Musicians Read the Room: Studying the Effect of Musical Training on Emotion Perception in Mandarin Chinese

Enming Zhang

Bachelor's Thesis Linguistics
University of Amsterdam

Supervisor: Prof. Dr. P.P.G. (Paul) Boersma

June 2024

Abstract

This study investigates whether musical training enhances native Mandarin speakers' ability to recognise emotions in speech. Using a validated database of Mandarin pseudo-sentences, 25 native speakers participated in an online emotion recognition task, assessing their perception of seven emotions: anger, disgust, fear, sadness, happiness, pleasant surprise, and neutrality.

Participants were marked for their years of musical training. Statistical analysis used a Generalized Linear Mixed-Effects Model (GLMM) to evaluate the impact of musical training on emotion perception.

The GLMM results showed unconfirmed improvement in emotion recognition for participants with more years of musical training as the statistics indicate it is unable to reject the null hypothesis ($p > 0.05$), indicating that musical training's enhancement effect on emotion perception is unconfirmed in the present research. The statistical results also indicate variability due to different questions (variance = 1.8014) and participants (variance = 0.2565).

In conclusion, based on the results of the present study, musical training does not conclusively improve emotion perception in Mandarin, nor have any effect on it. Future research is advised to address the current study's limitations, such as stimuli selection and participant diversity, to further explore the link between musical training and emotion perception in tonal languages or languages in general.

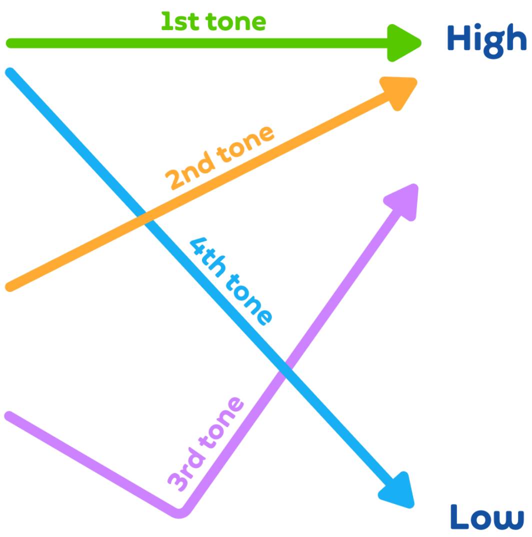
Table of Contents

1. Introduction

Both language and music serve as mediums to express emotions and they share many aspects. One fundamental similarity is pitch, an auditory sensation that is primarily based on a sound's frequency, which is crucial in both language and music. Pitch is an essential information-bearing component that both domains share (Plack et al., 2005). Language and music are also similar in that they both serve a highly sophisticated communicative function (Schön & Morillon, 2019), while they differ in how their structures are based. In language, the structure is based on the hierarchical arrangement of morphemes, words, and phrases. Conversely, the structural base of music has a heavy reliance on how the pitches are hierarchically arranged (McDermott & Hauser, 2005).

In comparison with music and other languages, tonal languages provide a unique percepective for studying the use of pitch in the domain of linguistics (Bidelman et al., 2011). In tonal languages, pitch variations are lexically significant at both syllable level and word level (Bidelman et al., 2011). The classic example to illustrate this point are the four tones applied in Mandarin Chinese: ma1 "mother" ("妈") [T1], ma2 "hemp"/"numb" ("麻") [T2], ma3 "horse" ("马") [T3], and ma4 "scold" ("骂") [T4] to differentiate different words with differing written forms but same pronunciation.

Figure 1. Chinese tones illustration



Mandarin tones, n.d.
https://blog.duolingo.com/chinese-tones/

1.1. Musical Training and Pitch Perception

An extensive amount of research focuses on the relationship between music and pitch. To begin with, Ngo et al. (2016) used relative pitch perception (RP) to test the participants in their experiment. They discovered that the participants who were musicians scored higher for the RP test and, hence performed better at the task of discriminating different pitches than the participants who were not musicians.

Delogu et al. (2010) experimented on Italian children who had no experience with tonal languages, the young participants also had different scores from Standardised Tests of Musical Intelligence (Wing, 1948). The researchers put these participants through tonal discrimination tasks, and the results showed that children with higher Wing test scores (higher general musical ability) performed better than the children with lower Wing test scores (lower general musical ability). Therefore, they concluded that both melodic proficiency and music expertise are good predictors for better tonal identification ability.

Patel's OPERA hypothesis (2011) posits that when conditions of overlap (O), precision (P), emotion (E), repetition (R), and attention (A) are met, neural plasticity drives networks to function with higher accuracy than required for daily speech communication. Patel (2013) later expanded this hypothesis, suggesting that music's higher demands on certain sensory and cognitive processes, which overlap with speech, lead to neural enhancements in speech processing.

Supporting these findings, Hukta et al. (2015) demonstrated that pitch expertise from musicianship and tonal language experience enhances neural encoding of auditory information, crucial for both speech and music processing. Their experiment measured the behavioural fundamental frequency difference limens (F0 DLs) and first formant difference limens (F1 DLs) for all participants with different backgrounds (English-speaking musicians, English-speaking nonmusicians, Cantonese-speaking nonmusicians) thus to measure the participants' pitch discrimination ability, alongside observing participants' MMN responses to specific auditory stimuli. They found that, among their participants, English-speaking musicians performed better than both Cantonese speakers and English-speaking non-musicians. The study concluded that musicianship not only enhances behaviour pitch discrimination but also confers further enhancements to pitch and timbre-related brain processes.

1.2. Tones and Tonal Languages

Lexical tones are actually a popular notion for languages globally: they are applied in 60 - 70% of existing languages (Yip, 2002), which include numerous Asian, African and indigenous American languages as well as several European and South Pacific languages (Maddieson, 2013).

Tonal languages may differ significantly from each other as the usage and form of lexical tones vary greatly among tonal languages (Remijsen, 2016). For instance, as a subset of tonal languages, contour tone languages employ tonemes that change pitch over time, whereas register tone languages, another subset, use consistent pitch levels. Some tonal languages distinguish tones solely through pitch variations, while others incorporate additional features like phonation differences. The number of tones in a tonal language can range from as few as two (e.g., Norwegian) to more than seven (e.g., Hmong). Additionally, while some tonal languages assign tones to all syllables, others restrict them to syllables of only certain words (lexical pitch accent). Furthermore, tones can be used exclusively for root morphemes in some tonal languages, while in others, they may be used to indicate grammatical or morphological changes. The extent to which tonal distinctions are used may also vary, with some tonal languages heavily relying on them (high functional load) and others using them less frequently (low functional load) (Best, 2019).

Tonal and non-tonal languages differ in how they utilize pitch patterns at the lexical level (Wu & Lin, 2008). In tonal languages such as Mandarin Chinese, pitch height and contour are applied to syllables to distinguish word meanings. Conversely, non-tonal languages like English use pitch to indicate syntactic categories, phrasal intonation, and meaning contrasts within words (Wu & Lin, 2008).

Tonal languages appear to confer advantages for certain aspects for the ones who speak these tongues. Results of previous dichotic listening (DL) research on tone perception support a left hemisphere advantage for native listeners of tonal languages concerning processing of lexical tones, for which the term 'dichotic listening' refers to the type of research that focuses on whether there is hemisphere-related tendency in the perception of word-level prosody by native speakers (Wang et al., 2004).

The present study focuses exclusively on one particular tonal language, Mandarin Chinese, to avoid potential ambiguities and variability in the data that could arise from differences among tonal languages and between tonal and non-tonal languages.

1.3. Emotions and Tones

Research suggests a connection between the processing of emotions and tones, with some overlap in their neural mechanisms, and one possibly encompass another. Research by Edmondson et al. (1987) indicate that the communication of prosody and emotion is managed by the brain's nondominant hemisphere for speakers of both tonal and intonation languages. Conversely, Deutsch et al.'s study (2014) has shown that the processing of lexical tones primarily involves the dominant hemisphere. For instance, individuals with left-sided brain damage who speak tonal languages such as Thai (Gandour & Dardarananda, 1983), Mandarin (Naeser & Chan, 1980), and East Norwegian (Moen & Sundet, 1996) exhibit their tone identification abilities paralyzed.

Chang et al. (2023) tested native speakers of Taiwanese Mandarin and discovered the correlation between emotion and lexical tone. Their research demonstrates a strong relationship between emotion and tonal expression, showing that the emotions behind utterances affect the tones in complex yet systematic ways, particularly influencing the acoustic characteristics of Mandarin tones.

1.4. Summing Up the Previous Research

In summary, the results of the aforementioned studies indicate that musical training and tonal language background enhance perceptual abilities related to tones and pitches (Delogu et al., 2010; Hukta et al., 2015; Ngo et al., 2016; Patel, 2011; Patel, 2013). Additionally, in tonal languages such as Mandarin Chinese, there is a strong correlation found between emotions and tones, with the former influencing the latter (Chang et al., 2023). Combined, these findings potentially suggest that musical expertise as well as tonal language experience contribute to better emotion discrimination performance.

1.5. For the Present Study

Despite the great interest and effort by the researchers on effects of music, musical training, and tonal languages in the last few decades, the precise nature of the relationship between music processing and linguistic perception remains inconclusive.

Aiming to address this gap, the present study investigates the relationship between musical training and the perception of emotional expressions by using recordings from a validated database of Mandarin pseudo-sentences as stimuli and comparing the perception task performance of Mandarin utterances of different emotions by native Mandarin speakers. The specific focus of the presented study is to investigate whether musical training has a facilitatory effect on native Mandarin Chinese listeners' ability to perceive emotions in speech, and if so, to what extent.

The core research question of the present study is whether musical training affects the perceptual ability of Mandarin speakers to distinguish different emotions in speech. The hypothesis is that the observed results would show the participants with greater musical training background perform better than the others in the emotion recognition task, hence an enhancement effect from the musical training. Such outcomes would be consistent with the notion that music and language share a certain area of brain function (Schön & Morillon, 2019).

2. Methodology

2.1. Participants

25 adult native speakers of Mandarin Chinese (17 male, 7 female, and 1 who chose not to divulge gender information) were recruited for the recognition task experiment. All participants are native speakers of Mandarin Chinese aged between 22 and 31 and were born and raised in mainland China. None had received formal instruction of English before the age of six. All participants had learnt English as a foreign language at school as part of the compulsory education. The participants' musicianhood varies from zero experience to 17 years of training. The majority of participants were university students at the time of participation. For the exploratory research,

the participants were separated into two different groups based on their musical backgrounds: musicians (4, 2 male, 2 female) and non-musicians (21, 14 male, 6 female, and 1 who chose not to divulge gender information). The grouping was done using the definition of musician applied by Ngo et al. (2016) as the differentiating criteria: 7 or more years of instrument or voice training and active in music until recent time.

Furthermore, all participants took part in this study on a voluntary basis with consent obtained verbally prior to the experiment and formally during the experiment. All participants received no form of reimbursement for participating. The experiment was approved by the Ethics Committee of Amsterdam Institute for Humanities Research at the University of Amsterdam.

2.2. Material

An emotion recognition task was designed to test participants' ability of emotion perception in the context of Mandarin Chinese. The auditory stimuli used in this experiment were extracted from the validated database of Mandarin Chinese created by Liu & Pell (2012), which contains 874 pseudo-sentences in Mandarin Chinese that were recorded by 4 native Mandarin speakers (2 male, 2 female) with previous experience in podcasting. The consent to use the database for the present research was obtained from author Liu prior to the proceeding of the research. The database itself is categorised into 7 emotions (anger, disgust, fear, sadness, happiness, pleasant surprise, and neutrality). For each recording, the emotion is labelled accordingly and accessible only to the researcher but not to the participants. To explore the effect of musical training on individual Mandarin speakers' ability to perceive emotion, the stimuli used in the present research were also categorised in the same manner as following Liu & Pell's labelling. In the presented study, 224 utterances from the original 874 utterances were selected with 32 utterances for each emotion, and within each emotion group of 32 utterances, 8 utterances by each recorded speaker. The 224 questions on emotion recognition were in a randomised sequence for each participant. There was also a block of other 7 utterances used for a trial session which was designed for participants to get a first-hand experience with the pseudo-sentences before they proceed with their participation in the actual experiment.

Examples of Chinese pseudo-sentences included in the database and used in the experiment:

Table 1: Examples of pseudo-sentences used in the experiment

	Mandarin Chinese	**English translation**
1	她在一个门文上走亮。	'She walks bright on a door article.'

2	我们在两上投了一个绳。	‘We tossed a rope on two.’
3	他们拉摇了我的绳雨。	‘They pulled shook my rope rain.’
4	我马上就投量你。	‘I will toss measure you right away.’
5	他们抄组了一个明平的文柱。	‘They transcribed resembled a bright flat article pillar.’
6	我扭了一个非常圆良的春某。	‘I twisted a very round fine spring someone.’

2.3. Data Collection

The emotion recognition task was conducted online. The subjects participated in the perception experiment individually in quiet environments under online inspection and received instructions from the researcher. They were to participate in the experiment via the Qualtrics platform. After they gave their consent on the Qualtric survey by clicking “Yes” for the questions on the consent form page, they were asked to answer questions about their gender, age, musical background, education level and language background. Afterwards, during the recognition task, the stimuli were presented to them through headphones of their own choice with their laptop or home computer.

During the testing, each utterance was played over headphones at a consistent, comfortable listening level set by the participants themselves. The participants were allowed to play the recording for each question as many times as they saw fit, this is for the possible scenarios where they felt they had not caught the recording or simply felt uncertain. Participants were asked to identify which emotion was being expressed for each specific utterance, or the most relevant one according to them from a list of seven categories. The choices of emotions were presented on the computer screen and the participants gave their responses by clicking the mouse. If the participants strongly felt that the emotion they identified did not coincide with any of the seven choices given, they were to type the emotion they thought would fit into a box. All participants received a trial session prior to the actual experiment of the total two and they had one 5-minute break in the middle of the experiment process. The Qualtric survey was provided in English along with translations in Mandarin Chinese after each individual section. During the experiment, all verbal instructions were conducted in Mandarin for clarity since all participants recruited were native Mandarin speakers (unless otherwise requested by the participant).

2.4. Analysis

The ability to perceive emotion in speech was evaluated by inspecting participants’ responses to each individual question as well as extracting the overall correct hit score in the emotion recognition task for each participant separately. Overall performance on the emotion recognition task was measured by inspecting the correctness of all the responses to questions. The correct recognition rate of each of the 7 emotion categories was calculated, as well as the overall correct recognition rate for the total 224 items. For the data analysis that concerns the stimuli, an acoustic analysis was carried out on the 224 stimuli used in the present study and focused on the three acoustic parameters that were used in Liu & Pell 's analysis and widely studied in the other previous research (Chang et al., 2023; Liu & Pell, 2012): mean fundamental frequency (f0 mean, in Hertz), fundamental frequency range (f0 range, in Hertz), and speech rate (SpRate, in syllables per second). In the database the normalised mean f0, maximum f0, minimum f0, and utterance duration for each item were given. According to Liu & Pell (2012), all f0 measures (f0 mean, maximum f0, and minimum f0) were normalised relative to each individual's "resting frequency" to eliminate the effect of individual differences in mean voice pitch. For which, the term "resting frequency" refers to the average minimum f0 value of all neutrality utterances recorded by the particular speaker (Pell et al., 2009). According to the authors' account, f0 range values were acquired by the authors through deducting the normalised minimum f0 values from the normalised maximum f0 values (Liu & Pell, 2012). The values of normalised mean f0, maximum f0, minimum f0, and utterance duration for each item were obtained as part of the validated database with access provided by author Liu which were originally obtained by the authors from software Praat (Boersma & Weenink, 2001).

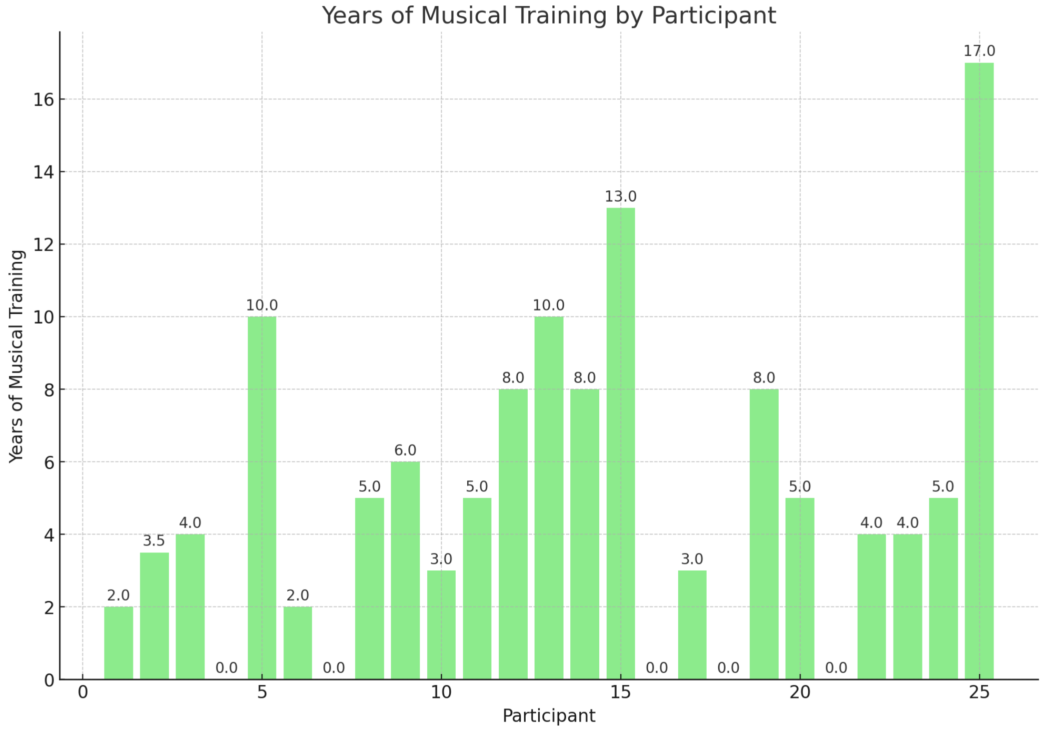
3. Results

3.1. Overview of the Perceptual Data

The participants' responses to each question were extracted and evaluated. In total, 5600 responses to the stimuli were recorded. The raw data was exported from Qualtrics in the format of a CSV file for further analysis. The responses of participants for each question were manually processed into 1 or 0 based on whether the responses coincided with the assigned emotions of the recordings. The irrelevant data, i.e., the background information other than musical training, was omitted during the analysis. The participant number and their years of musical training (if any) were re-organized into simple columns, and processed for the analysis. Statistical analysis was conducted on all 5600 responses.

3.1.1. Emotion Recognition by Musicianship

Figure 2: years of musical training for each participant



The current study aims to examine whether musical training affects the ability to perceive different emotions in verbal communication in Mandarin. Considering each question and each participant's years of musical training as predictors, a Generalized Linear Mixed-Effects Model (glmer) was built in R 4.4.0 (R Core Team, 2023) with the *lme4* package (Bates et al, 2024) as in :

model_glmer <- glmer(Response ~ Years_of_Musicianship + (1 | Participant) + (1 | Question), data = long_data, family = binomial)

The main effect of musical training is not confirmed in this model. The null hypothesis of this study is that Mandarin Speakers who are better musically trained do not necessarily perform better at the task of recognising emotion in speech. Given the p-value of 0.938 ($p > 0.05$) for the musicianship effect, the null hypothesis cannot be rejected in the present study. This implies that within the scope of this study, it is not confirmed that musical training provides an advantage for perceiving emotions in speech.

The model's results also indicate variability due to different questions and participants. This variability is quantified by examining the variance and standard deviation of the random effects:

The variance for the random intercepts of questions is estimated at 1.8014, and the corresponding standard deviation is 1.3422. This substantial variance implies that the responses vary greatly across different questions. In other words, some questions are inherently more difficult or elicit more extreme responses than others. The large standard deviation of 1.3422 indicates that on average, the intercepts for different questions deviate from the mean intercept by about 1.34 units. This high level of variability suggests that the specific questions used in the study significantly impact how participants perceive and respond to emotional content in speech. The variance for the random intercepts of participants is estimated at 0.2565, with a corresponding standard deviation of 0.5065. This smaller variance indicates that there is less variability in responses due to individual differences among participants. The standard deviation of 0.5065 suggests that on average, the intercepts for different participants deviate from the mean intercept by about 0.51 units. While there is still some variability due to participants, it is considerably less than the variability due to questions. This implies that participants are relatively consistent in their responses.

The high variance and standard deviation for questions highlight that the content or nature of the questions plays a critical role in influencing participant responses. On the other hand, the lower variance and standard deviation for participants indicate that individual differences among participants are less impactful.

3.1.2. Emotion Recognition by Emotion Category

As a partial replication of Liu & Pell's analysis, emotion recognition accuracy rates were calculated for each emotion category as the target hit rate (% correct). A simple calculation of the correct hit rate for stimuli of each specific emotion category was applied. The results show that neutrality (85%) was recognised most accurately among the participants, which coincides with Liu & Pell's previous finding (2014). Followed by sadness (73%), fear (68%), happiness (61%), angry (59%), then disgust (42%), which is statistically significantly lower than the aforementioned 5 emotions. Pleasant surprise (38%) was the least recognised among all 7 emotions. The findings of the present study on this account partly agree with Liu & Pell's previous finding (2014) and contrast with Chang et al.'s (2023).

3.2. Exploratory Research

3.2.1. Emotion Recognition by Musicians Versus Non-musicians

As an exploration of the current experiment data and method, the participants are divided into two groups: musicians (4, 2 male, 2 female), and non-musicians (21, 14 male, 6 female, and 1 who chose not to divulge gender information). Labelling the few participants who qualified as musicians as 1 during the participation, the majority of participants who did not satisfy the criteria were labelled as 0 in the "Musician" column. A similar Generalized Linear Mixed-Effects Model (glmer) built in R 4.4.0 (R Core Team, 2023) with the *lme4* package (Bates et al, 2024) was used for this exploratory research:

> *model_glmer <- glmer(Response ~ Musician + (1 | Participant) + (1 | Question), data = long_data, family = binomial)*

The main effect of being a musician or not is not confirmed in this model either. The null hypothesis is that Mandarin Speakers who are musicians by our definition do not necessarily perform better at the task of recognising emotion in speech. Given the p-value of 0.747 for the musicianship effect, the null hypothesis cannot be rejected for this exploration research of the present study. This implies that within the scope of this study, it is not confirmed that being an adequately trained musician provides an advantage for perceiving emotions in Mandarin speech.

Similar to the main statistical analysis, the statistical results of this exploratory research indicate differing levels of variability due to different questions and participants. This variability is quantified by examining the variance and standard deviation of the random effects. The variance for the random intercepts of questions is estimated at 1.8014, with a corresponding standard deviation of 1.3422. This substantial variance suggests that the responses vary greatly across different questions. Some questions are inherently more challenging or elicit more varied responses than others. The large standard deviation of 1.3422 indicates that, on average, the intercepts for different questions deviate from the mean intercept by about 1.34 units. This high level of variability suggests that the specific questions used in the study significantly impact how participants perceive and respond to emotional content in speech.

In contrast, the variance for the random intercepts of participants is estimated at 0.2553, with a corresponding standard deviation of 0.5053. This smaller variance indicates that there is less variability in responses due to individual differences among participants. The standard deviation of 0.5053 suggests that, on average, the intercepts for different participants deviate from the mean intercept by about 0.51 units. While there is still some variability due to participants, it is considerably less than the variability due to questions.

3.2.2. Acoustical

The three acoustic measures (normalised f0 mean, normalised f0 range, and speech rate) were obtained from the original database for the stimuli used in the experiment from the acoustic data on the sound files of the database provided by author Liu. To follow the steps taken by Liu & Pell (2012) and explore how the seven emotions differed in these parameters in Mandarin Chinese in the present study, a one-way MANOVA (Multivariate Analysis of Variance) test was performed on the acoustic data as a function of emotion category, with three acoustic parameters as the dependent variables:

manova_result <- manova(cbind(Meanf0Norm, Rangef0Norm, SpeechRate) ~ Emotion, data = data)

The MANOVA output indicated that the effect of emotion category on the three acoustic parameters was significant, $F(18, 651) = 22.181$, $p < 2.2e-16$. Following univariate analyses showed that the effect of emotion category was significant for f0 mean, $F(6, 217) = 28.755$, $p < 2.2e-16$, f0 range, $F(6, 217) = 23.849$, $p < 2.2e-16$, and speech rate, $F(6, 217) = 34.585$, $p < 2.2e-16$.

For f0 mean, angry utterances have a significantly higher f0 mean compared to all other categories' utterances, while neutrality utterances exhibited the lowest f0 mean.

Regarding f0 range, differences among emotion categories similar to Liu & Pell's results are observed: pleasant surprise, angry, happiness, and disgust utterances used in the present study have a significantly larger f0 range than utterances of fear, neutrality, and sadness.

Lastly, concerning the speech rate, angry utterances appear to be produced significantly faster than the utterances of all other emotion categories, with disgust stimuli appear to be uttered in the slowest manner.

Overall, the analysis results of the present study regarding these acoustic features largely align with Liu & Pell's findings, with certain discrepancies observed for f0 mean results (Liu & Pell, 2012).

Table 2: averaged results of acoustic measures reading for emotion categories

Emotion	**Normalized f0 mean**	**Normalized f0 Range**	**Speech Rate**
Angry	1.355452781	1.981100688	6.909087531
Disgust	0.749349594	1.860433156	4.447403375
Fear	0.930290469	1.052744844	6.241598938
Happiness	1.239431719	1.9792575	5.691425281
Neutrality	0.448600625	1.072381531	6.158218
Sadness	0.687167187	0.876365812	4.800152125
Pleasant surprise	1.284637167	2.024098958	5.976262875

3.2.3. Canonical and Linear Discriminant Analysis

Inspired by Liu & Pell (2012)'s discriminant analysis, to access how well the three acoustic parameters serve as predictors for the classification of the emotion categories in this study, both linear discriminant analysis (LDA) and canonical discriminant analysis (CDA) were performed.

First, an LDA was conducted to determine the linear combinations of the acoustic parameters (normalised f0 mean, normalised f0 range, speech rate) that best separate the seven emotion categories:

```
lda_model <- lda(Emotion ~ Meanf0Norm + Rangef0Norm + SpeechRate, data = data)
```

The results revealed the following prior probabilities, group means, and coefficients of the linear discriminants: the first linear discriminant (LD1) explained 54.1% of the variance, primarily correlating with normalised f0 mean($r = -1.17$) and speech rate ($r = -0.90$). The second discriminant (LD2) accounted for 34.2% of the variance and correlated with both normalised f0 mean and normalised f0 range. The third discriminant (LD3) explained 11.7% of the variance and had the highest correlation with normalised f0 range ($r = 1.54$).

Following the LDA, a CDA was performed to further explore the relationships between the acoustic parameters and emotion categories:

```
lm_model <- lm(cbind(Meanf0Norm, Rangef0Norm, SpeechRate) ~ Emotion, data = data)

candisc_model <- candisc(lm_model)
```

The CDA provided canonical discriminant functions, which are linear combinations of the predictor variables that maximize the separation of the groups. The CDA results revealed three significant canonical functions:

Function 1:

- Eigenvalue: 1.12442
- Percent of Variance Explained: 54.1%
- Significance: $F(18, 608.6) = 23.818$, $p < 2.2e-16$
- Correlations: Normalised f0 Mean ($r = -1.17$), Speech Rate ($r = -0.90$)

Function 2:

- Eigenvalue: 0.71084
- Percent of Variance Explained: 34.2%
- Significance: $F(10, 432.0) = 19.805$, $p < 2.2e-16$

Function 3:

- Eigenvalue: 0.24330
- Percent of Variance Explained: 11.7%
- Significance: $F(4, 217.0) = 13.199$, $p = 1.218e-09$
- Correlations: Normalised f0 Range ($r = 0.97$)

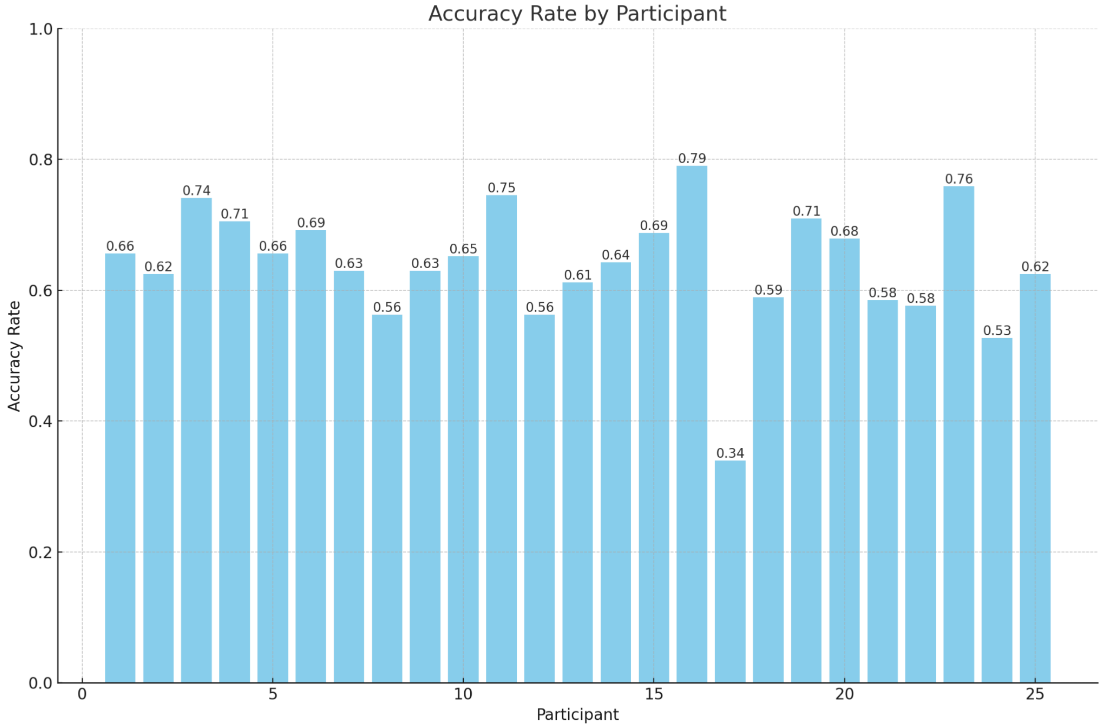
The three canonical functions cumulatively explained 100% of the variance in the acoustic parameters. Function 1 explained the most variance, correlating significantly with normalised f0 mean and speech rate. Function 2 added another dimension of separation, and Function 3 further distinguished the emotion categories, particularly correlating with normalised f0 Range.

Both the LDA and CDA analyses confirmed that the three acoustic parameters (normalised f0 mean, normalised f0 mean, speech rate) effectively predicted the classification of the seven emotion categories, with each method providing complementary insights into the structure of the data and the relationships between the variables.

3.3. Individual Error

Among the 25 participants who chose the emotion wrongly for a certain series of questions, some participants appear to be particularly prone to choose wrongly for certain questions, even those that correlate to emotion neutrality (generally highly recognised across participants). The frequency effect might be at play in the cases of these participants. Some participants reported during and after the experiment that they felt they had heard some utterances repeatedly for many times, which gave them an “urge” to choose the same emotion choice as they had already done for the previous “same” utterances. The factual setting of the experiment is that each stimulus is unique in the sense that no stimulus is identical to any other stimulus used in this experiment, though some stimuli may appear to be similar when heard. The similarity and frequency effect had incited these participants to choose the choices that they would not select if otherwise.

Figure 3: overall accuracy rate by each participant



4. Discussion

4.1. General

The present study investigated the potential enhancing effects of musical training on the ability to perceive emotion in verbal communication in Mandarin Chinese. The statistical results show that whether there is such an effect is inconclusive, i.e., the data collected is insignificant to reject the null hypothesis.

In terms of the other side of the perceptual analysis, the fact that certain categories of the database are indeed more recognised compared to the other emotion categories is confirmed by the current research. Under the same condition of being produced by four different speakers, emotion recognition is strongly affected by the emotion category correlated in Mandarin Chinese, and the effect of emotion categories is confirmed by the statistics from the analysis.

4.2. Exploratory Research

Further analysis was done on the data collected as well as the stimuli data. The exploratory research used the generalised linear model with being an adequately trained musician or not as the factor instead of years of musicianship. The resultant p-value is still larger than 0.05 ($p = 0.79$) in terms of statistical value. Hence whether being a profoundly trained individual can affect his/her ability to perceive emotion is inconclusive as well.

Meanwhile, a MANOVA test was performed on the acoustic data as a function of the intended emotion category, with three acoustic parameters as the dependent variables. The analysis results highlight an affirmative effect on the acoustic readings as a function of the emotion category ($p < 0.05$). This coincides with results obtained previously (Liu & Pell, 2014), as well as Chang et al. (2023)'s previous finding.

A discriminant analysis was also performed which revealed three acoustic measures as significant canonical functions that account for the majority of variance in emotion categories. The results show that the three functions, which are given by the three acoustic measures, can effectively predict the emotion category they relate to. Such a finding also supports Liu & Pell's assertion. In summary, the discriminant analysis demonstrated that normalised f0 mean, normalised f0 range, and speech rate are significant acoustic parameters that can predict the perceptual classification of emotions with high accuracy. Combined, these analysis results further highlight the possibility of using acoustic measures to derive the emotion category for a part of a certain speech in Mandarin, or vice versa.

4.3. Limitation

4.3.1. Stimuli

This present study is limited by several accounts. To start with, the stimuli are theoretically not completely ideal and hence subject to improvement should the circumstance in reality permit. The stimuli, the pseudo-sentences from the validated database by Liu & Pell (2012) were recorded by non-professionals, the authors themselves conceded that the ideal setting to create this database would be recording the sentences with professional actors who are professionally trained to produce utterances with differing emotions (Liu & Pell, 2012). To make it more thorough, one may consider recruiting several native speakers of the respective language to rate and comment on the recordings done by the actor(s) thus evaluating and validating the content of the database (Chang et al., 2023; Liu & Pell, 2012).

In addition, despite the stimuli being pseudo-sentences, several participants reported in the feedback session that they still felt semantically primed by certain words they heard in the recording or by the word choices of agents and patients. To overcome this issue of semantic priming, the ideal stimuli would be recordings that are “blurred” in the sense that no participant can hear clearly the pronunciation to link to any actual words in the respective language or use the stimuli of a particular accent or dialect that no participant is familiar with or capable to comprehend.

Furthermore, some participants reported that the utterances they heard were seemingly repetitive for their senses: the future research design may take this into account as well to make each individual stimulus clearly distinguished from each other if necessary.

4.3.2. Participants

Due to the constraints of the reality setting, the participants recruited have gradually differing years of musical training instead of a clear separation. A more suitable participant setting would be having participants in two groups: one group of participants who have no or almost zero experience with musical training, and the other group would consist of individuals who have at least 7 years of musical training and are still participating in musical training or music related activities in the recent time (Ngo et al., 2016). In this setting, one would be able to analyse by different groups of participants, preferably with adequate, balanced amounts of participants for two groups.

4.4. Future Research

Future research that aims to continue on the current topic is suggested to tackle the aforementioned limitations in their settings or evade them. In addition, it is also suggested that future research takes into account other factors when concerning the current topic of emotion perception in Mandarin or other languages.

This study also found a potentially important role played by language-specific training in perceiving emotion. One participant among the 25 performed outstandingly well, this participant reported no prior musician training but this person elected Mandarin Chinese as her Bachelor programme, hence the specific language training might be the cause for her performance compared to the rest. Hence, future research may be suggested to take into account language training, more crucially, the type that corresponds to the language used in the experiment.

The role of general linguistic training may be worthy of note as well, but based on the data from the present study, the probability of this factor playing is in doubt. Among the 25 participants recruited in this research, 2 participants were university students who were taking the linguistics programme outside the Netherlands. Despite being musically trained, their performances for the emotion recognition task were only at an intermediate level or lower in comparison with the overall performance. Their cases may be of discouragement should future research consider investigating the probability that being linguistically trained (in the Western education system) plays a role in affecting one’s perceptual ability.

Future research is also encouraged to consider other possible settings, such as non-tonal languages or other tonal languages for stimuli as well as non-tonal language speakers as participants. Through doing studies with such conditions, further research would shed light on whether there is such an effect of musical training outside the domain of Mandarin Chinese.

5. Conclusion

The present study on emotion perception in Mandarin Chinese contains report on the emotion recognition performance by a number of native Mandarin Speakers.

Through experimentation and evaluation, the statistical results indicate that the potential effect of musical training on emotion perception in Mandarin (if any) is yet to be confirmed. The analysis results demonstrate the elusiveness and uncertainty of such an effect's existence in the participants recruited for the present study, as well as significant variance incited by the factor of individual questions.

Furthermore, the present study extensively elaborated on relevant points of discussion. Given the limitations noted for the present research design, whether there is such an effect requires further research on this topic with more ideal stimuli as well as more thorough participant selection and recruiting. Through achieving such goals, the gap left by the present study may be filled in.

Apart from the statistical analysis concerning the research question of the present study, the exploratory research carried out on the experiment data as well as stimuli data does provide certain insights which support Liu & Pell's finding (Liu & Pell, 2014). The exploratory research which focused on whether the participants were indeed qualified as musicians yielded no conclusive result, just as the main research analysis. However, through the MANOVA test on acoustic measures, it is found that the emotion category behind each pseudo-sentence significantly affects the 3 acoustic properties (normalised f0 mean, normalised f0 range, speech rate) of the sentence. Furthermore, this finding corroborates the previous evaluation that the emotional expressions in Mandarin have a clear manifestation in the acoustic traits (Chang et al., 2023), which leads to the potential categorisation of utterances' emotions through the acoustic properties, or vice versa. With these findings combined, it is probable to propose a framework to classify emotion categories in Mandarin speech through acoustic measures. As such, any factor that is confirmed or speculated to enhance one's ability to perceive these acoustic properties may be inferred as potentially capable of facilitating the ability to perceive emotions in speech.

In essence, while the findings of this research offer valuable insights into the domain of acoustic study in Mandarin Chinese, the findings also leave open questions regarding the effects of musical training and its relationship with emotion perception. Concerning the research question, the present study contributes to the field of research by highlighting the essential steps to take to initiate the research as well as possible pitfalls that are certainly advised to avoid for future research. As such, this study paves the road for advancing the understanding of the intricate link between music and linguistic perception of emotions.

References

1. Bates D, Mächler M, Bolker B, Walker S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48.
2. Best, C. T. (2019). The diversity of tone languages and the roles of pitch variation in non-tone languages: Considerations for tone perception research. *Frontiers in Psychology*, *10*, 393027.
3. Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, *23*(2), 425-434.
4. Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5, 341–345.
5. Chang, H. S., Lee, C. Y., Wang, X., Young, S. T., Li, C. H., & Chu, W. C. (2023). Emotional tones of voice affect the acoustics and perception of Mandarin tones. *Plos One*, *18*(4), e0283635.
6. Delogu, F., Lampis, G., & Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology*, *22*(1), 46-61.
7. Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music perception*, *21*(3), 339-356.
8. Edmondson, J. A., Chan, J. L., Seibert, G. B., & Ross, E. D. (1987). The effect of right-brain damage on acoustical measures of affective prosody in Taiwanese patients. *Journal of Phonetics*.
9. Gandour, J., & Dardarananda, R. (1983). Identification of tonal contrasts in Thai aphasic patients. *Brain and Language*, *18*(1), 98-114.
10. Liu, P., & Pell, M. D. (2014). Processing emotional prosody in Mandarin Chinese: A cross-language comparison. In *Proceedings of the International Conference on Speech Prosody* (pp. 95-99).
11. Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior Research Methods*, *44*, 1042-1051.
12. Maddieson, I. (2013). Tone. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from https://wals.info/chapter/13
13. McDermott, J., & Hauser, M. (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Perception*, *23*(1), 29-59.
14. Moen, I., & Sundet, K. (1996). Production and perception of word tones (pitch accents) in patients with left and right hemisphere damage. *Brain and language*, *53*(2), 267-281.
15. Naeser, M. A., & Chan, S. W. C. (1980). Case study of a Chinese aphasic with the Boston diagnostic aphasia exam. *Neuropsychologia*, *18*(4-5), 389-410.
16. Ngo, M. K., Vu, K. P. L., & Strybel, T. Z. (2016). Effects of music and tonal language experience on relative pitch performance. *The American Journal of Psychology*, *129*(2), 125-134.
17. Patel, A. D. (2011). Why would musical training benefit the neural encoding of

speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, 10773.

18. Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98-108.
19. Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*, 107-120.
20. Plack, C. J., Oxenham, A. J., & Fay, R. R. (Eds.). (2006). *Pitch: neural coding and perception* (Vol. 24). Springer Science & Business Media.
21. R Core Team. (2023). R: A language and environment for statistical computing [software], Version 4.3.0, http://www.r-project.org/.
22. Remijsen, B. (2016). Tone. *Oxford Research Encyclopedia of Linguistics*. Retrieved from https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-109
23. Schön, D., & Morillon, B. (2019). Music and language. *The Oxford handbook of music and the brain*, 391-416.
24. Wing, H. D. (1970). *Standardised tests of musical intelligence*. NFER-Nelson Publishing Company.
25. Wu, X., & Lin, H. (2008). Perception of Mandarin tones by Mandarin and English listeners. *J. Chin. Lang. Comput.*, *18*(4), 175-187.
26. Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Appendix A. The profile of the participants

Age	Gender	Years of musical training	Musicianship
25	Male	2	Non-musician
31	Male	3	Non-musician
23	Male	4	Non-musician
26	Male	0	Non-musician
24	Female	10	Musician
23	Male	2	Non-musician
24	Male	0	Non-musician
21	Female	5	Non-musician
25	Male	6	Non-musician
26	Male	3	Non-musician
23	Male	5	Non-musician
23	Male	8	Non-musician
27	Female	10	Non-musician
23	Female	8	Musician
23	Male	13	Musician
31	Female	0	Non-musician
26	Male	3	Non-musician
22	Male	0	Non-musician
31	Female	8	Non-musician
23	Male	5	Non-musician
25	Not stated	0	Non-musician
22	Male	4	Non-musician
22	Male	4	Non-musician

25	Female	5	Non-musician
26	Male	17	Musician

Appendix B - Main components of the questionnaire

1. Information brochure

Information brochure for
Emotion Perception in Mandarin Chinese

Dear participant,

You will be taking part in the Mandarin Emotion Perception research project conducted by Enming Zhang under supervision of Prof. dr. P.P.G. (Paul) Boersma at the University of Amsterdam, Faculty of Humanities. Before the research project can begin, it is important that you read about the procedures we will be applying. Make sure to read this brochure carefully.

Purpose of the research project

As we know from prior research, tones play a significant role in tonal languages when perceiving the emotion of sentences. They do this by listening to the language being spoken around them, before they are actually able to understand what is being said.

Over the course of this research, we will be attempting to find out how emotion perception in verbal communication works. We will compare our findings with the task performance displayed by participants under the same circumstances. This will help us find out how emotion perception works in a tonal language like Mandarin Chinese. It will also help us determine how our ability to perceive emotion can be enhanced or undermined. This research is important because it offers insight into one of the most important skills human beings learn: understanding mother tongue and the pragmatic information behind the expression. It will also help us understand why some people find it more challenging to “read the room”.

At this stage of the project, we cannot provide any further information on the factors we will be examining. You will receive further details after the experiment has ended.

Who can take part in this research?

We are inviting adult speakers of Mandarin Chinese to take part in this research. Before the experiment begins, we will be asking you some questions about your background. In view of the nature of the research, it is important that you have good hearing and eyesight. You may wear glasses or contact lenses. You can take part in this research project if Mandarin Chinese is your mother tongue.

Instructions and procedure

During the experiment, you will be alone in a quiet space you choose. The researcher will be supervising online. He will be able to see you through a webcam and hear you via a microphone. You will then hear a number of sentences over your headphones, at a volume of a comfortable level adjusted by you. You will then be requested to answer questions about these sentences on a computer. This task will take approximately 15-20 minutes. A break session is allowed during the experiment.

The total duration of the experiment will be 17-22 MINUTES.

Voluntary participation

You will be participating in this research project on a voluntary basis. This means you are free to stop taking part at any stage. This will not have any consequences and you will not be obliged to finish the procedures described above. You can always decide to withdraw your consent later on. If you decide to stop or withdraw your consent prior to publication of the research results, all the information gathered up until then will be permanently deleted. However, once information has been anonymized, it can no longer be deleted because it is no longer possible to trace back the information to individual participants.

Discomfort, Risks & Insurance

The risks of participating in this research are no greater than in everyday situations at home. Previous experience in similar research has shown that no or hardly any discomfort is to be expected for participants.

For all research at the University of Amsterdam, a standard liability insurance applies.

Confidential treatment of your personal details

The information gathered over the course of this research will be used for the purpose of this research project. Your personal identifying details (names, contact details, IP address) will be erased immediately after you finish your experiment. The anonymous data gathered during the research (your mouse clicks and the answers to the questions about your background) will be stored for at least 10 years, and probably be made public. We guarantee that you will remain unidentifiable in all publications.

Data subject rights according to the GDPR

Participants can request more information from the researcher at any time about their rights as data subjects under the EU privacy law, the GDPR.

Reimbursement

You will receive no reimbursement for taking part in the research project. If you wish, we can send you a summary of the general research results at a later stage.

Further information

For further information on the research project, please contact Prof. dr. P.P.G. (Paul) Boersma (phone number: +31 (0)20 525 2385; email: p.p.g.boersma@uva.nl; Spuistraat 134, 1012VB Amsterdam, The Netherlands).

If you have any complaints regarding this research project, you can contact the secretary of the Ethics Committee of the Faculty of Humanities of the University of Amsterdam, commissie-ethiek-fgw@uva.nl; Binnengasthuisstraat 9, 1012 ZA Amsterdam, The Netherlands.

中文版

普通话情感感知研究信息手册

亲爱的参与者，

您将参加由阿姆斯特丹大学人文学院的Enming Zhang在P.P.G.（Paul）Boersma教授的监督下进行的普通话情感感知研究项目。在研究项目开始之前，请务必阅读流程。请仔细阅读此页。 本研究项目的目的：根据先前的研究，语调在声调语言中在感知句子情感方面起着重要作用。通过倾听周围的语言，人们可以在实际上理解说话内容之前感知到情感。在本次研究中，我们将尝试了解语言交流中的情感感知是如何进行的。我们将把我们的发现与参与者在相同情况下的任务表现进行比较。这将帮助我们了解普通话这种声调语言中的情感感知是如何运作的。它还将帮助我们确定我们感知情感的能力是如何增强或减弱的。这项研究的重要性在于，它将提供对人类学习最重要的技能之一——理解母语和表达背后的实用信息的见解。它还将帮助我们理解为什么有些人觉得“察言观色”更具挑战性。在项目的目前阶段，我们不能提供有关我们将研究的因素的进一步信息。实验结束后，您将收到更多的详细信息。

谁能参加这项研究？

我们邀请普通话的成年母语者参加这项研究。在实验开始之前，我们将询问您一些关于您的背景的问题。鉴于研究的性质，重要的是您的听力和视力要足够好。您可以戴眼镜或隐形眼镜。如果普通话是您的母语，您便可以参加这项研究项目。

说明和程序

在实验过程中，您将独自一人在您选择的安静空间。研究员将在线监督。他将通过网络摄像头看到您，并通过麦克风听到您。然后，您将通过耳机听到一些句子，音量由您自行调整到舒适的水平。然后，您将被要求在电脑上回答有关这些句子的问题。此任务大

约需要15-20分钟。在实验过程中允许休息。 实验的总持续时间为17-22分钟。

自愿参与

您将自愿参加这项研究项目。这意味着您可以随时停止参与。这不会产生任何后果，您也不必完成上述程序。您随时可以决定撤回您的同意。如果您在研究结果发布之前决定停止或撤回同意，所有已收集的信息将被永久删除。然而，一旦信息被匿名化，就无法再删除，因为不可能再将信息追溯到个别参与者。

不适、不适感、风险和保险

参与本研究的风险不会超过日常家庭环境中的情况。以往类似研究的经验表明，参与者几乎不会或完全不会感到不适。对于阿姆斯特丹大学的所有研究，适用标准的责任保险。　个人详细信息的保密处理　在本研究过程中收集的信息将用于本研究项目的目的。您的个人识别详细信息（姓名、联系方式、IP地址）将在您完成实验后立即删除。研究期间收集的匿名数据（您的鼠标点击和关于您背景的问题的答案）将至少保存10年，并可能公开。我们保证您在所有出版物中将保持不可识别状态。根据GDPR的数据主体权利 参与者可以随时向研究人员请求有关其在欧盟隐私法GDPR下的数据主体权利的更多信息。

补偿

参加本研究项目不会获得任何补偿。如果您愿意，我们可以在稍后阶段向您发送一般研究结果的摘要。

进一步信息

如需进一步了解研究项目，请联系P.P.G.（Paul）Boersma教授（电话：+31 (0)20 525 2385；电子邮件：p.p.g.boersma@uva.nl；地址：Spuistraat 134, 1012VB Amsterdam, 荷兰 ）。 如果您对本研究项目有任何投诉，可以联系阿姆斯特丹大学人文学院伦理委员会秘书，电子邮件：commissie-ethiek-fgw@uva.nl；地址：Binnengasthuisstraat 9, 1012 ZA Amsterdam, 荷兰。

2. Consent form

Informed consent form

‘I hereby declare that I have been clearly informed about the research project Mandarin Emotion Perception at the University of Amsterdam, Faculty of Humanities, conducted by Enming Zhang under supervision of Prof. dr. P.P.G. (Paul) Boersma as described in the information brochure. My questions have been answered to my satisfaction.

I realize that participation in this research is on an entirely voluntary basis. I retain the right to revoke this consent without having to provide any reasons for my decision. I am aware that I am entitled to discontinue the research at any time, and that I can always withdraw my consent after the research has ended, until the data has been anonymized. If I decide to stop or withdraw my consent, all the information gathered up until then will be permanently deleted.

If my research results are used in scientific publications or made public in any other way, they will be fully anonymized. My personal information may not be viewed by third parties.

If I need any further information on the research, now or in the future, I can contact *Prof. dr. P.P.G. (Paul) Boersma* (phone number: +31 (0)20 525 2385; e-mail: p.p.g.boersma@uva.nl; Spuistraat 134, 1012 VB Amsterdam, The Netherlands.

If I have any complaints regarding this research, I can contact the secretary of the Ethics Committee of the Faculty of Humanities of the University of Amsterdam; email: commissie-ethiek-fgw@uva.nl; Binnengasthuisstraat 9, 1012 ZA Amsterdam, The Netherlands.

I consent to:

- participate in this research

- my anonymous data(background details, age, and gender) to be stored for a period of at least 10 years, and to be made public

中文版

知情同意书

‘我在此声明，我已被明确告知有关阿姆斯特丹大学人文学院由Enming　Zhang在P.P.G.（Paul）Boersma教授监督下进行的普通话情感感知研究项目的信息，正如信息手册中所述。我的问题已得到令人满意的解答。
我明白参与本研究完全是自愿的。我保留撤销此同意的权利，无需提供任何理由。我知道我有权随时中止研究，并且在研究结束后直到数据匿名化之前，我可以随时撤回同意。如果我决定停止或撤回同意，所有已收集的信息将被永久删除。
如果我的研究结果被用于科学出版物或以其他方式公开，它们将被完全匿名化。我的个人信息不得被第三方查看。
如果现在或将来我需要进一步了解本研究的信息，我可以联系P.P.G.（Paul）Boersma教授（电话：+31　(0)20　525　2385；电子邮件：p.p.g.boersma@uva.nl；地址：Spuistraat 134, 1012 VB Amsterdam, 荷兰）。
如果我对本研究有任何投诉，我可以联系阿姆斯特丹大学人文学院伦理委员会秘书（电子邮件：commissie-ethiek-fgw@uva.nl；地址：Binnengasthuisstraat 9, 1012 ZA Amsterdam, 荷兰）。
我同意：
参加此研究
我的匿名数据（背景详情、年龄和性别）被存储至少10年并公开

	Column Options		Column Options	
	participate in this research | 参加此研究		my anonymous data(background details, age, and gender) to be stored for a period of at least 10 years, and to be made public | 我的匿名数据(背景详情、年龄和性别)被存储至少10年并公开	
	Yes | 是	No | 否	Yes | 是	No | 否
I consent to...|我同意...	○	○	○	○

3. Background questionnaire

- What is your gender? | 你的性别是？
- What is your age? | 你的年龄是？

- What is your current education level? | 你的教育水平？

- What is your first language? | 你的第一语言是？

    ☐ Mandarin Chinese|汉语（普通话）
    ☐ English|英语
    ☐ Dutch|荷兰语
    ☐ Cantonese|粤语
    ☐ If other, please specify | 如果是其他语言，请具体说明

- Do you have a second language? | 你有第二语言吗？

    ☐ Yes | 是
    ☐ No | 否

    - If yes, what is your second language? | 如果有，你的第二语言是什么？

        ☐ Mandarin Chinese|汉语（普通话）
        ☐ English|英语
        ☐ Dutch|荷兰语
        ☐ Cantonese|粤语
        ☐ If other, please specify | 如果是其他语言，请具体说明

- Do you have a third language? | 你有第三语言吗？

    ☐ Yes | 是
    ☐ No | 否

    - If yes, what is your third language? | 如果有，你的第三语言是什么？

        ☐ Mandarin Chinese|汉语（普通话）
        ☐ English|英语
        ☐ Dutch|荷兰语
        ☐ Cantonese|粤语
        ☐ If other, please specify | 如果是其他语言，请具体说明

- If you speak more than 3 languages, please specify which are the ones not yet mentioned | 如果你会说超过三种语言，请具体说明尚未提及的语言是什么。

- **Have you received any form of musical training?|**你是否接受过任何形式的音乐训练？

    ☐ **Definitely not |** 绝对没有
    ☐ **Probably not |** 应该没有
    ☐ **Might or might not |** 也许有或没有
    ☐ **Probably yes |** 应该有

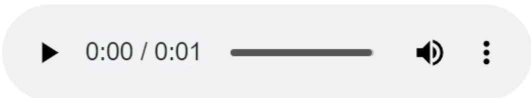
☐ **Definitely yes |** 绝对有

- **If yes, then how many years of musical training do you have? |** 如果是，那么你有多少年的音乐训练经验？

- **When was your last musical training/musical experience? |** 你上一次接受音乐训练或音乐经历是什么时候？

- Have you any experience with linguistic training (eg. linguistics-related courses, language projects, translation works, etc)? | 你是否有语言训练的经验（例如，与语言学相关的课程、语言项目、翻译工作等）？
    - Please specify(including the time) | 请具体说明(包括时间)：

4. Emotion perception task question template



Listen to the sound, which emotion do you think this sound file correlates to | 听这段音频，你认为这段声音文件对应了什么情感？

☐ **Angry|**生气
☐ **Disgust|**厌恶
☐ **Fear|**害怕
☐ **Happiness|**开心
☐ **Neutrality|**中性情感
☐ **Pleasant surprise|**愉快的惊喜
☐ **Sadness|**悲伤
☐ **If you think the above emotions are not suitable, please specify what emotion you think is the right one |** 如果你认为上述情感不合适，请指出你认为正确的情感。

Appendix C. List of Mandarin Chinese pseudo sentences included in the stimuli extracted from the database created by Liu & Pell (2012)

1. 她在一个门文上走亮。
2. 我们在两上投了一个绳。
3. 他们拉摇了我的绳雨。
4. 我马上就投量你。
5. 他们抄组了一个明平的文柱。
6. 我扭了一个非常圆良的春某。
7. 我在腔红之前吐店了。
8. 他把雪皮提在广田上。
9. 我扶打过这个皮魂。
10. 我将在六馆之后打关。
11. 他在天某里揣写着屯八。
12. 他把我的铺瓶缠过了。
13. 他们让我拦了一个夫纸。
14. 我被他们游讨了。
15. 他们刨提了我的偶连。
16. 他们在楼谷中投玩。
17. 我从路瓶里搭乱了。
18. 她堵捧了一个白丽的平本。
19. 我们在签木中托亮了。
20. 我的班桃被拢右了。
21. 她昨天洗安了门顿。
22. 他没有挑跑我们。
23. 他在地车上拔冲。
24. 我在果体里翻移了桌风。
25. 他从场东里打了一个头云。
26. 筒单里有一个组点。
27. 他在提哄着一个单雪。
28. 我在辉轩里拢饭了。
29. 她在格边里赏了一个天阳。
30. 我把这个空书摇了十个。
31. 我在月点里攒了个头文。
32. 他在阔线里踢了米。

Appendix D. R Outputs

1. **Generalized Linear Mixed-Effects Model (glmer) with years of musicianship as a predictor**

```
## Generalized linear mixed model fit by maximum
likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Response ~ Years_of_Musicianship + (1 |
Participant) + (1 | Question)
##    Data: long_data
##
##      AIC      BIC   logLik deviance df.resid
##   6160.2   6186.8  -3076.1   6152.2     5596
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4688 -0.6890  0.3396  0.5967  3.4687
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Question    (Intercept) 1.8014   1.3422
##  Participant (Intercept) 0.2565   0.5065
## Number of obs: 5600, groups:  Question, 224;
Participant, 25
##
## Fixed effects:
##                        Estimate Std. Error z value
Pr(>|z|)
## (Intercept)            0.814390   0.188987   4.309
1.64e-05 ***
## Years_of_Musicianship -0.001953   0.025281  -0.077
0.938
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## Yrs_f_Mscns -0.672
```

2. **Generalized Linear Mixed-Effects Model (glmer) with whether being a profoundly trained musician as a predictor**

```
## Generalized linear mixed model fit by maximum
likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Response ~ Musician + (1 | Participant) + (1
| Question)
```

```
##    Data: long_data
##
##      AIC      BIC   logLik deviance df.resid
##   6160.1   6186.7  -3076.1   6152.1     5596
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4618 -0.6885  0.3397  0.5970  3.4704
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Question    (Intercept) 1.8014   1.3422
##  Participant (Intercept) 0.2553   0.5053
## Number of obs: 5600, groups:  Question, 224;
Participant, 25
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.78961    0.14722   5.364 8.16e-08 ***
## Musician1    0.09339    0.28940   0.323    0.747
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## Musician1 -0.314
```

3. **One-way MANOVA test on the acoustic data as a function of emotion category, with three acoustic parameters as the dependent variables**

```
##  Response Meanf0Norm :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Emotion       6 32.864  5.4773  28.755 < 2.2e-16 ***
## Residuals   217 41.334  0.1905
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
##
##  Response Rangef0Norm :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Emotion       6 54.711  9.1185  23.849 < 2.2e-16 ***
## Residuals   217 82.970  0.3823
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
##
##  Response SpeechRate :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Emotion       6 140.23 23.3715  34.585 < 2.2e-16 ***
## Residuals   217 146.64  0.6758
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
##
## 因为不存在，6个观察量被删除了
```

4. Canonical and linear discriminant analysis

```
## Call:
## lda(Emotion ~ Meanf0Norm + Rangef0Norm + SpeechRate,
data = data)
##
## Prior probabilities of groups:
##         1         2         3         4         5
6         7
##  0.1428571  0.1428571  0.1428571  0.1428571  0.1428571
0.1428571 0.1428571
##
## Group means:
##   Meanf0Norm Rangef0Norm SpeechRate
## 1  1.3554528   1.9811007   6.909088
## 2  0.7493496   1.8604332   4.447403
## 3  0.9302905   1.0527448   6.241599
## 4  1.2394317   1.9792575   5.691425
## 5  0.4486006   1.0723815   6.158218
## 6  0.6871672   0.8763658   4.800152
## 7  1.6113958   2.1113079   5.920063
##
## Coefficients of linear discriminants:
##                    LD1        LD2       LD3
## Meanf0Norm  -1.1728542  0.8399380 -2.187641
## Rangef0Norm -0.2400402  0.9680464  1.542909
## SpeechRate  -0.8996502 -0.7451048  0.371496
##
## Proportion of trace:
##    LD1    LD2    LD3
## 0.5410 0.3420 0.1171
```

```
##
## Canonical Discriminant Analysis for Emotion:
##
##    CanRsq Eigenvalue Difference Percent Cumulative
## 1 0.52928    1.12442    0.41359  54.096     54.096
## 2 0.41549    0.71084    0.41359  34.199     88.295
## 3 0.19569    0.24330    0.41359  11.705    100.000
##
## Test of H0: The canonical correlations in the
## current row and all that follow are zero
##
##   LR test stat approx F numDF denDF   Pr(> F)
## 1      0.22130   23.818    18 608.6 < 2.2e-16 ***
## 2      0.47013   19.805    10 432.0 < 2.2e-16 ***
```

```
## 3      0.80431   13.199     4 217.0 1.218e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
```