, Singapore. 20

Lee, K. Y., van Hasselt, C. A., and Tong, M. C. (2010). Lexical tone perception ability of profoundly hearing-impaired children: Performance of cochlear implant and hearing aid users. *Otology & Neurotology*, 31(7):1079–1087. 25

Lee, K. Y., Van Hasselt, CA., Chiu, SN., and Cheung, D. M. (2002). Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. *International journal of pediatric otorhinolaryngology*, 63(2):137–147. 112

Lee, K. Y. S. and van Hasselt, C. A. (2004). Long-term outcome on spoken word recognition ability of young children with cochlear implants. In *International Congress Series*, volume 1273, pages 300–303. Elsevier. 25

Lee, Y.-S., Vakoch, D. A., and Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25(5):527–542. 15

Li, A., Wang, N., Li, J., Zhang, J., and Liu, Z. (2014a). Mandarin lexical tones identification among children with cochlear implants or hearing aids. *International Journal of Pediatric Otorhinolaryngology*, 78(11):1945–1952. 25

Li, C. N. and Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2):185–199. 21

Li, G., Zhu, H., Dodd, B., Jiang, Peng, D., and Shu, H. (2000). Phonological Acquisition of Putonghua-Speaking Children. *Acta Psychologica Sinica*, 32(2):170–176 (in Chinese). 21

Li, L. (2018). A study on the level of phonetic development in Mandarin Chinese among Shanghai children between 23 to 30 months (in Chinese). Master's thesis, Shanghai Normal University. 29, 110

Li, Q. and Chen, Y. (2016). An acoustic study of contextual tonal variation in Tianjin Mandarin. *Journal of Phonetics*, 54:123–150. 14

Li, Q., Xia, S., Zhao, F., and Qi, J. (2014b). Functional changes in people with different hearing status and experiences of using Chinese sign language: An fMRI study. *Journal of Communication Disorders*, 50:51–60. 4

Li, Y.-L., Lin, Y.-H., Yang, H.-M., Chen, Y.-J., and Wu, J.-L. (2018). Tone production and perception and intelligibility of produced speech in Mandarin-speaking cochlear implanted children. *International Journal of Audiology*, 57(2):135–142. 15

Liang, Q. and Mason, B. (2013). Enter the dragon – China's journey to the hearing world. *Cochlear Implants International*, 14(sup1):S26–S31. 5, 26

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1):1–36. 166

Lin, M. (1965). The pitch indicator and the pitch characteristics of tones in Standard Chinese. *Acta Acustica*, 2(01):8–15 (in Chinese). 13

Lin, T. (1985). *The Record of Phonetics Experiment for Beijing Mandarin*. Peking University Press. 13

Lin, T. (1996). Phonetic studies and teaching Chinese as a foreign language. *Teaching Chinese in the World*, 37(3):20–23 (in Chinese). 23

Liu, C. (2007). A study on the phonetic development of Standard Chinese among Shanghai children between 18 to 23 months (in Chinese). Master's thesis, Shanghai Normal University. 29, 110

Liu, C.-J., Liu, S.-Y., Huang, G.-Y., Kuo, Y.-C., and Wang, N.-M. (2010). Psychosocial Benefits of the Adults with Cochlear Implants and the Related Factors to Predict These Outcomes. *Journal of the Speech-Language-Hearing Association of Taiwan*, 25(1):1–22 (in Chinese). 26

Liu, F. (1924). *Record of Experiments on the Four Tones*. Qunyi Shushe (Reprinted by Zhonghua Book Company in 1950), Shanghai. 13

Liu, J. (2004). Perceiving the boundary between the lexical rising tone and the falling-raising tone. In Shi, F. and Shen, Z., editors, *Joy of Research a Festschrift in Honor of Professor William S-Y. Wang on His Seventieth Birthday*, pages 222–233. Nankai University Press, Tianjin. 15, 16

Liu, L., Chen, A., and Kager, R. (2017). Perception of tones in Mandarin and Dutch adult listeners. *Language and Linguistics*, 18(4):622–646. 159

Liu, Q., Huang, Z., Chen, L., Du, X., and Sun, X. (2011). On a Comparison of Phonemic Contrast Recognition Between Children with Cochlear Implants,Hearing Aids and Normal Hearing. *Chinese Journal of Special Education*, (02):25–29 (in Chinese). 165

Liu, Q., Zhang, X., Zhang, G., Yu, H., Zheng, T., and Zheng, Q. (2019). Speech Recognition Ability of Students with Visual Impairment in Quiet and Noisy Environment. *Chinese Scientific Journal of Hearing and Speech Rehabilitation*, 17(06):433–437 (in Chinese). 17

Liu, S. and Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and speech*, 47(2):109–138. 165

Ma, J., Zhu, J., Yang, Y., and Chen, F. (2021). The development of categorical perception of segments and suprasegments in Mandarin-speaking preschoolers. *Frontiers in Psychology*, page 2427. 79

Ma, X., Liu, X., Zhang, S., Zhang, M., and Cao, X. (2014). The Development and Prospect of Hearing Aids. *Journal of MUC (Natural Sciences Edition)*, 23(1):39–43 (in Chinese). 31

Macmillan, N. A. and Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Lawrence Erlbaum Associates, Mahwah, N.J, 2nd ed edition. 16, 81

Macrae, J. H. and Dillon, H. (1996). Gain, frequency response, and maximum output requirements for hearing aids. *Journal of rehabilitation research and development*, 33(4):363. 30

Majorano, M., Guerzoni, L., Cuda, D., and Morelli, M. (2020). Mothers' emotional experiences related to their child's diagnosis of deafness and cochlear implant surgery: Parenting stress and child's language development. *International journal of pediatric otorhinolaryngology*, 130:109812. 164

Mäki-Torkko, E. M., Vestergren, S., Harder, H., and Lyxell, B. (2015). From isolation and dependence to autonomy–expectations before and experiences after cochlear implantation in adult cochlear implant users and their significant others. *Disability and rehabilitation*, 37(6):541–547. 164

Mangiafico, S. (2023). Rcompanion: Functions to Support Extension Education Program Evaluation. 50

Mao, Y., Chen, H., Xie, S., and Xu, L. (2020). Acoustic Assessment of Tone Production of Prelingually-Deafened Mandarin-Speaking Children With Cochlear Implants. *Frontiers in Neuroscience*, 14. 25

Mao, Y., Li, G., Wang, h., Zhang, X., Cai, ., , å., , å., and , . (2017). Research of the Consonant and Vowel Recognition Conditions of Children with Cochlear Implantation at Different Stages. */Chinese and Foreign Medical Research*, 15(10):24–25 (in Chinese). 29, 110

Martines, F., Martines, E., Mucia, M., Sciacca, V., and Salvago, P. (2013). Prelingual sensorineural hearing loss and infants at risk: Western Sicily report. *International Journal of Pediatric Otorhinolaryngology*, 77(4):513–518. 4

Mattock, K. and Burnham, D. (2006). Chinese and English Infants' Tone Perception: Evidence for Perceptual Reorganization. *Infancy*, 10(3):241–265. 20

Mattock, K., Molnar, M., Polka, L., and Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, 106(3):1367–1381. 20

Mayberry, R. I. and Kluender, R. (2018). Rethinking the critical period for language: New insights into an old question from American Sign Language. *Bilingualism: Language and Cognition*, 21(5):886–905. 19

McGarr, N. S. and Osberger, M. J. (1978). Pitch deviancy and intelligibility of deaf speech. *Journal of communication Disorders*, 11(2-3):237–247. 25

Meadow-Orlans, K. P. (1987). An analysis of the effectiveness of early intervention programs for hearing-impaired children. In Guralnick, M. J. and Bennett, F. C., editors, *The Effectiveness of Early Intervention for At-Risk and Handicapped Children*, pages 325–362. Academic Press. 3

Michels, A., Oukheira, Y., Brendel, M., Aschendorff, A., Arndt, S., and Wesarg, T. (2022). Effect of adaptive beamforming and noise reduction algorithms on speech intelligibility and noise tolerance in bimodal cochlear implant users. *Cochlear Implants International*, 23(3):148–157. 164

Miracle, W. C. (1989). Tone production of American students of Chinese: A preliminary acoustic study. *Journal of the Chinese Language Teachers Association*, 24(3):49–65. 23, 24

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. CRC Press, Boca Raton. 54, 73

Mitchell-Innes, A., Morse, R., Irving, R., and Begg, P. (2017). Implantable microphones as an alternative to external microphones for cochlear implants. *Cochlear Implants International*, 18(6):304–313. 164

Mok, M., Holt, C. M., Lee, K. Y. S., Dowell, R. C., and Vogel, A. P. (2017). Cantonese Tone Perception for Children Who Use a Hearing Aid and a Cochlear Implant in Opposite Ears. *Ear & Hearing*, 38(6):e359–e368. 9

Morton, K. D., Torrione Jr, P. A., Throckmorton, C. S., and Collins, L. M. (2008). Mandarin Chinese tone identification in cochlear implants: Predictions from acoustic models. *Hearing research*, 244(1-2):66–76. 25

Moulines, E. and Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9(5-6):453–467. 42

Ning, M. (2018). The effect of aging on speech perception: A review. *Modern Linguistics*, 6(05):714–722 (in Chinese). 22

O'Connor, E. F. and O'Connor, A. F. (2010). Hearing preservation surgery: Current opinions. In *Cochlear Implants and Hearing Preservation*, volume 67, pages 108–115. Karger Publishers. 164

Oh, D., Walker, M., and Freeman, J. B. (2021). Person knowledge shapes face identity perception. *Cognition*, 217:104889. 86

Park, L. R., Teagle, H. F., Gagnon, E., Woodard, J., and Brown, K. D. (2019). Electric-acoustic stimulation outcomes in children. *Ear and Hearing*, 40(4):849–857. 164

Peasgood, A., Brookes, N., and Graham, J. (2003). Performance and benefit as outcome measures following cochlear implantation in non-traditional adult candidates: A pilot study. *Cochlear Implants International*, 4(4):171–190. 26

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., and Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, 38(4):616–624. 15, 16, 17, 52

Peng, S.-C., Lu, H.-P., Lu, N., Lin, Y.-S., Deroche, M. L. D., and Chatterjee, M. (2017). Processing of Acoustic Cues in Lexical-Tone Identification by Pediatric Cochlear-Implant Recipients. *Journal of Speech, Language, and Hearing Research*, 60(5):1223–1235. 31

Peng, S.-C., Tomblin, J. B., Cheung, H., Lin, Y.-S., and Wang, L.-S. (2004). Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. *Ear and hearing*, 25(3):251–264. 25

Pénicaud, S., Klein, D., Zatorre, R. J., Chen, J.-K., Witcher, P., Hyde, K., and Mayberry, R. I. (2013). Structural brain changes linked to delayed first language acquisition in congenitally deaf individuals. *NeuroImage*, 66:42–49. 19

Pillsbury III, H. C., Dillon, M. T., Buchman, C. A., Staecker, H., Prentiss, S. M., Ruckenstein, M. J., Bigelow, D. C., Telischi, F. F., Martinez, D. M., and Runge, C. L. (2018). Multicenter US clinical trial with an electric-acoustic stimulation (EAS) system in adults: Final outcomes. *Otology & Neurotology*, 39(3):299. 164

Profillidis, V. and Botzoris, G. (2019). Chapter 6 - trend projection and time series methods. In Profillidis, V. and Botzoris, G., editors, *Modeling of Transport Demand: Analyzing, Calculating, and Forecasting Transport Demand*, pages 225–270. Elsevier. 53

Qin, P. and Yang, J. (2021). Citation and Disyllabic Tone Patterns of Shouyang Dialect. *Modern Chinese*, (06):42–47 (in Chinese). 15

Qiu, S., Zou, Y., Zhang, Y., and Xie, W. (2021). Analysis on rehabilitation effect of cochlear implant in 50 hearing-impaired children. *Journal of Audiology and Speech Pathology*, 29(02):211–213 (in Chinese). 15

Rader, T., Döge, J., Adel, Y., Weissgerber, T., and Baumann, U. (2016). Place dependent stimulation rates improve pitch perception in cochlear implantees with single-sided deafness. *Hearing Research*, 339:94–103. 5

Rader, T., Fastl, H., and Baumann, U. (2013). Speech perception with combined electric-acoustic stimulation and bilateral cochlear implants in a multisource noise field. *Ear and Hearing*, 34(3):324–332. 164

Rapport, F., Lo, C. Y., Elks, B., Warren, C., and Clay-Williams, R. (2022). Cochlear implant aesthetics and its impact on stigma, social interaction and quality of life: A mixed-methods study protocol. *BMJ open*, 12(3):e058406. 164

Ren, C., Zhao, Z., Lin, Y., and Zha, D. (2022). Evaluation of Hearing, Speech and Quality of Life after Cochlear Implantation in Postlingually Deafened Adults. *Journal of Audiology and Speech Pathology*, 30(02):115–118 (in Chinese). 27

Riestenberg, K. J. (2017). *Acoustic Salience and Input Frequency in L2 Lexical Tone Learning: Evidence from a Zapotec Revitalization Program in San Pablo Macuiltianguis*. PhD thesis, Georgetown University. 161

Rong, R. (2012). The Devoicing Process of the Middle Chinese Voiced Initials in Mandarin Dialects. *Nankai Linguistics*, 20(02):85–95+163 (in Chinese). 16

Rong, R. and Shi, F. (2013). Perception of the Boundary between Mandarin Tone 1 and Tone 3:Effects of Pitch and Duration. *Linguistic Sciences*, 12(01):17–26 (in Chinese). 16

Sahli, S., Arslan, U., and Belgin, E. (2009). Depressive emotioning in adolescents with cochlear implant and normal hearing. *International journal of pediatric otorhinolaryngology*, 73(12):1774–1779. 27

Sahli, S. and Belgin, E. (2006). Comparison of self-esteem level of adolescents with cochlear implant and normal hearing. *International Journal of Pediatric Otorhinolaryngology*, 70(9):1601–1608. 27

Schatzer, R., Koroleva, I., Griessner, A., Levin, S., Kusovkov, V., Yanov, Y., and Zierhofer, C. (2015). Speech perception with interaction-compensated simultaneous stimulation and long pulse durations in cochlear implant users. *Hearing Research*, 322:99–106. 30

Schatzer, R., Vermeire, K., Visser, D., Krenmayr, A., Kals, M., Voormolen, M., Van de Heyning, P., and Zierhofer, C. (2014). Electric-acoustic pitch comparisons in single-sided-deaf cochlear implant users: Frequency-place functions and rate pitch. *Hearing Research*, 309:26–35. 5

Schoonahd, J. W., Gould, J. D., and Miller, L. A. (1973). Studies of Visual Inspection. *Ergonomics*, 16(4):365–379. 50

Schramm, D., Fitzpatrick, E., and Séguin, C. (2002). Cochlear Implantation for Adolescents and Adults with Prelinguistic Deafness:. *Otology & Neurotology*, 23(5):698–703. 26

Schwartz, J.-L., Boë, L.-J., and Abry, C. (2007). Linking the dispersion-focalization theory (DFT) and the maximum utilization of the available distinctive features (MUAF) principle in a perception-for-action-control theory (PACT). In Solé, M., Beddor, P., and Ohala, M., editors, *Experimental Approaches to Phonology*, pages 104–124. Oxford University Press, Oxford. 166

Shao, J., Lau, R. Y. M., Tang, P. O. C., and Zhang, C. (2019). The Effects of Acoustic Variation on the Perception of Lexical Tone in Cantonese-Speaking Congenital Amusics. *Journal of speech, language, and hearing research*, 62(1):190–205. 81

Shen, J., Deutsch, D., and Rayner, K. (2013a). On-line perception of Mandarin Tones 2 and 3: Evidence from eye movements. *The Journal of the Acoustical Society of America*, 133(5):3016–3029. 16

Shen, X. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association*, 24(3):27–47. 24

Shen, X., Guan, Y., Jiang, G., Yu, L., and Zhou, J. (2013b). Characteristics of tones in vowels in prelingually hearing-impaired children after cochlear implant. *Chinese Journal of Rehabilitation Theory and Practice*, 19(07):608–611 (in Chinese). 112

Shen, X., Guan, Y., , G., Yu, L., and Zhou, J. (2013c). Characteristics of Tones in Vowels in Prelingual Hearing-impaired Children after Cochlear Implant. *Chinese Journal of Rehabilitation Theory and Practice*, 19(07):608–611 (in Chinese). 15

Shi, F. (1986). Analysis of bisyllabic tone in Tianjin dialect. *Studies in Language and Linguistics*, (01):77–90 (in Chinese). 14

Shi, F. and Liao, R. (1994). *The Selected Papers in Phontics (in Chinese)*. Beijing Language and Culture Institute Press, Beijing. 14

Shi, F. and Wang, P. (2006). A statistic analysis of the tones in Beijing Mandarin. *Studies of the Chinese Language*, (01):33–40+96 (in Chinese). 15

Shi, H. (2019). An Experimental Phonetic Analysis of the Tone Pattern and Tone Variation of Xihua Dialect. *Journal of Zaozhuang University*, 36(06):38–43 (in Chinese). 15

Shih, C. and Lu, H.-Y. D. (2015). Effects of talker-to-listener distance on tone. *Journal of Phonetics*, 51:6–35. 14

Srinivasan, A. G., Padilla, M., Shannon, R. V., and Landsberger, D. M. (2013). Improving speech perception in noise with current focusing in cochlear implant users. *Hearing research*, 299:29–36. 164

Stagray, J. R. and Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguistics*, pages 143–163. 23

Straatman, L. V., Huinck, W. J., Langereis, M. C., Snik, A. F. M., and Mulder, J. J. (2014). Cochlear Implantation in Late-Implanted Prelingually Deafened Adults: Changes in Quality of Life. 35(2):7. 26

Studebaker, G. A. (1985). A ”Rationalized” Arcsine Transform. *Journal of Speech, Language, and Hearing Research*, 28(3):455–462. 90

Sun, C. (2006). *Chinese: A Linguistic Introduction*. Cambridge University Press. 1

Sun, X., Zhang, F., Huang, H., and Zhang, L. (2009). Assessment of speech and hearing in children with hearing impairment. *Journal of Audiology and Speech Pathology*, 17(04):327–329 (in Chinese). 37

Talkin, D. (2015). REAPER: Robust Epoch And Pitch EstimatoR. 117

Tan, J., Dowell, R., and Vogel, A. (2016). Mandarin lexical tone acquisition in cochlear implant users with prelingual deafness: A review. *American journal of audiology*, 25(3):246–256. 25

Tang, P., Yuen, I., Rattanasone, N. X., Gao, L., and Demuth, K. (2021). Longer Cochlear Implant Experience Leads to Better Production of Mandarin Tones for Early Implanted Children. *Ear and Hearing*, 42(5):1405–1411. 14, 118

Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., and Demuth, K. (2019). The acquisition of Mandarin tonal processes by children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 62(5):1309–1325. 14, 25, 31

Tao, D., Deng, R., Jiang, Y., Galvin III, J. J., Fu, Q.-J., and Chen, B. (2015). Melodic pitch perception and lexical tone perception in Mandarin-speaking cochlear implant users. *Ear and Hearing*, 36(1):102. 112

Team, R. (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 50, 53, 123, 125

Tejani, V. D., Kim, J.-S., Oleson, J. J., Abbas, P. J., Brown, C. J., Hansen, M. R., and Gantz, B. J. (2021). Residual hair cell responses in electric-acoustic stimulation cochlear implant users with complete loss of acoustic hearing after implantation. *Journal of the Association for Research in Otolaryngology*, 22(2):161–176. 164

Toh, X. R., Lau, F., and Wong, F. C. K. (2022). Individual differences in nonnative lexical tone perception: Effects of tone language repertoire and musical experience. *Frontiers in Psychology*, 13:940363. 24

Tomlinson, J. M. and Ronderos, C. R. (2021). Does intonation automatically strengthen scalar implicatures? *Semantics and Pragmatics*, 14(4). 86

Tsao, F.-M. (2008). The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *Chinese Journal of Psychology*, 50(2):111–124. 20, 21

Tupper, P., Leung, K., Wang, Y., Jongman, A., and Sereno, J. A. (2020). Characterizing the distinctive acoustic cues of Mandarin tones. *The Journal of the Acoustical Society of America*, 147(4):2570–2580. 13, 14, 118, 154

Van Hessen, A. J. and Schouten, M. E. H. (1999). Categorical perception as a function of stimulus quality. *Phonetica*, 56(1-2):56–72. 159

Van Leussen, J.-W. and Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in psychology*, 6:1000. 7

van Schoonhoven, J., Sparreboom, M., van Zanten, B. G., Scholten, R. J., Mylanus, E. A., Dreschler, W. A., Grolman, W., and Maat, B. (2013). The effectiveness of bilateral cochlear implants for severe-to-profound deafness in adults: A systematic review. *Otology & Neurotology*, 34(2):190–198. 164

Von Ilberg, C. A., Baumann, U., Kiefer, J., Tillein, J., and Adunka, O. F. (2011). Electric-acoustic stimulation of the auditory system: A review of the first decade. *Audiology and Neurotology*, 16(Suppl. 2):1–30. 30

Wang, D. and Peng, G. (2012). The effects of pitch range and duration on tone categorical perception. In *Proceedings of the 3rd International Symposium on Tonal Aspects of Languages*, pages 26–29. 16

Wang, P., Shi, F., Rong, R., Chen, X., Li, S., and Wang, X. (2014). The perceptional category of tone3 in Standard Mandarin. *Studies of the Chinese Language*, 361(04):359–370+384 (in Chinese). 16

Wang, R., He, X., and Zeng, B. (2021a). Visual Benefit in Lexical Tone Perception in Mandarin: An Event-related Potential Study. *Neuroscience*, 466:196–204. 80

Wang, S., Liu, B., Dong, R., Zhou, Y., Li, J., Qi, B., Chen, X., Han, D., and Zhang, L. (2012). Music and lexical tone perception in Chinese adult cochlear implant users. *The Laryngoscope*, 122(6):1353–1360. 6

Wang, W., Zhou, N., and Xu, L. (2011). Musical pitch and lexical tone perception with cochlear implants. *International Journal of Audiology*, 50(4):270–278. 6, 31

Wang, W. S. (1976). Language Change. *Annals of the New York Academy of Sciences*, 208:61–72. 15

Wang, X., Wang, S., Fan, Y., Huang, D., and Zhang, Y. (2017a). Speech-specific categorical perception deficit in autism: An event-related potential study of lexical tone processing in Mandarin-speaking children. *Scientific reports*, 7(1):1–12. 17

Wang, Y. (2016). A Survey on Satisfaction of Post-support for Adults with Cochlear Implants (in Chinese). Master's thesis, Chongqing Normal University. 164

Wang, Y., Jongman, A., and Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, 113(2):1033–1043. 23, 24

Wang, Y. and Li, M. (2010). The Effects of Tone Pattern and Register in Perceptions of Tone 2 and Tone 3 in Mandarin. *Acta Psychologica Sinica*, 42(09):899–908 (in Chinese). 16, 23

Wang, Y., Shi, Y., Fu, Y., Wang, Q., Fu, Y., and Xi, X. (2015). Study of ceiling effect of commonly used Chinese recognition materials in post-lingual deafened patients with cochlear implant. *Journal of Clinical Otorhinolaryngology Head and Neck Surgery*, 29(04):298–303 (in Chinese). 17

Wang, Y. and Tan, X. (2015). The Identification and Discrimination of Tone2 and Tone3 in Mandarin:The Influence of Experimental Design on Categorical Perception of Tones. *Linguistic Sciences*, 14(04):337–352 (in Chinese). 15, 16

Wang, Y., Yang, X., Ding, H., Xu, C., and Liu, C. (2021b). Aging Effects on Categorical Perception of Mandarin Lexical Tones in Noise. *Journal of Speech, Language, and Hearing Research*, 64(4):1376–1389. 80

Wang, Y., Yang, X., and Liu, C. (2017b). Categorical Perception of Mandarin Chinese Tones 1–2 and Tones 1–4: Effects of Aging and Signal Duration. *Journal of Speech, Language, and Hearing Research*, 60(12):3667–3677. 16, 22, 48, 80, 161

Wei, C., Cao, K., Jin, X., Chen, X., and Zeng, F.-G. (2007). Psychophysical performance and Mandarin tone recognition in noise by cochlear implant users. *Ear and hearing*, 28(2 Suppl):62S. 17

Wei, C.-G., Cao, K., and Zeng, F.-G. (2004). Mandarin tone recognition in cochlear-implant subjects. *Hearing research*, 197(1-2):87–95. 6, 31

Wei, K. (2014). A Statistic Ayalysis of the tones of Isolated Words in Kucong Language. *Journal of Pu'er University*, 30(01):5–9 (in Chinese). 15

Wei, T. and Simko, V. (2021). R package 'corrplot': Visualization of a Correlation Matrix. 125

Wen, H. (2019). A Study of Chinese Tone Identification between Chinese and Indonesian Subjects. Master's thesis, Jinan University. 29

Whalen, D. H. and Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of phonetics*, 23(3):349–366. 41

Wie, O. B., Torkildsen, J. v. K., Schauber, S., Busch, T., and Litovsky, R. (September/October 2020). Long-Term Language Development in Children With Early Simultaneous Bilateral Cochlear Implants. *Ear and Hearing*, 41(5):1294–1305. 169

Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics*, 40(1):141–151. 21

Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *The Journal of the Acoustical Society of America*, 133(1):434–443. 21

Wong, P., Schwartz, R. G., and Jenkins, J. J. (2005). Perception and Production of Lexical Tones by 3-Year-Old, Mandarin-Speaking Children. *Journal of Speech, Language, and Hearing Research*, 48(5):1065–1079. 21

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1):3–36. 123

Wooi Teoh, S., Pisoni, D. B., and Miyamoto, R. T. (2004). Cochlear Implantation in Adults with Prelingual Deafness. Part I. Clinical Results:. *The Laryngoscope*, 114(9):1536–1540. 26

Wu, T. and Xu, Z. (1979). A Preliminary Analysis of the Records of Speech Development in Children from Birth to Three Years Old. *Acta Psychologica Sinica*, 24(2):153–165 (in Chinese). 21

Wu, X. (2006). *(Non-) Categorical Perception of Mandarin Tones: A Comparison between Speakers of Tone and Non-Tone Languages*. PhD thesis, University of Victoria. 23

Wu, Y. (2007). *The Analysis of Speech Recognition Abilities of Deafness Children Treated with Cochlear Implants or Hearing Aids under Different Noisy Condition (in Chinese)*. MM thesis, Central South University. 165

Wu, Z. and Lin, M. (1989). *An Outline of Experimental Phonetics*. Higher Education Press, Beijing. 13

Xi, J., Jiang, W., Zhang, L., and Shu, H. (2009). Categorical Perception of VOT and Lexical Tones in Chinese and the Developmental Course. *Acta Psychologica Sinica*, 41(07):572–579 (in Chinese). 21

Xi, J., Zhang, L., Shu, H., Zhang, Y., and Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, 170(1):223–231. 80

Xi, X. (2013). A 10-year development tour of cochlear implant related Chinese speech audiometry materials. *Chinese Journal of Otology*, 11(03):413–417 (In China). 17

Xia, J., Guan, Y., and Xue, Y. (2012). Comparison of Phonetic Articulation between Hearing-impaired and Normal Children in 4.5˜5.5 Years Old Groups. *Chinese Journal of Rehabilitation Theory and Practice*, 18(8):707–709 (in Chinese). 29, 110

Xiao, R., Liang, D., and Li, S. (2020). Effects of aging on the Mandarin lexical tone perception: Evidence from ERPs. *Acta Psychologica Sinica*, 52(01):1–11 (in Chinese). 22

Xie, C., Wang, L., Jia, X., Ding, S., Wang, G., and Liu, H. (2019). Clinical Analysis of 86 Adult Patients Postlingual Deafness with Cochlear Implantation. *Chinese Scientific Journal of Hearing and Speech Rehabilitation*, 17(05):346–348 (In Chiese). 26

Xie, J. and Xia, F. (2018). New Progress in the Education of Children with Hearing Impairment (2013-2014). In Xie, J. and Xia, F., editors, *New Progress in Special Education of China (2013-2014) (In Chinese)*, pages 12–26. China Renmin University Press, Beijing. 5

Xu, K. and Shi, F. (2022). The Shortcut to Chinese Phonetic Teaching——The Elaboration of Simplified Phonology of Mandarin. *International Chinese Language Education*, 7(03):5–12 (in Chinese). 161

Xu, M., Jing, SHAO., Liu, B., Wang, L., Ding, H., and Zhang, Y. (2022). Aging-Related Decline in Phonated and Whispered Speech Perception Not Compensated for by Increased Duration and Intensity: Evidence from Mandarin-Speaking Adult Listeners. 165

Xu, Y. (1997). Contextual Tonal Variations in Mandarin. *Journal of phonetics*, 25(1):61–83. 1

Xu, Y., Gandour, J. T., and Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, 120(2):1063–1074. 48, 78

Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L., and Demuth, K. (2018). Five-year-olds' acoustic realization of Mandarin tone sandhi and lexical tones in context are not yet fully adult-like. *Frontiers in psychology*, 9:817. 21

Xue, J. (2013). *The Study on Mandarin Tone2and Tone3by American and Thai Speakers*. PhD thesis, Peking University, Beijing. 23

Yan Feng, Gang Peng, and Shi-Yuan Wang, W. (2022). Categorical Perception of Lexical Tones in Mandarin-Speaking Seniors. *Journal of Speech, Language & Hearing Research*, 65(8):2789–2800. 22, 48

Yang, B. (2012). The gap between the perception and production of tones by American learners of Mandarin–An intralingual perspective. *Chinese as a Second Language Research*, 1(1):33–53. 24

Yang, B. (2015). *Perception and Production of Mandarin Tones by Native Speakers and L2 Learners*. Springer, Berlin, Heidelberg. 13, 23, 112, 154

Yang, J., Davis, B. L., and Diehl, R. L. (2021). The development of tonal duration in Mandarin-speaking children. *JASA Express Letters*, 1(8):085202. 112

Yang, J., Diehl, R., and Davis, B. (2008). The development of tonal duration in Mandarin-speaking children. *The Journal of the Acoustical Society of America*, 124(4):2498–2498. 21

Yang, J. and Liu, C. (2012). Categorical Perception of Lexical Tone in 6 to 8-year-old Monolingual and Bilingual Children. *Int. J. Asian Lang. Process.*, 22(2):49–62. 21

Yang, X., Wang, Y., Xu, L., Zhang, H., Xu, C., and Liu, C. (2015). Aging effect on Mandarin Chinese vowel and tone identification. *The Journal of the Acoustical Society of America*, 138(4):EL411–EL416. 22, 29

Yang, Y., Zhao, H., Sun, X., Lu, X., Cui, Y., and Wang, K. (2013). The recognition of consonants, vowels, tones and spondaic words by hearing-impaired children in the electro-acoustic stimulation mode. *Chinese Journal of Special Education*, (5):32–36 (in Chinese). 29, 110

Yaru Meng, Fei Chen, Yan Feng, Gang Peng, and Wei Zheng (2022). Age-Related Differences of Mandarin Tone and Consonant Aspiration Perception in Babble Noise. *Journal of Speech, Language & Hearing Research*, 65(9):3438–3451. 22

Yeung, H. H., Chen, K. H., and Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*. 20

Yin, J. (2021). An Acoustic Experimental Study on the Errors of the Intermediate Chinese Learners in Cambodia. Master's thesis, Yunnan Normal University. 23

Yin, L., Tu, W., Gao, S., Tong, D., Teng, ., Xu, Z., Cao, Y., Hu, B., and Zeng, F. (2016). Deafness and Hearing Aid or Prostheses Selection. *China Medical Device Information*, 22(05):23–29+63 (in Chinese). 25

Yip, M. J. (1980). *The Tonal Phonology of Chinese*. PhD thesis, Massachusetts Institute of Technology. 24

Yu, A. C. L. (2010). Tonal effects on perceived vowel duration. *Laboratory Phonology*, 10:151–168. 2

Zan, F. and Tang, S. (1998). The study of Phonetic Development of Hearing Impaired Children. *China Journal of Special Education*, 5(1):12–19. 25

Zeng, F.-G., Tang, Q., and Lu, T. (2014). Abnormal Pitch Perception Produced by Cochlear Implant Stimulation. *PLoS ONE*, 9(2):e88662. 31

Zhang, G., Li, N., and Lin, X. (2017). The Assessment of the Effect after Cochlear Implant of Postlingual Deafness. *Chinese Scientific Journal of Hearing and Speech Rehabilitation*, 15(03):191–194 (in Chinese). 26

Zhang, H., Zhang, J., Ding, H., and Zhang, Y. (2020a). Bimodal Benefits for Lexical Tone Recognition: An Investigation on Mandarin-speaking Preschoolers with a Cochlear Implant and a Contralateral Hearing Aid. *Brain Sciences*, 10(4):238. 52, 80, 163

Zhang, H., Zhang, J., Peng, G., Ding, H., and Zhang, Y. (2020b). Bimodal Benefits Revealed by Categorical Perception of Lexical Tones in Mandarin-Speaking Kindergarteners With a Cochlear Implant and a Contralateral Hearing Aid. *Journal of Speech, Language, and Hearing Research*, 63(12):4238–4251. 16, 17, 47, 52, 80, 163

Zhang, J., Wang, W., Wang, Y., Chen, C., Sun, G., and Chen, Y. (2020c). Assessment of speech perception and quality of life after cochlear implant in postlingual deaf adults. *Journal of Clinical Otorhinolaryngology Head and Neck Surgery*, 34(11):1002–1004 (in Chinese). 27

Zhang, L. (2005). Experimental Research on the Development of Chinese Phonological Awareness of Hearing-impaired students. *Chinese Journal of Special Education*, 65(11):22–27 (in Chinese). 25

Zhang, L., Huang, Z., Hu, L., Zhu, Q., and Yi, L. (2012a). On the Tone Pronunciation for Preschool Children with Hearing Impairments. *Chinese Journal of Special Education*, 140(02):21–25 (in Chinese). 13, 15

Zhang, Y. (2014). *Chinese-Speaking Children's Early Language Development*. The Commercial Press, Beijing. 21

Zhang, Y., Zhang, L., Shu, H., Xi, J., Wu, H., Zhang, Y., and Li, P. (2012b). Universality of categorical perception deficit in developmental dyslexia: An investigation of Mandarin Chinese tones. *Journal of Child Psychology and Psychiatry*, 53(8):874–882. 17

Zheng, H., Wong, L. L., and Hickson, L. (2022). Barriers to hearing aid adoption among older adults in mainland China. *International Journal of Audiology*, pages 1–12. 164

Zheng, Q. (2014). Effects of Vowels on Mandarin Tone Categorical Perception. *Acta Psychologica Sinica*, 46(09):1223–1231 (in Chinese). 29

Zheng, W., Lu, X., and Guo, Y. (2016). A longitudinal study of the consonant and vowel recognition scores in 60 hearing-impaired children with cochlear implants (in Chinese). *Chinese Scientific Journal of Hearing and Speech Rehabilitation*, 14(3):184–187 (in Chinese). 29, 110

Zhou, N., Huang, J., Chen, X., and Xu, L. (2013). Relationship between tone perception and production in prelingually-deafened children with cochlear implants. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, 34(3):499. 26

Zhou, Q., Chen, Y., Chen, Y., Zhang, H., Wei, J., and Dang, J. (2016). Tongue performance in articulating Mandarin apical syllables by prelingual deaf adults using ultrasonic technology: Two case studies. In *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pages 1–5. 27

Zhou, X., Chen, ., Qian, Y., and Guo, M. (2008). Time and frequency domains in the tone of Mandarin monosyllabic words using Cool Edit Pro 2.0 software. *Journal of Clinical Rehabilitative Tissue Engineering Research*, 12(35):6827–6830 (in Chinese). 15

Zhu, H. (2002). *Phonological Development in Specific Contexts: Studies of Chinese-speaking Children*. Number 3 in Child Language and Child Development. Multilingual Matters, Clevedon : Buffalo, N.Y. 21

Zhu, H. and Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, 27(1):3–42. 21

Zhu, S., Wong, L. L., and Chen, F. (2014a). Tone identification in Mandarin-speaking children with profound hearing impairment. *International Journal of Pediatric Otorhinolaryngology*, 78(12):2292–2296. 25

Zhu, S., Wong, L. L., and Chen, F. (2014b). Tone identification in Mandarin-speaking children with profound hearing impairment. *International Journal of Pediatric Otorhinolaryngology*, 78(12):2292–2296. 112