Newcastle University

The effect of immediate and simple feedback on the perception of Mandarin lexical tones by English speakers

Supervisors: Dr. Biao Zeng & Prof. Martha Young-Scholten
September 2012

Dissertation submitted in partial fulfilment of the MA in Linguistics and Language Acquisition at Newcastle University. I declare that this work is entirely my own and that in all cases where I have drawn on the work of any other author, either directly or indirectly, this is fully and specifically acknowledged in the text of my dissertation and the work cited in the bibliographical references listed at the end of the Dissertation.

Abstract

Lexical tone is an important feature in Mandarin. However, it caused substantial difficulty for the non-Mandarin speakers. Since lexical tones are not completely illegible to non-Mandarin speakers, many methods are employed to improve the perception of Mandarin lexical tones by non-Mandarin speakers. One method is perceptual learning. Perceptual learning is a learning style through which people pick up "previously unused information" (Gibson & Gibson1955). Most perceptual learning studies on Mandarin lexical tone perception have used perceptual training as the condition for the participants to pick up information (e.g. Wang et al. 1999; Wang et al. 2003; Francis et al. 2008). Feedback, which is another important tool for perceptual learning, is nevertheless understudied in Mandarin lexical tone perceptual learning studies. Since feedback has many categories, in the current study, only the immediate and simplest form of feedback is examined. Whether the perception of the lexical tones by non-Mandarin speakers can be improved by receiving immediate and simple feedback is the focus of this study. The aims of this study are: 1) to investigate the effect of feedback and provide future studies with experimental grounds on the use of feedback; 2) to suggest the use of feedback in Computer Assisted Language Learning through investigating the effect of feedback; 3) to contribute to the existing theories, such as Autosegmental Theory (Goldsmith 1979), Categorical Perception (Best 1995), Noticing Hypothesis (Schmidt 1990), with empirical evidence.

The current study examined the perception of Mandarin lexical tones by 24 native

British English speakers, with 5 native Mandarin speakers as one of the control groups. The experiment made use of an AX discrimination task which required the participants to make judgments on whether the tones of 160 pairs of stimuli were the same or not, regardless of the consonants or vowels. The experiment group consisted of 12 native British English speakers. This group received simple feedback which only indicated the incorrectness on the incorrect judgments immediately after the judgment was made. As a control group, the other 12 participants did not receive any feedback. The results of the experiment showed that simple immediate feedback did not have a significant effect on the perception of Mandarin lexical tones by English speakers, both in terms of accuracy and reaction time. The reasons for the results are mainly discussed in terms of feedback types, individual differences on perceptual learning, and the influence from musical experience on lexical tone perception.

The first chapter of this dissertation is a general introduction. In the second chapter, some background information about the phonology of Mandarin syllables is introduced. The third chapter reviews the previous studies on lexical tone perception, perceptual learning and feedback, to establish the basis for this study. A pilot experiment and a full-scale experiment are reported in the fourth and the fifth chapter respectively. The last chapter is dedicated to the general discussion, conclusion and suggestions for future studies.

Key words: Mandarin lexical tone; perceptual learning; feedback

Table of Contents

List of Figures and Tables

1. Introduction

1.1 The issues in discussion

Lexical tone, an important feature in Mandarin Chinese, is one of the most difficult features for English speakers to acquire (e.g. Kiriloff 1969; Bluhme and Burr1971; Francis et al. 2008; Leather 1990; Wang et al. 1999). One reason is that the interaction of the two suprasegmental features of Mandarin, i.e. lexical tones and intonation, is complex and not yet thoroughly understood (Peabody and Seneff 2006). Despite of the influence of sentence-level prosodic interaction, even the tones of the single mono-syllabic Chinese characters are difficult for English speakers to distinguish.

The reasons for such difficulty are mainly due to the different methods that native Mandarin speakers and non-Mandarin speakers use in perceiving Mandarin lexical tones. According to Autosegmental Theory (Goldsmith 1979), Mandarin lexical tones belong to suprasegments while consonants and vowels belong to segments. Native Mandarin speakers are able to perceive tones separately from the consonants and vowels while non-Mandarin speakers find it difficult to do so. Moreover, native speakers could perceive Mandarin lexical tones categorically, while non-Mandarin speakers fail to perform categorical perception of the pitches in speech.

Since non-Mandarin speakers are not tone-deaf, efforts have been made to improve either the perception or production of Mandarin lexical tones by previous studies. One way of improving their perception is through perceptual learning. For example, non-tonal language speakers do not pay attention to lexical tones because tones do not make any difference in their native language; nevertheless, after having more experience, such as receiving training or feedback, their perception of tones will improve. Training, as the most common perceptual learning method, has been employed in many studies that investigated the effect of perceptual learning on music perception and speech perception, especially on segmental phonology perception (e.g. Bradlow et al. 1997; Iverson et al. 2003; Kraljic and Samuel 2005). Recently, a few studies have also been done on training non-Mandarin speakers to perceive or produce Mandarin lexical tones (e.g. Francis et al. 2008; Wang et al. 1999; Wang et al. 2003). However, feedback, as a more efficient perceptual learning method, has been understudied.

Feedback has many forms. In terms of its timing, there are immediate feedback and delayed feedback (e.g. Kulhavy 1977; Kulik and Kulik 1988); with regards to its complexity, feedback ranges from simply indicating the correctness of a response, to the most elaborated answer which includes the review of the detailed elaboration on the related knowledge (e.g. Mason and Bruning 2001). In order to avoid the influence of the recapitulation of former knowledge, the current study only examines the simplest form of feedback, which only informs the participants of their errors.

1.2 The current study

In light of these concerns, the present study examined the effect of immediate and simple feedback on the perception of Mandarin lexical tones by non-Mandarin speakers. It tested 24 native British English speakers, together with 5 native Mandarin speakers as controls, with an AX discrimination task (Cutler and Chen 1997; Zeng 2008) on their perception of Mandarin lexical tones. The following questions will be addressed:

1) Will immediate and simplest feedback have any effect on the perception of Mandarin lexical tones by non-Mandarin speakers?

2) Which aspect(s) will the effect be about, if any? Will it be about the accuracy, or reaction time? Or both?

To answer the above questions, two hypotheses have been formed:

Hypothesis 1: The error rate of English speakers' perception of Mandarin lexical tones will decrease as more feedback is received.

Hypothesis 2: The reaction time of English speakers' discrimination of Mandarin lexical tones will be shortened as more feedback is received, even after the accuracy rate stops improving.

1.3 Aims

The aims of the present study revolve around three points. The first is to examine whether the feedback will make any difference in improving the perception of non-Mandarin speakers. If feedback has any significant effect on the perception of Mandarin tones, more future studies can make use of short-term feedback instead of long-term training, which will greatly simplify the process of the experiment.

The second aim of this study is to give support to Computer Assisted Language Learning (CALL). CALL is becoming an increasingly popular teaching and learning method nowadays. Whether providing immediate and simple feedback could improve the learners' perception of Mandarin lexical tones is an important issue in Mandarin CALL. If the results support the hypotheses, then CALL could develop more learning schemes that take advantage of feedback; if the results of this study do not support the hypotheses, CALL may need to reduce the use of immediate and simple feedback or change the ways of giving feedback.

Besides the previous two aims, the results of the current study may perform as a piece of evidence for some theories, such as Autosegmental Theory (Goldsmith 1979), Categorical Perception (Best 1995), Noticing Hypothesis (Schmidt 1990), and so on.

1.4 Overview

The second chapter of this dissertation will introduce the background of this study in detail, namely the phonology of Mandarin. The third chapter will cover the review of previous studies, with regards to three topics: 1) lexical tone perception; 2) perceptual learning; 3) feedback. In the fourth and fifth chapters, a pilot as well as a full-scale experiment will be reported respectively. The last chapter will be on the discussion of the experiment results, the conclusion and the suggestions for future studies.

2. Background: Phonology of Mandarin syllables

2.1 Introduction

Traditionally, a Mandarin syllable is considered to consist of three parts: the initial, the final and the tone (Cheung 1973; Huang 1992). However, more and more scholars tend to categorize them as onset, rime and tone. (Duanmu 2007; Lin 2007) The difference between these two categorization methods lies on the glide. In the first categorization, the glide belongs to the final, while it belongs to the onset in the second categorization (Triskova 2011). Regardless of the names, the onset (or the initial) and the rime (or the final) are composed of consonants and vowels. They are therefore segmental features. The tone, which is indisputably separate in both methods, is a suprasegmental feature. The present study will adopt the names of "onset" and "rime" for the segments. In this chapter, the segmental features, i.e. the onsets and the rimes, of Mandarin syllables and suprasegmental features of Mandarin syllables will be introduced.

2.2 Segmental features of Mandarin syllables

Mandarin is a monosyllabic language. A Mandarin syllable usually has an onset and a rime. The onset is either a consonant (C) or it is omitted; the rime can be monophthongs, diphthongs, or vowel (V) + nasal ([n] or [ŋ]) or vowel+ liquid ([ɹ]). Therefore, the syllable structure of Mandarin could be V, VC (nasal/ liquid), CV, CVV, CVC (nasal/ liquid), or CVVC (nasal/ liquid).

2.2.1 Onsets

Most of the onsets in Mandarin exist in English, for example, [m], [n], [p^{h}], [k^{h}], etc. Some others have subtle differences with those in English, such as /f/ is pronounced as [f^{h}] in Chinese, which is different from [f] in English. These subtle differences are on the level of allophones, which will not influence the perception of the syllable. However, there are a few onsets that the English speakers find difficult to perceive and produce, for instance, the alveolar-palatals ([tɕ], [tɕ^{h}], [ɕ]), the fricatives ([ts], [ts^{h}]) and the affricates ([tʂ], [tʂ^{h}], [ʂ^{h}], [ʐ]). These sounds will be excluded in the present study since the focus is on the suprasegmental features rather than the segmental features. /w/ and /j/, considered by the scholars who support the Onset-Rime Model as glides, are also excluded.

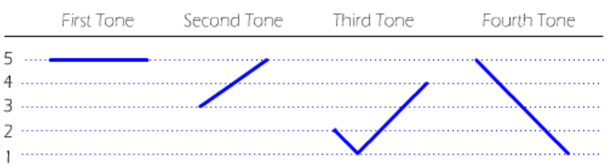
2.2.2 Rimes

Mandarin has a small repertoire of vowel phonemes, /a/, /o/, /i/, /u/ and /y/. Except for [y], the other four phonemes all exist in English, including their allophones. As for the V + nasal and V+ liquid, such combinations are legal in English phonology. The most difficult rimes for the English speakers are the apicals, [ɿ] and [ʅ].Therefore, in the design of present study, [y], [ɿ] and [ʅ] are excluded. /u/ and /i/ are also excluded, but only when they are in positions in which the scholars supporting the Initial-Final Model consider them as glides.

2.3 Suprasegmental features of Mandarin syllables

The lexical tone is the only suprasegmental feature on syllable level in Mandarin. There are five lexical tones, high level, rising, dipping, falling, and neutral (Li and Thompson 1977). The first four are the most important ones for comprehension; the neutral tone, however, only occurs when the syllable is unstressed. Chao (1930) designed a tone letter system that could represent the tones according to their relative pitch levels. In this system, the lowest level is "1" and "5" is the highest. The pitch levels of the four tones are 55, 35, 214 and 51 respectively. When the syllables are pronounced in isolated syllables under ideal circumstances, the first four tones will have the contours as shown in Figure 1. The four tones resemble music notes to some extent, but the pitches of music tones are absolute, while those of Mandarin lexical tones are comparatively relative. In this study, neutral tone, which is commonly referred to as "no tone", is excluded, due to its many allotones, which vary according to the tone of the preceding syllable.

Figure 1: Four Mandarin lexical tones (Chao 1930)



Tone plays an important role in the semantics of a syllable. In Mandarin, one character may have one or several pronunciations. But it is always monosyllabic no matter which pronunciation it adopts. Each syllable has one of the five lexical tones. Change of lexical tones of a syllable may result in either referencing another character, or varying the meaning of the same character altogether. Different characters may have the same syllable and the same tone. The corresponding relationship between the tones and characters may be many-to-one or one-to-many. For example, in Table 1, a syllable of a certain tone can have more than one corresponding characters; one character can also have different tones with the same syllable, such as “吗” has the second, third and neutral tones; one character can have different syllables with different tones as well, for instance, “抹” can also be pronounced as *mǒ* [mo]214 (‘erase’), or *mò* [mo]51 (‘plaster’).

Table 1: Example of the relationship between tone and character in Mandarin

(*Pinyin*: Romanized representation of Mandarin pronunciation)

Pronunciation		Chinese characters
Pinyin	IPA	
mā	[A̧] 55	妈(‘mother’), 抹(wipe)[2]
má	[A̧] 35	麻(‘hemp’), 吗(‘what’)[1]
mǎ	[A̧] 214	马(‘horse’), 吗(‘morphine’)[1]
mà	[A̧] 51	骂(‘scold’)
ma	[A̧]	吗(a particle used at the end of questions)[1]

Yang et al. (1988) calculated that there are approximately 1300 syllable sounds in Mandarin, while more than 50,000 characters are used in the writing system. If there are no tones, only 411 syllable sounds are left in Mandarin. Therefore, perceiving the lexical tones correctly is an important task in learning Mandarin.

2.4 Summary

A Mandarin syllable is composed of segmental features and suprasegmental features. Segmental features include onsets and rimes. In the current study, some segmental sounds in both onset and rime positions that may cause too much perceptual difficulty for English speakers are excluded. The suprasegmental feature on syllable level only consists of lexical tones. Lexical tones are closely related to the meaning of Mandarin characters; therefore, the perception of lexical tones is a very important issue. In the next chapter, a detailed review of previous study on lexical tone perception, perceptual learning, and feedback will be presented.

3. Literature review: mechanisms and effectiveness of different means of lexical tone perception, perceptual learning, and feedback

3.1 Introduction

The previous chapter briefly introduced Mandarin phonology, including its segmental features and suprasegmental features. In this chapter, a more detailed account of lexical tone perception, perceptual learning, and feedback will be reviewed from various aspects.

In Section 3.2, the perception of lexical tones will be reviewed in terms of the native speakers' perception and the influence of different linguistic backgrounds on lexical tone perception, as well as the interaction between lexical tones and the segmental and suprasegmental features of Mandarin syllables. To improve the perception of Mandarin lexical tones by non-Mandarin speakers, perceptual learning is a frequently used method. In Section 3.3, various aspects of perceptual learning will be examined, including the mechanisms of perceptual learning and previous studies on perceptual learning of lexical tones. Since the current study is investigating the effect of feedback on the perception of Mandarin lexical tones, Section 3.4 will review the different aspects of feedback, including its mechanisms, the types and the effectiveness of feedback in perceptual learning in the previous studies.

3.2 Lexical tone perception

3.2.1 Mandarin lexical tone perception by native Mandarin speakers

Fundamental frequency (f0) is the lowest frequency of a periodic waveform, which decides the pitches of speech and music. Tone perception mainly depends on the perception of different f0 values, among which the f0 contour and f0 height are the most crucial factors (Wang et al. 1999; Peabody and Seneff 2006).

F0 contour has always been considered as the primary cue for tone perception since early studies in 1970s, such as Abramson (1975) and Gandour (1983). Figure 2a is the canonical f0 contour of the first four lexical tones in Mandarin when pronounced in isolation, which serves as an acoustic support for the tone letter system developed by Chao in 1930 according to the pitch heights of the tones (as shown in Figure 2b). Studies from various perspectives supported the importance of f0 contour to Mandarin lexical tone perception (e.g. Gårding et al. 1986; Shen and Lin 1991; Whalen and Xu 1992; Moore and Jongman 1997; Xu 1997; Fu et al. 1998; Wang 1999; Liu and Samuel 2004; Francis et al. 2008).

Shen and Lin (1991) found that the turning point of f0 contour acted as an important cue in differentiating the second tone and third tone in Mandarin. Two experiments were conducted in this study. The first experiment presented two f0 continua of stimuli, each of whose turning point occurred at different percentage of the whole stimulus, to native Mandarin speakers. The participants perceived the tones according to the degree of the initial fall, which was not as steep in the second tone as in the third tone. The timing of the turning point therefore is considered as a cue for the differentiation between the second tone and the third tone. The second experiment made use of a tone identification task, which concluded that the mistakes made by the native-speaker participants occurred when the timing of the turning point and the degree of the initial fall were not correlated.

Figure 2a: canonical f0 contour of the first four lexical tones

in Mandarin when pronounced in isolation (Francis et al. 2008)

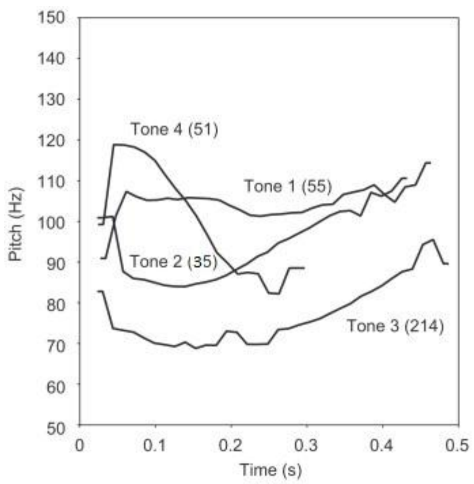
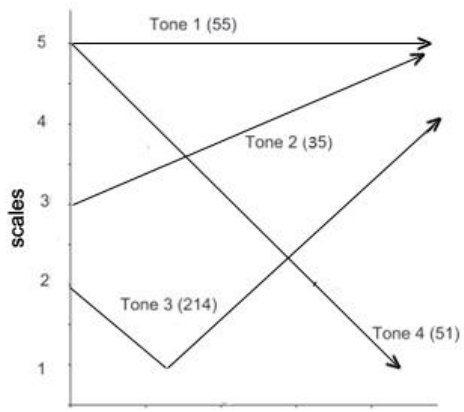


Figure 2b: tone letter system by Chao (1930)



F0 height is also a crucial cue in Mandarin lexical tone perception (Gandour 1983; Massaro and Cohen 1985; Lee 2009). Massaro and Cohen (1985) assessed the importance of both f0 height and f0 contour to the vowel in each stimulus. The result showed that f0 height was as important as f0 contour in the perceptual recognition of Mandarin lexical tones. The tones were normally recognized according to both cues; however, when one cue was not as clear as the other one, the other cue would be the most influential.

Despite the importance of f0 information, when this primary cue is missing, the perception of Mandarin lexical tones is also possible by judging from secondary cues such as amplitude, duration, or glottalization (Sagart 1986; Whalen and Xu 1992; Fu et al. 1998; Fu and Zeng 2000; Kong and Zeng 2006; Francis et al. 2008). Liu and Samuel (2004) found native Mandarin speakers could judge the tones at a high accuracy rate even when the f0 information was neutralized. When critical portion of f0 information was removed, the native speakers could still have an accuracy rate between 75% and 80%; when the f0 information was completely removed, the participants' performance was still surprisingly good. These results showed that f0 information was not the only cue.

Whalen and Xu(1992) reported that amplitude was a secondary cue in Mandarin tone recognition. In this study, the researchers used signal-correlated noise stimuli, which contained no f0 information but only amplitude contour and duration. The results showed that the native speakers had a recognition rate of 80% after the removal of f0 information and the second, the third and the fourth tone were able to be perceived only from amplitude contour.

Fu et al. (1998) found that duration, which was represented by temporal envelope cue, had an influence in tone recognition. This study divided the stimuli into different frequency analysis bands. The temporal and amplitude cues were preserved in each frequency band while the spectral detail within each band was removed. When the duration cue was added, the recognition rate was higher by 10% to 20% than when there was only the amplitude cue. The result indicated that the duration cue was a factor that influenced tone perception. However, some other studies, such as Whalen and Xu (1992) and Fu and Zeng (2000) reported that the duration had little effect.

3.2.2 Lexical tone perception and linguistic experience

Linguistic experience is an important factor in speech perception. Many studies have showed that linguistic experience has influence on the perception of segmental phonology (e.g. Miyawaki et al. 1975; Logan et al. 1991; Kuhl et al. 1992; Polka and Werker1994; Guion et al. 2000; Werker and Tees 2002).Similarly, linguistic experience also has influence on the perception of suprasegmental phonology. In this section, how speakers with different linguistic experience perceive lexical tones will be reviewed.

3.2.2.1 Lexical tone perception by speakers of tonal languages

F0 contour is not a unique focus of native Mandarin speakers; it is also the focus for the speakers of other tonal languages. Previous studies on the perception of tones covered different tonal languages such as Cantonese, Mandarin, Taiwanese Southern Min, Thai and Yoruba (e.g. Gandour and Harshman1978; Gandour 1983; Lee et al. 1996; Cutler and Chen 1997; Gandour et al. 2000; Hallé et al. 2004; Kaan 2007; Francis et al. 2008; Kaan 2008; Sun 2012; Zheng 2012). Many of them showed that speakers of other tonal languages tend to use f0 contours to distinguish Mandarin lexical tones as native Mandarin speakers do while speakers of non-tonal languages use other cues such as the pitch height of onsets and offsets.

Gandour and Harshman (1978) examined speakers of two tonal languages, the contour-tonal language Thai and the register-tonal language Yoruba, as well as speakers of a non-tonal language, i.e. American English. The results showed that the speakers of both tonal languages perceived tones according to the pitch contour direction while the English speakers paid more attention to the pitch height of the onsets or offsets. Gandour (1983) extended his previous study to the perceptions of tones by native speakers from four tonal languages, i.e. Cantonese, Mandarin, Taiwanese Southern Min, and Thai as well as English. Again the results converged with the previous conclusion.

Sun and Huang (2012) also discovered that tonal language speakers' perception of Taiwanese Southern Min were influenced by their phonological system while American English speakers depended more on the psychoacoustic features such as pitch height which is also important in intonation languages.

In Lee et al. (1996), English speakers, Cantonese speakers, and Mandarin speakers performed a same-different judgment task on Mandarin lexical tones and Cantonese lexical tones respectively. The results demonstrated that Cantonese speakers distinguished Mandarin lexical tones better than English speakers did, whereas Mandarin speakers' perception of Cantonese lexical tones was on a par with English speakers. Since Cantonese has a larger lexical tone inventory than Mandarin, the result clearly revealed that linguistic experience indeed affected tone perception.

3.2.2.2 Lexical tone perception by speakers of non-tonal languages

Other cross-linguistic studies on the perception of Mandarin lexical tones by non-tonal speakers also found that non-tonal speakers' perception is different form tonal language speakers. (e.g. Leather 1990; Lee et al.1996; Klein et al. 2001; Wong 2002; Gandour et al., 2003; Wang et al. 2004; Krishnan et al. 2005; Xu et al. 2006; Huang 2007; Chandrasekaran et al. 2010; Ding et al. 2011; So and Best 2011; Hao 2012).

Xu et al. (2006) carried out an identification task and a discrimination task on native English speakers perceiving Mandarin lexical tones. The results illustrated strong categorical perception for Mandarin speakers, who perceived Mandarin lexical tones mainly depending on the directions of the pitches, while English speakers, who perceived Mandarin lexical tones according to the pitch height, did not perceive them categorically.

Huang (2007) also examined English speakers and gained similar results. An AX discrimination task and an AX difference rating task were used in this study. English speakers again discriminated the Mandarin lexical tones according to the onset and offset pitch heights.

Studies on other non-tonal languages have converging results. Ding et al. (2011) examined German speakers' perception of Mandarin lexical tones with an identification task. Although English and German are both intonation languages, German has a simpler and less steep intonation inventory than English (Jilka 2005, cited in Ding et al 2011). Despite the differences between German and English, the study found that German speakers, as English speakers, identify Mandarin lexical tones with the knowledge of their intonation categories.

Different from English and German, which are "stress-timed or stress accented languages" (Beckman 1986, cited in So and Best 2011), French is a "syllable-timed language without an accent system" (Fox 2000, cited in So and Best 2011). Because of their difference, there might be a different perceptual pattern by the speakers of these two languages. Hallé et al. (2004) used an ABX identification task and an AX discrimination task to test the French speakers. The results indicated that the French speakers were able to perceive Mandarin lexical tones based on the psychoacoustic cues, such as onset and offset pitch heights, from the knowledge of their native intonation. In the study of So and Best (2011), English and French speakers were asked to categorize Mandarin lexical tones into their native intonation categories. This study echoed Hallé et al. (2004) by reaching the conclusion that non-native phonology would be assimilated into native prosodic system such as intonation.

3.2.3 Lexical tone perception and syllable segments and suprasegments

In terms of phonology, Mandarin lexical tones, which are suprasegments, are separate from the consonants and vowels, which are segments. However, the consonants and vowels as well as the tones themselves have an influence on the perception of Mandarin lexical tones. In this section, how the perception of Mandarin lexical tones is influenced by different onset and rime contexts and in different tone pairs will be examined.

3.2.3.1 Lexical tone perception and syllable segments

The interaction between Mandarin lexical tones and the onsets is understudied; the existing studies have shown some influence of the onsets on the lexical tones. Xu and Xu (2003) examined the effect of consonant aspiration on the f0 of Mandarin lexical tones. The results showed that the aspiration changed the starting point of the following f0; nevertheless, the ending points of the contours in both conditions were identical, no matter how big the difference was at the starting point.

Studies on the interaction between Mandarin lexical tones and the rimes are slightly more. Since vowels of different heights have different f0 in many non-tonal languages such as English (Zeng 2008), the interaction of lexical tones and vowel quality have been studied by various scholars.

Zeng (2008) reviewed some studies on the interaction between vowel height and Mandarin lexical tones. Xu (1957, cited in Zeng 2008) discovered that vowel height could be raised by a high tone and lowered by a low tone. However, Yip (2002) argued that it was the tones that influenced the vowel height. Regardless of the controversy, there was certain interaction between vowels and lexical tones.

Zeng (2008) further examined the interaction of tones and segments by applying the same-different paradigm. He divided the stimuli into four categories, i.e. same-onset/ same-rime (SOSR), same-onset/ different-rime (SODR), different-onset/ same-rime (DOSR), different-onset/ different-rime (DODR). The result of the experiment showed that when the rimes were the same, English speakers performed much better than when the rimes were different, as shown in Table 2. These results supported that the rimes, or vowels, had an influence on the perception of tones.

Table 2

Mean reaction times and accuracy rates

for Chinese and English participants in Zeng (2008)

Condition	Chinese	English
SOSR	375 (98%)	500 (99%)
SODR	737 (80%)	965 (50%)
DOSR	645 (88%)	738 (84%)
DODR	799 (77%)	988 2%)

3.2.3.2 Lexical tone perception and syllable suprasegments

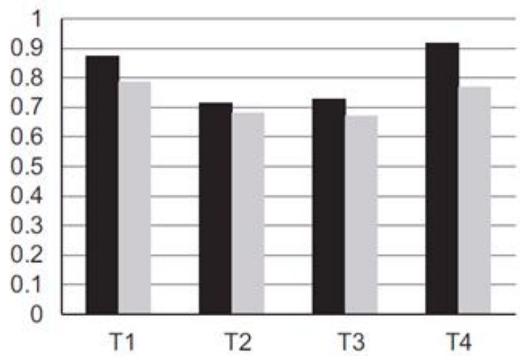
Although tone is the only suprasegment in Mandarin syllables, the four tones have different degrees of difficulty to non-Mandarin speakers. The third tone (214) is considered as the most difficult one by some studies. Wong et al. (2005) found that native Mandarin children acquired the third tone much slower than other tones. Hao (2012) reported the third tone was the most difficult tone for English speakers. Native English speakers and native Cantonese speakers performed an identification task in this study. The result, as shown in Figure 3, revealed that the third tone was more difficult than the first and fourth tone but at a similar difficulty level for both groups.

Figure 3: The mean accuracy rates for each Mandarin tone

in the identification task (Hao 2012)

(Black bars: Native English participants;

gray bars: Native Cantonese participants.)



Moreover, different tone pairs pose different difficulties to non-native speakers during tone discrimination tasks. The tone pair of the second tone (35) and third tone (214) is the most confusing pair in terms of both perception and production to non-Mandarin speakers (e.g. Kiriloff 1969; Shen and Lin 1991; Best 1995; Moore and Jongman 1997; Wang et al. 1999; Wang et al. 2003; Hume and Johnson 2003; Best and Tyler 2007; So and Best 2010; Hao 2012).

Hao (2012) had three tasks, i.e. identification task, mimicry task, and reading task, in all of which the native English and native Cantonese speakers performed the worst in the tone pairs between the second tone (35) and the third tone (214). One of the reasons for the confusion is the acoustic features of the two tones. As shown in Figure 2a above, the f0 contours of the second tone (35) and the third tone (214) are similar, both containing a falling-rising trend (Moore and Jongman 1997; Hao 2012). Since f0 contour is the primary cue for the perception of lexical tones for Mandarin speakers, it is not surprising that even native Mandarin speakers had confusions over these two tones (e.g. Chuang et al. 1972; Shen and Lin 1991). The other similarity is the f0 onset (Moore and Jongman 1997; Hao 2012). Since the non-tonal speakers mainly depend on the pitch height of onsets and offsets to distinguish the tones, the similar f0 onsets confuses them to a large extent.

3.2.4 Reasons for the difficulties in tone perception by non-native speakers

The review in the previous two sections showed that speakers of tonal languages and non-tonal languages have different ways of perceiving tones. However, the causes of these differences were not investigated. In this section, two possible reasons for the differences proposed by previous studies will be reviewed.

3.2.4.1 Categorical perception

One reason that leads to different perception abilities by tonal speakers and non-tonal speakers may be due to the categorical perception. Categorical perception is a concept in accordance with Perceptual Assimilation Model (Best 1995). It means that speakers map non-native phonology into their native categories. For example, Mandarin speakers will map different pitches into lexical tone systems while English speakers map them according to their intonation system. Many studies on lexical tone perception support this view (e.g. Lee et al. 1996; Hallé et al. 2004; Kaan et al. 2008; Francis et al. 2008; Zheng et al. 2011)

Stagray and Downs (1993) examined the differential sensitivity for frequency of English and Mandarin speakers with a same-different discrimination task. They discovered that English speakers could have more differential sensitivity towards small f0 contour variations than Mandarin speakers. This was because there was not any lexical tone category in English. English speakers therefore did not have to perceive frequencies categorically while the Mandarin speakers had to categorize all frequencies into their native tonal categories.

Similarly, the ERP study of Zheng et al. (2011) found that Cantonese speakers could make finer distinctions among different tones. Since Cantonese have more tonal categories than Mandarin, it was not surprising that Cantonese speakers would have more categories for different pitches. This study thus was also in line with Lee et al. (1996), both acting as evidence for categorical perception.

Evidence from neuroscience also support categorical perception of Mandarin lexical tones. Gandour et al. (2000) and Klein et al. (2001) both conducted positron emission tomography (PET) studies on lexical tone perception. Their results indicated that speakers from tonal languages only demonstrated activations in the left brain hemisphere, the area that processes linguistic related stimuli, when they dealt with their native tones while non-native speakers did not show any activation in the left hemisphere. These results illustrated that native speakers categorized lexical tones into language matters while non-native speakers could not.

In the ERP study of Kaan et al. (2008), English listeners were more sensitive to early f0 contour differences while Mandarin and Thai speakers were more sensitive to late f0 contour differences. However, having received tone category training, the early f0 contour sensitivity was suppressed. This indicated that English speakers gained the ability of categorical perception after training, which in turn provided evidence for categorical perception.

3.2.4.2 Autosegmental theory

The other possible reason for the different abilities of the perception of lexical tones by tonal and non-tonal speakers may be ascribed to the Autosegmental Theory. Goldsmith (1979) advanced the Autosegmental Theory, in which he contended that tonal components and segmental components belong to separate tiers.

In support of the Autosegmental theory, Zeng and Mattys (2011) discovered that the native Mandarin speakers perceived Mandarin lexical tones separately from the segmental features of syllables while the English speakers perceived them as a whole. They conducted a migration task, in which the participants heard a pair of dichotic syllables one pair at a time. This task was based on a psychological concept, illusory-conjunction phenomenon, which is represented by accidental combinations of two objects into one object (Treisman and Schmidt1982). Native English speakers and native Mandarin speakers were examined in this task. The results demonstrated that native Mandarin speakers had high migration rates for tones and rime + tone but no migration for rimes alone while native English speakers showed no tone migration but high migration in the other two conditions. These results indicated that Mandarin speakers automatically perceived lexical tones separately from the tone-bearing rhyme, but English speakers could not.

Since the non-tonal language speakers cannot perceive the Mandarin lexical tones separately with the segments, the conditions of consonants and vowels will very likely affect their perception. Therefore, the non-tonal language speakers perceive lexical tones differently from the tonal language speakers.

3.3 Perceptual learning

3.3.1 Introduction

Perceptual learning is "extracting previously unused information" (Gibson & Gibson1955), which is available to be picked up by people (Goldstone 1998). It "…

encompasses parts of the learning process that are independent of conscious forms of learning and involves structural and/or functional changes in the primary sensory cortices" (Fahle and Poggio 2002).

Studies on perceptual learning have provided rich behavioural evidence to the research on brain plasticity. Traditionally, it was believed that human brain would lose its plasticity in adulthood (Seitz and Watanabe 2005). However, studies from neuroscience have had evidence that are against the traditional idea (Buonomano and Merzenich 1998; Das et al. 2001). Perceptual learning supported the new belief by providing evidence of adults learning without being consciously trained.

Perceptual training has a common paradigm. Fiser (2009) described that, in a typical perceptual learning paradigm, the experimenters explain the well-defined task to the participants verbally; then, after a repetitive training (usually with feedback), the performance of the participants will improve. Through this paradigm, the participants unconsciously undergo a learning process.

3.3.2 Mechanisms of perceptual learning

In Goldstone's (1998) review of perceptual learning, he summarized that perceptual learning has four mechanisms: "differentiation", "unitization", "attentional weighting" and "stimulus imprinting". "Differentiation" is to separate the previous perceptually inseparable stimuli while "unitization" is to integrate or categorize the once separated stimuli. "Attentional weighting" is to shift the focus of the attention to certain features or dimensions of the stimuli. “Stimulus imprinting” means that the perceptual system will develop “detectors” (or “receptors”) that are specialized for the stimuli. All mechanisms are related to categorical perception.

“Differentiation”, including five sub-categories (differentiation of whole stimuli, psychophysical differentiation, differentiation of complex stimuli, differentiation of categories and differentiation of dimensions), is the process through which the stimuli that are difficult to distinguish become separate as experience grows. For instance, in a meta-analysis of face discrimination study, Shapiro and Penrod (1986) concluded that people were better at discriminating faces from their own races. This result supported “differentiation” in that the perceptual separability increased along with the familiarity.

On the contrary to “differentiation”, “unitization” is the mechanism that allows people to perceive thing in an integral manner. For example, LaBerge’s (1973) concluded that repeated exposure could result in unitization or chunking process in perception. This study used real letters and novel letters to form different patterns. The participants reacted slower at first when the pattern was not attended. However, after a period of repeated exposure, the participants were able to perceive the pattern in a unitized manner so that the non-attending condition gradually disappeared (similar studies: Salasoo et al. 1985; Czerwinski et al. 1992).

“Attentional weighting” means that the attention will be directed away from the undesired features or dimensions of the stimuli towards the desired features or dimensions of the stimuli as experience increases. This mechanism is closely related to categorical perception. For instance, Pisoni et al. (1982) examined the perception of voiced, voiceless unaspirated and voiceless aspirated stops by monolingual English speakers. The participants could notice the differences of non-native sound contrasts and perceive them in a categorical manner after a few minutes' training. Moreover, in the subsequent perceptual tests, the participants revealed consistent labelling and categorical discrimination functions for all 3 voicing categories.

The last mechanism, "stimulus imprinting", includes three sub-categories, i.e. whole stimulus storage, feature imprinting, and topological imprinting. This mechanism is to develop new detectors for the stimuli. For instance, Palmeri (1997) reported that the participants could tell the number of random patterns that were made up with 6 to 11dots after training. The participants' reaction time for the 6-dot pattern was at the same level with the 11-dot pattern if the patterns were presented in the same way as they were in the training sessions. The results showed that participants developed special detectors to perceive the patterns by storing the specific arrangements of the dots.

3.3.3 Previous studies on perceptual learning

Perceptual learning is a widely studied concept. Researches about perceptual learning have been carried out in many fields, such as neurophysiology (e.g. Crist et al. 2001; Carcagno and Plack 2011), psychophysics (e.g. Dosher and Lu 1998; Chung et al. 2005), neuroimaging (e.g. Schiltz et al. 1999; Schwartz et al. 2002), speech studies (e.g. Wang et al. 1999; Wang et al. 2003; Kraljic and Samuel 2005; Francis et al. 2008) and so on. However, for the past 50 years, most studies were about visual processing, making use of tasks such as face discrimination (e.g. Shapiro and Penrod 1986; de Heering and Maurer 2011), texture discrimination (e.g. Karni and Sagi 1991; Hussain et al. 2009), orientation discrimination (e.g. Schiltz et al. 1999; Zhang et al. 2010), and motion direction discrimination (e.g. Ball and Sekuler 1987; McDevitt et al. 2012).

In recent years, studies on speech perception and production have been increasing in number. Samuel and Kraljic (2009) reviewed the studies on perceptual learning for speech by dividing all the studies into two themes, including five sub-categories, i.e. perceptual learning of non-native phonetic contrasts, perceptual learning of accents and idiolects, perceptual learning of degraded speech input, as well as lexically induced perceptual learning and audio-visually induced perceptual learning.

As regards to perceptual learning of non-native phonetic contrasts, a commonly studied subject is Japanese speakers' perception or production of English /r/ and /l/ (Logan et al. 1991; Lively et al. 1993; Lively and Pisoni. 1994; Bradlow et al. 1997; Bradlow et al. 1999; Aoyama et al. 2004; Cutler et al. 2006). Other studies include Mandarin speakers' perception of English /t/ and /d/ in the word final position (Flege 1995), Dutch speakers' identification of Japanese fricative /s/ (Sadakata and McQueen 2011), American speakers' discrimination of German vowels (Kingston 2003) and so on. The results of these studies all supported that perceptual learning could improve the perception or production of non-native phonetic contrasts.

The process of getting used to different accents and idiolects is also perceptual learning. Samuel and Kraljic (2009) mainly reviewed three studies, i.e. native English listeners' perception of heavy Chinese-accent speech (Bradlow and Bent 2008), native English listeners' perception of Spanish- and Chinese- accent speech with shorter training (Clarke and Garett 2004) and the perception of the speech by deaf talkers (McGarr 1983). All three studies revealed that the listeners developed better perception after a period of exposure.

For degraded speech input, perceptual learning is also effective. Degraded speech is intentionally modified speech, such as stimuli compressed by computer software (e.g. Dupoux and Green 1997; Pallier et al. 1998), or vocoding speech with noise (e.g. Davis et al. 2005).

Lexically induced perceptual learning and audiovisually induced perceptual learning use lexical context and visual cues to induce learning. For example, Norris et al. (2003) used an ambiguous sound between [f] and [s]. In the training, the participants that were given context of real words could categorize the ambiguous sound as the sound that was needed in the word whereas the participants who heard non-words could not. Samuel and Kraljic (2009) concluded that lexically induced perceptual learning was more stable than the visually induced perceptual learning.

As reviewed above, most studies on speech perception and production through perceptual learning focus on the segmental features. Only a few studies have been done on suprasegmental features such as pitches and lexical tones (Wang et al. 2003; Kraljic and Samuel 2005; Francis et al. 2008; Carcagno and Plack 2011).

Carcagno and Plack (2011) examined the discrimination of three pitches that had the same f0 contours with three Mandarin lexical tones, and monitored the changes of the frequency-following response (FFR). After 10 hours of training over the course of 34 days in average, the trained participants showed significant improvement in the odd-one-out task, compared with the untrained participants in the control group. The FFR strength was also partly enhanced. The results indicated that the ability of discriminating pitches could be improved by getting more experience.

As pitch discrimination could be improved by perceptual training, the perception of Mandarin lexical tones could also be trained. Wang et al. (1999) successfully trained American English speakers to identify Mandarin lexical tones through perceptual learning. The researchers trained eight American English speakers for eight sessions over the course of two weeks. In every training session, the American English speakers were asked to perform a forced-choice identification task, which required the participants to identify the tones in tone pairs. After the participants made a judgment, there would be a neutral voice telling the participants the correct answer in English, with a repetition of the trial following. Immediately after each training session, the post-tests were carried out. The accuracy of the participants' identification was improved by 21% from the pre-test to the post-test. In the generalization tests, the participants made an 18% improvement to the new stimuli produced by the same speaker, and a 25% improvement to the new stimuli produced by a new speaker. Six months after training, a retention test was performed, in which the participants still had a 21% improvement than the pre-test.

Wang et al. (2003) further examined the perceptual learning of Mandarin lexical tones by American speakers in terms of the production as an extension of the Wang et al. (1999). The major procedures were the same with Wang et al. (1999), with production tests after the perception pre-test and the perception post-tests. The identification rate of the American English speakers' tone productions judged by native Mandarin speakers improved by 18% from the pre-test to the post-test. The acoustic analysis of both pre-test productions and post-test productions also proved the improvement.

Studies of other lexical tones also support the effectiveness of perceptual training. Francis et al. (2008) also found perceptual learning effective by examining the recognition of Cantonese lexical tones by Mandarin and English speakers. The researchers trained the Mandarin speaking participants for approximately ten hours over the course of ten days and the English speaking participants for 16 hours over the course of 30 days. The procedures of the training session were the same with Wang et al. (1999). The results, converging with previous studies, revealed improvements for both the Mandarin speaking participants and the English speaking participant. The identification accuracy rate of the Mandarin speaking participants improved from 63.7% in the pre-test to 73.0% in the post-test and that of the English speaking participants increased from 66.0% to 82.7%.

These studies all involved training sessions for perceptual learning. However, some studies showed that perceptual learning could happen under much simpler circumstances. Norris et al. (2003) discovered that the participants had significant improvement on categorizing the non-native phoneme after only listening to 20 ambiguous sounds in a 200-word stimuli inventory. This result indicated that it is not necessary to have too much training for perceptual learning to happen. The participants in the study of Pisoni et al. (1982) also revealed categorical perception after a few minutes' training, which converged with the results in Norris et al. (2003).

Longer training is not only unnecessary but also has a reverse effect. In the study of Molley et al. (2012), shorter training sessions showed better learning results than longer training sessions. The researchers divided the participants into four groups, T100, T200, T400 and T800, with T100 receiving a daily 100-trial training for over 8 days, T200 receiving a daily 200-trial training for over 8 days, T400 receiving a daily 400-trial training for over 4 days, and T800 receiving a daily 800-trial training for over 2 days. The results showed that the participants in T100 learnt at the highest rate whereas the participants in the longest training sessions showed over-training effect in the end.

Some studies even found perceptual learning effective without any training or feedback. Petrov et al. (2006) reported that in an orientation discrimination task, the participants received no prior training or any feedback during the test. Surprisingly, they showed robust learning in terms of both accuracy and learning rate.

These studies demonstrated that training was not necessary. Therefore, the present study will examine the effect of the feedback only, without any prior training.

3.4 Feedback

3.4.1 Mechanisms of feedback

In the literature of feedback study, there are three explanations for feedback. The first one, reinforcement, is under the behaviourism framework in the early 20th century; the second one, consciousness-raising, is a mechanism viewed from a cognitive point of view; the third one, which is more recent, makes use of neuro-technology to lend support to the old theory, categorical perception.

3.4.1.1 Reinforcement

Under the behaviourism framework, early studies examined feedback and regarded feedback as a reinforcement of the right answers or a correction for the wrong answers (e.g. Pressey 1950; Skinner 1954; Holland 1960). The leading scholar in behaviourism, Skinner, also stated in his 1954 article that teachers should reinforce students' right answers in classroom instead of delaying for too long and losing the effectiveness of the feedback (Skinner 1954). Holland (1960) also considered feedback as an "immediate reinforcement for correct answers". These scholars who believed reinforcement was the mechanism that underlay feedback, also supported that immediate feedback worked more efficiently than the delayed feedback because delayed feedback failed to reinforce the correct response. They also mainly worked in the domain of classroom teaching, due to the limitation of technology development in the early 20th century.

3.4.1.2 Consciousness-raising

Since 1970s, however, the belief has changed towards the cognitive perspective. According to the information-processing theory, mistakes are representations of the learners' cognitive processes (Bruning et al. 1999). This process, with the name of "consciousness raising", "noticing", "attention" or other resembling terms, refers to a similar concept that the salience of underlying structures of the test items is addressed (Rutherford and Sharwood Smith 1985; Rutherford 1987; Nagata and Swisher 1995). In this process, feedback can help the learners' raise their awareness of the misconceptions, and make sure of their understanding and the expectations of their performance (Mason and Bruning 2001).

Kulhavy and Stock (1989) reviewed written feedback from this perspective. They concluded that effective feedback could provide the learners with both "verification" and "elaboration". The "verification" part of the feedback decides the correctness of a response while the "elaboration" part provides information for the learners to modify their next response towards a right direction. Findings from many studies support that only the combination of both parts could produce effective feedback (e.g. Roberts and Park 1984; Rosa and Leow 2004)

3.4.1.3 Categorical perception

The 21st century sees new technological development. More and more evidence from neuroscience suggest that feedback facilitates categorical perception. The fMRI study of Poldrack et al. (2001) found that striatum, the largest component of the basal ganglia, was activated by feedback-based learning while not activated in the observational learning without feedback. Shohamy et al. (2004) also found converging data from both neuroimaging and neuropsychology to prove this point. Seger (2008) further pointed out that feedback benefited learning in terms of two ways of involving with basal ganglia: "First, feedback results in dopamine signals that project to the striatum and affect synaptic plasticity at the corticostriatal synapse (Reynolds and Wickens, 2002). Second, feedback is a signal used by the executive and motivational corticostriatal loops to modulate activity in their associated cortical regions (Kimura and Graybiel, 1995)." Some other studies discovered that basal ganglia were responsible for categorization tasks (e.g. Vogels et al. 2002; Seger et al. 2000; Nomura et al. 2007; Cincotta and Seger 2007). Therefore, feedback happens in the brain area that deals with categorization. Evidence from neuroscience could lend support to the categorical perception mechanism of feedback.

3.4.2 Types of feedback

Viewed from different perspectives, feedback can have many types. In terms of the feedback timing, there are immediate feedback and delayed feedback(e.g. Kulhavy 1977; Kulik and Kulik 1988; Opitz et al. 2011); in terms of the feedback modes, there are written feedback, spoken feedback, or audio-visual feedback(e.g. De Bot1983; Jones1998; Sachs and Polio 2007); in terms of the feedback quality, there are error-message-only feedback and informative feedback(e.g. PujolÀ 2001; Miller et al. 2005); in terms of feedback clarity, there are explicit feedback and implicit feedback (e.g. Carroll and Swain1993; Ellis et al. 2006); in terms of the psychological processing of feedback, there are positive feedback and negative feedback (e.g. Aljaafreh and Lantolf 1994; Brandl1995).

The same feedback sometimes has different names in different disciplines depending on their focuses. For example, Lyster (1997) listed the different names for “negative feedback” considered by psychologists (e.g. Annett1969; Long et al. 1998) in other disciplines: it is called “negative evidence” by linguists (e.g. White 1989; Perfors 2010), “corrective feedback” by second language teachers (e.g. Fanselow1977; Saito and Lyster2011), “repair” by discourse analysts (e.g. Kasper 1985; Liebscherand Dailey-O'Cain 2003),and as “focus-on-form” in classroom second language acquisition (e.g. Lightbown and Spada 1990; Loewen 2011). In this section, two ways of categorizing the feedback that are most relevant to the current study will be reviewed, i.e. the feedback types by the timing and the complexity.

3.4.2.1 Feedback types by the timing

Feedback can be divided into two types according to the timing of the feedback: immediate feedback, and delayed feedback. According to Dempsey et al. (1993), both immediate feedback and delayed feedback can be further divided into six types respectively as listed in Table 3.

Table 3: Types of feedback (Dempsey et al. 1993)

Immediate feedback	Delayed feedback
Item-by-item	Item-by-item
Learner-controlled	Logical content break
Logical content break	Less than 1 hour (end of session)
End-of-module (end of session)	1-24 hours (end of session)
Break by learner	1-7 days (end of session)
Time-controlled (end of session)	Extended delay (end of session)

Scholars have been arguing about which is more effective since the 1920s. The scholars who supported the immediate feedback argued that errors should be corrected before the students remember them (Pressey1932; Mason and Bruning 2001) and the correct response should be reinforced immediately (Skinner 1954; Renner 1964); the scholars who supported delayed feedback, however, believed that delayed feedback could reduce proactive interference so that the incorrect information could be forgotten before inputting the correct information (Kulhavy and Anderson 1972).

Some studies, mostly early studies, were in favour of delayed feedback. For example, Kulhavy and Anderson (1972) claimed that delayed feedback was superior to immediate feedback with evidence from a multiple-choice test. Kulhavy (1977) again supported his previous claim with evidence from a writing task. Bardwell (1981) also found delayed feedback was more effective than immediate feedback in terms of a school related learning.

Kulik and Kulik (1988) challenged Kulhavy's assertion that delayed feedback only works well in some special experimental situations. They made a meta-analysis on 53 studies and found that, in applied studies, nine out of eleven studies were in favour of immediate feedback with four having statistical significance; in experiments on acquisition of test content, 13 out of 14 studies supported that delayed feedback was more effective, with seven being significant. In the studies that included follow-up retention, 16 out of 27 studies were in favour of immediate feedback (with 10 significant ones) while the other 11 supported delayed feedback (4 significant ones). Based on these data, they further concluded that immediate feedback was more effective in situations that made use of actual classroom quizzes and real learning material while the delayed feedback worked better in some experimental studies. The mixing results were due to the specific features of the studies. More recent studies mainly support the effectiveness of immediate feedback. For example, Opitz et al. (2011) conducted an event-related potentials (ERP) experiment to examine the effect of immediate and delayed feedback in an artificial grammar learning (AGL) task. The results showed that immediate feedback was more effective for AGL task.

Nevertheless, some studies contended that the effect of immediate feedback and that of delayed feedback did not have significant difference. For example, Lopez (2009) used a computer-based test to examine which feedback, immediate feedback or delayed feedback, worked better in a science final exam. The results showed no significant difference between different types of feedback after five exams under each of the conditions.

Mason and Bruning (2001) concluded from several previous studies that the effectiveness of immediate feedback could be higher in simpler tasks such as decision-making and novel information tasks (e.g. Jonassen and Hannum 1987) and lower-level knowledge-based tasks (e.g. Gaynor 1981). On the contrary, delayed feedback is more suitable for higher-level tasks such as abstract concepts and application or comprehension skills (e.g. Gaynor 1981; Jonassen and Hannum 1987). The current study examines the perception of Mandarin lexical tones through an AX discrimination task, which belongs to a simple task. Since immediate feedback is considered as having a better effect on this type of task, the current study will focus on the effect of immediate feedback.

3.4.2.2 Feedback types by the complexity

Mason and Bruning (2001) summarized the eight feedback types which were included in the computer-based instruction (CBI) literature: no-feedback, knowledge-of-response, answer-until-correct, knowledge-of-correct-response, topic-contingent, response-contingent, bug-related, and attribute-isolation. Hsieh and O'Neil (2002) considered three, knowledge of response, knowledge of correct response, and elaborated feedback, as the most often used generic feedback types.

"Knowledge of response", which is also called "error-message-only feedback" (e.g. PujolÀ 2001; Miller et al. 2005), only tells the learners whether their answers are "right" or "wrong". This is the simplest form of feedback. "Knowledge of correct response" is slightly more complicated than the previous type by providing the learners with the correct answers. "Elaborated feedback" in Hsieh and O'Neil (2002) is the same with "response-contingent feedback" in the categorization in Mason and Bruning (2001), which gives explanations for the learners' correct or incorrect responses.

In terms of the effectiveness of all types of feedback, many studies showed an effect on learning (Gilman1969; Roberts and Park, 1984;Waldrop et al. 1986; McKendree1990; Pridemore and Klein, 1991; Morrison et al. 1995; Whyte et al. 1995; Rosa andLeow2004), while some others found no differences, with or without feedback (Feldhusen and Birt 1962; Moore and Smith 1964; Rosenstock et al. 1965; Hodes 1985; Merrill 1987; Park and Gittelman 1992; Clark 1993; Mory 1994; Wentling, 1973).

Knowledge of response, the only type of feedback that has only "verification" but not "elaboration", was supposed to be the least effective type of feedback. Some studies did show that knowledge of response was not as efficient as other types of feedback.

Jaehnig and Miller (2007) reviewed the results of the effects of different feedback types on learning. Four studies were comparing the effect of knowledge of response and no feedback, i.e. Gilman1969; Moore and Smith1964; Roper1977; Rosa and Leow 2004. Two studies, Moore and Smith (1964) and Roper(1977), examined the effect of knowledge of correct response and knowledge of response. Seven studies (Gilman1969; Roberts and Park, 1984;Waldrop et al. 1986; McKendree1990; Pridemore and Klein, 1991; Rosa and Leow 2004) involved elaborated feedback and knowledge of response. In all these studies, knowledge of response was never more effective than any other types of feedback.

However, the current study is not trying to decide which kind of feedback is the most effective one. Instead it is trying to examine whether feedback has any effect on the perception of Mandarin lexical tones. To exclude the possible effect of all other factors, such as reviewing or relearning the previous learnt knowledge, the current study will make use of the simplest type of feedback, i.e. knowledge of response.

3.4.3 Effect of feedback on perceptual learning

The studies on whether feedback has any influence on perceptual learning have mixed results. Although most studies on perceptual learning involve feedback that has an effect on the perceptual learning results, some other studies also argue in several other ways. Some studies found that feedback is necessary in perceptual learning under certain circumstances, while some others proved that feedback on its own could induce perceptual learning. Some other studies argued that perceptual learning could happen without any external feedback. In this section, studies of each stance will be reviewed.

Besides the studies which support that perceptual learning can be improved by feedback, some other studies believe that feedback is necessary under certain circumstances (e.g. Shiu and Pashler 1992; Herzog and Fahle 1997; Seitz et al. 2006). For example, Shiu and Pashler (1992) found little improvement for the group that received no feedback in an orientation discrimination task, while there was significant improvement with feedback. In the same study, they changed their stimuli in another experiment and found that even when perceptual learning was successful without feedback, it still needed feedback to make perceptual learning happen when the stimuli were difficult.

Feedback not only can facilitate perceptual learning, recent studies have argued that feedback can in fact induce perceptual learning (Herzog and Fahle 1999; Herzog et al. 2006; Shibata et al. 2009; Choi and Watanabe 2012). Herzog and Fahle (1999) and Herzog et al. (2006) examined the role of fake feedback in vernier discrimination tasks. They found that if the participants kept receiving fake feedback, which indicated opposite directions of the real directions, their performance would be negatively affected. However, this effect disappeared immediately after the correct feedback was given. Choi and Watanabe (2012) re-examined the fake feedback's role in a visual orientation task and found retained perceptual learning induced by fake feedback. This result was also in line with the result in Shibata et al. (2009).

However, many studies challenged the impression that feedback must affect perceptual learning by arguing that perceptual learning could happen even without external feedback (e.g. McKee and Westheimer 1978; Ball and Sekuler 1987; Karni and Sagi 1991; Poggio et al. 1992; Fahle and Edelman 1993; Crist et al. 1997; Watanabe et al. 2001; Petrov et al. 2006; Liu et al. 2012). For example, Petrov et al. (2005; 2006) studied a visual orientation discrimination task under non-stationary context without giving feedback. Robust perceptual learning was found in the task. They therefore supported the belief that perceptual learning could happen even in the absence of feedback.

The aforementioned studies argued for or against feedback's significance in perceptual learning in many different ways. However, most studies were conducted in vision researches. The role of feedback in the perceptual learning in speech studies, especially about the segmental features, is extremely understudied. The present study therefore will examine the effect of immediate, simple feedback's effect on the perception of Mandarin lexical tones by English speakers.

3.5 Summary

The perceptual pattern of lexical tones by native Mandarin speakers is different from that of the non-Mandarin speakers. Linguistic experience, i.e. tonal language experience or non-tonal language experience, affect the perception of lexical tones. Tonal language speakers perceive lexical tones in other languages, e.g. Mandarin lexical tones, in similar ways as native speakers; however, non-tonal language speakers perceive tones in different ways.

Apart from linguistic experience, the interaction between lexical tones and the segmental features or suprasegmental features of the syllables will also influence the perception of lexical tones by non-Mandarin speakers to some extent. Because of these interactions, different tones or tone pairs will have different levels of difficulty to non-Mandarin speakers.

Why native Mandarin speakers and non-Mandarin speakers perceive lexical tones differently are mainly due to two possible reasons. The first reason may be that non-Mandarin speakers cannot categorize the pitches of the speech in the same way as native Mandarin speakers can; the second reason may be that non-Mandarin speakers cannot perceive the Mandarin lexical tones as a separate tier from the segmental features of Mandarin syllables.

Although non-Mandarin speakers do not categorize Mandarin lexical tones in the same way as native Mandarin speakers do, nor do they deal with lexical tones on the same tiers as native speakers, the reality is that non-Mandarin speakers are not totally "deaf" to lexical tones. They only have different weighting patterns for different cues with the native speakers. In order to improve the perceptual ability of non-Mandarin speakers, methods of changing the weightings of their attention to different cues could be used. One way is through perceptual learning.

There are four mechanisms of perceptual learning, "differentiation", "unitization", “attentional weighting” and “stimulus imprinting” (Goldstone 1998). All these mechanisms are relevant to categorical perception, which is one main reason that leads to the difference in the perception of Mandarin lexical tones by native Mandarin speakers and non-Mandarin speakers.

Previous studies have made efforts to improve the perception of Mandarin lexical tones through perceptual learning. However, most studies made use of training for several sessions rather than feedback.

The mechanisms of feedback have developed from “reinforcement” (e.g. Skinner 1954) in early times to “consciousness-raising” (e.g. Mason and Bruning 2001) and “categorical perception” (e.g. Seger 2008) in recent studies. From the recent mechanisms, connections between feedback and lexical tone perception and perceptual learning could be easily found.

Although feedback in general has similar mechanisms with lexical tone perception and perceptual learning, different feedback types can cause different influences on perception. According to its timing and complexity, feedback can be divided into different categories. The effectiveness of delayed feedback and immediate feedback has been under debate for years. In the current study, the effect of immediate feedback will be examined. As for the complexity of feedback, the simplest feedback form (knowledge of response) will be examined in order to avoid the effects of other factors. In the next chapter, a pilot experiment will be reported.

4. Pilot experiment

4.1 Introduction

The aims of the pilot experiment were to improve the experiment design and to assess the value of the hypotheses. If any problems occurred in conduction of the experiment or the data, changes would be made to perfect the experiment procedure so as to get the most suitable results. If the results of this pilot experiment could support the hypotheses to some extent, it would thus be worthwhile to carry out further study.

Compared with the full-scale experiment, the scale of the pilot experiment in discussion was smaller. The number of participants was 12, which was half of the full-scale experiment. The number of the stimuli, 80 trials, was also half of the full-scale experiment. The procedures of the experiment were not as strictly controlled as the full-scale experiment in terms of the environment of testing, the forms of the immediate feedback, and the limit of the time that the participants could have on making “same” or “different” judgment. Due to the different apparatuses used in the pilot experiment, which were only pen and paper, the reaction time could not be calculated. The hypothesis of the pilot experiment therefore was that the error rate of the perception of Mandarin lexical tones by non-Mandarin speakers would decrease as more feedback was received.

4.2 Participants

12 native speakers of English (11 British speakers and 1 American speaker) with no known hearing or speech problems participated in the experiment. All participants were undergraduate or graduate students at Newcastle University, who participated voluntarily without receiving any incentives. None had any knowledge of or experience with Mandarin or any other tonal language.

The participants were divided into two groups of 6 participants each: one experiment group, and one control group. The experiment group received feedback after every judgment, while the control group did not receive any feedback.

4.3 Stimuli

A total of 80 pairs of monosyllabic Mandarin morphemes were selected, all of which were composed of both an onset and a rime. Monophthongs and diphthongs in the rimes were evenly numbered and distributed in the stimuli list (Appendix 1).

Some speech sounds in Mandarin phonology inventory were excluded, for example, *zh* [tʂ], *ch* [tʂ^{h}], *sh* [ʂ], *j* [tɕ], *q* [tɕ^{h}] and *x* [ɕ] in the onset position as well as [ɿ], [ʅ], [y] in the rime position. The disputable glides were also excluded. The remaining phonemes are listed in Table 4. The reason for the exclusion is that non-native sounds or syllable structures may cause perceptual confusion for the non-Mandarin speakers (Martin and Peperkamp 2011). The results of the lexical tone perception are possible to be affected since interactions exist between the lexical tones and segmental features of Mandarin syllables as previously reviewed.

Table 4: List of chosen onsets and rimes in *pinyin* and IPA

Onsets	Rimes	
b [p]	monophthongs	diphthongs
p [p^h]		
m [m]	*a* [Ą]	*ao* [ɑu]
f [f^h]	*o* [o]	*ai* [ai]
d [t]	*e* [ɤ]	*ou* [ou]
t [t^h]	*i* [i]	*ei* [ei]
n [n]	*u* [u]	*an* [an]
g [k]		*en* [ən]
k [k^h]		*in* [in]
h [x]		*ang* [ɑŋ]
		eng [əŋ]
		ing [iŋ]
		ong [uŋ]

The stimuli were divided into two categories according to their segmental features, i.e. the contexts of their onsets and rimes. Based on the results of Zeng (2008) mentioned above, the current experiment selected the two conditions that were more difficult for English speakers, namely same-onset/ different-rime (SODR) and different-onset/ different-rime (DODR). Therefore, the rimes would be always different in each trial while there were two contexts for the onsets, either the same or different. Within each context, 20 stimuli trials were “yes” trials, which means the two stimuli in the trial were of a same tone, while the other 20 were “no” trials with stimuli of different tones.

Table 5 shows the numbers of stimuli trials in each category. In “Yes” trials, there were five trials for each tone. In “No” trials, we used the contrast between the third tone (214) and other tones. According to the literature reviewed in 3.2.3.2, the third tone (214) is the most difficult tone and the tone pair of the second tone (35) and the third tone (214) is the most confusing pair; this study therefore used more stimuli with the third tone (214) and second (35)-third (214) tone pair.

Table 5: Numbers and categories of stimuli

	DODR (40)		SODR (40)	
	Tone pairs	Number	Tone pairs	Number
“Yes” trials (20)	55 + 55	5	55 + 55	5
	35 + 35	5	35 + 35	5
	214 + 214	5	214 + 214	5
	51 + 51	5	51 + 51	5
“No” trials (20)	214 + 55	5	214 + 55	5
	214 + 35	10	214 + 35	10
	214 + 51	5	214 + 51	5

4.4 Apparatuses

All the stimuli trials were recorded by a female native Mandarin speaker with Sony

ICD-PX312 digital voice recorder. An ASUS U36J laptop was used for playing the stimuli, and the judgments were taken down with pen and paper.

4.5 Procedures

The experimenter first explicitly introduced what lexical tones were in Mandarin, using *ā* [Ą] 55, *á* [Ą] 35, *ǎ* [Ą] 214 and *à* [Ą] 51 as examples and then compared the lexical tones to musical notes.

Having had a thorough idea of what Mandarin lexical tones were, the participants were asked to perform an AX discrimination task. The participants heard two stimuli one by one at a time and they needed to make a judgment on whether the tones of the two stimuli were the same, regardless of the consonants and vowels.

The experimenter gave the participants in the experiment group a feedback after each judgment by verbally telling the participants whether the judgment was "right" or "wrong". For the control groups, however, the experimenter would immediately play the next trial without giving any feedback. The experimenter took notes of the correctness of the judgments of both the experiment groups and the control groups for data analysis.

4.6 Results

4.6.1 Descriptive data

The average error rate of the control group was 16.67% while that of the experiment group decreased to 7.71% (Figure 3). This indicated that feedback had an effect on the perception of Mandarin lexical tones.

With the stimuli divided into four blocks (i.e. First Block: the first 20 pairs of stimuli; Second Block: the second 20 pairs of stimuli; Third Block: the third 20 pairs of stimuli; Fourth Block: the fourth 20 pairs of stimuli), the average error rates by blocks of the control group were 21.67%, 19.12%, 11.67% and 14.17% respectively while those of the experiment group were15.00%, 10.00%, 1.67% and 4.17% respectively (Figure 4). These results indicated that as more feedback was received, the perception abilities of both groups were improving; moreover, the experiment group improved at a higher speed and reached a higher level than the control group.

Figure 3: Average error rates (pilot experiment)

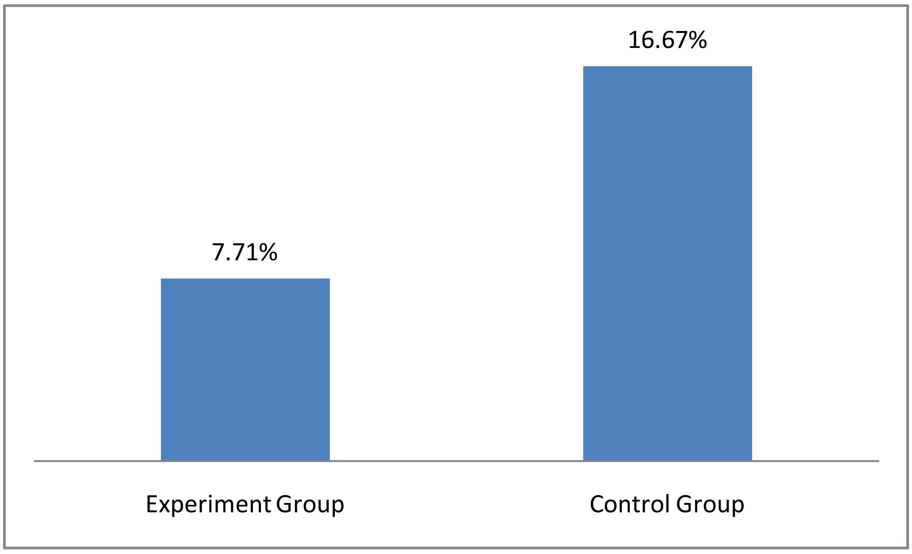
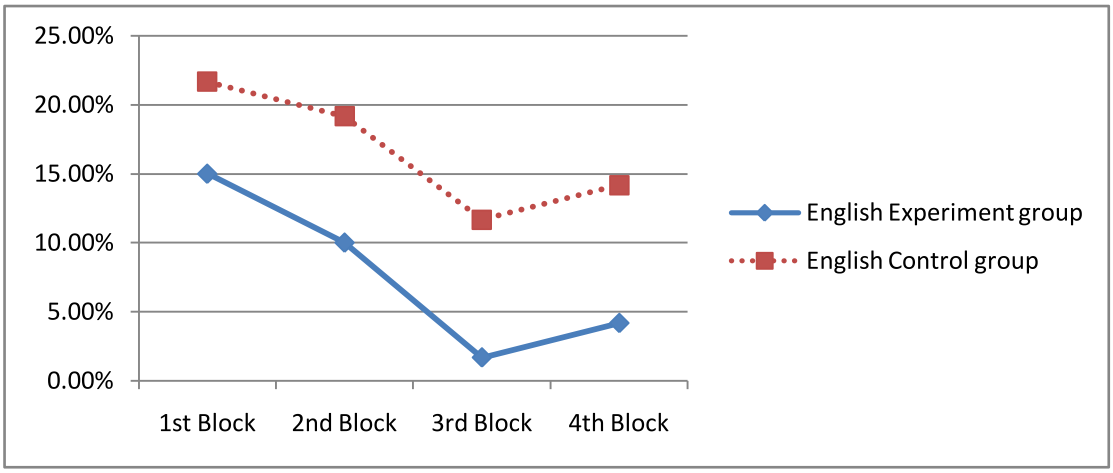


Figure 4: Average error rates by blocks (pilot experiment)



4.6.2 Statistical data analysis

A 2×2×4 repeated-measures ANOVA was performed on the error rate, with three factors Context (same onsets, different onsets), Block (1^{st}, 2^{nd}, 3^{rd}, 4th) and Feedback (with, without). Context and Block are within-subject factors and the between-subject factor was Feedback.

The results showed two significant effects, Block effect and Feedback effect. The main effect of Block was significant, $F(1, 10) = 5.218$, $p = .005$. The main effect of the between-group factor is also significant, $F(1, 10) = 9.163$, $p = 0.013$. However, the Context effect, $F(1, 10) = 3.959$, $p = .075$, was not statistically significant. There was not any significant interactional effect either.

The Block effect suggested that as the participants heard more stimuli, with or without feedback, their perception had been significantly improved. This result supported the theories on perceptual learning as the participants could perceive the lexical tones significantly better after a period of exposure without noticing the learning process.

The Feedback effect revealed that the experiment group receiving feedback perceived the Mandarin lexical tones significantly better than the control group that did not receive any feedback. Therefore, the hypothesis is supported by the data in this pilot study that feedback had significant effect on the perception of Mandarin lexical tones by English speakers in terms of accuracy.

4.7 Discussion

Although the results were in accordance with the hypothesis, there were two outstanding issues, namely the low error rate and the fourth block rebound.

The low error rate may be due to the easy task form. According to Zeng (2008), English speakers generally made more than 50% errors and even the native Mandarin speakers had an average error rate of 21.5%. However, in this pilot study, the control group, who made more mistakes than the experiment group, only had 16.67% of errors. This suggested that the task might have been too easy for the English speakers. There was not any limitation for their response so that they could think for as long as they were willing to. However, such delayed judgment may reveal the conscious processing of the Mandarin lexical tones rather than their perception. Therefore, in the full-scale experiment, the task form was changed to reach a higher difficulty level, in terms of many aspects including setting an expiry time for the participants' responses. In addition, a group of native Mandarin speakers were also recruited in the full-scale experiment to act as controls for the purpose of examining the difficulty of the task.

The error rate of the fourth block was expected to be the lowest among all four blocks; however, it rebounded from the 1.7% in the third block to 4.2%. This might be due to the fixed stimuli order since the control group had the same curve as shown in Figure 2. Therefore, in the full-scale experiment, the stimuli order was randomized for every participant. It may also be attributed to the concentration span of the participants. To avoid this problem, there will be breaks after each block in the design of the full-scale experiment. It may also be because of a ceiling effect. If so, this problem will need further research.

4.8 Summary

This chapter reported a pilot experiment that was conducted before the full-scale experiment in order to assess the value of the hypotheses in the full-scale experiment and improve the design of the full-scale experiment. It hypothesized that the error rate of the perception of Mandarin lexical tones by non-Mandarin speakers would decrease as more feedback was received. This pilot study tested the hypothesis with an AX discrimination task, examining the perception of Mandarin lexical tones by 12 native English speakers. The result showed that the participants who received immediate feedback had fewer errors than the participants who received no feedback in average. The statistical analysis revealed significant block effect and feedback effect, which meant that: 1) the participants' error rate decreased as they were more familiar with the stimuli; 2) the experiment group who received feedback perceived significantly better than the control group who received no feedback.

The outstanding issues, i.e. the low error rate in general and the fourth block rebound, indicated that improvements were needed in the full-scale experiment. Therefore, in the full-scale experiment, the task difficulty would be raised, the stimuli order would be randomized for each participant, and breaks would be added between each block. The next chapter will report the experiment design and the results of the full-scale experiment in detail.

5. Full-scale experiment

5.1 Introduction

The results of the pilot experiment showed that there was significant effect of feedback on the perception of Mandarin lexical tones by English speakers in terms of accuracy but the experiment task needed modification. Therefore, in the current full-scale study, the AX discrimination task (Cutler and Chen 1997; Zeng 2008) was performed on a computer with DMDX software to calculate the error rate as well as the reaction time. In order to raise the task difficulty, the stimuli number was doubled. The participant number also doubled in order to gain more accurate results. The hypotheses are: 1) the error rate of English speakers' perception of Mandarin lexical tones will decrease as receiving more feedback; 2) the reaction time of English speakers' discrimination of Mandarin lexical tones will be shortened as receiving more feedback, even after the accuracy rate stops improving.

5.2 Participants

24 British native English speakers and 5 Chinese native Mandarin speakers were recruited to participate in this experiment. All participants were undergraduate or post-graduate students at Newcastle University, who did not have any known hearing or speech problems. None of the British participants had any knowledge of or experience with Mandarin or any other tonal language. The participants received a small amount of incentive for their participation. The experiment lasted for

approximately 20 to 25 minutes for each person.

The British participants were divided into two groups: 12 English speakers formed an experimental group, while the other 12 English speakers acted as a control group. The experimental group received feedback after each judgment, while the control group did not. The Chinese participants acted as controls for the experiment, who did not receive any feedback either.

5.3 Stimuli and apparatuses

The stimuli were produced and recorded by a male native Mandarin speaker in a phonology laboratory, which were digitized with a sampling rate of 44 kHz. The stimuli consisted of two sets, one as pre-test stimuli and filler stimuli and the other as post-test stimuli. All stimuli were natural monosyllabic morphemes in Mandarin, all of which consisted of both an onset and a rime. The selection of onsets and rimes was the same with the pilot study (Table 4).

The post-test stimuli (Appendix 2) were composed of 172 trials monosyllabic Mandarin morphemes, with two syllables in each trial. The trials were divided into four blocks, after which the participants had a chance to have a break. The 43 trials in each block were evenly distributed according to their tone pairs, tone orders, rime features as in the pilot study. The proportion of each category was the same as those in the pilot study (Table 5), but the number of each category doubled. The order of the stimuli was randomized for every participant. There were not any same judgments for more than three consecutive trials in order to avoid cognitive inertia. Before each block, there were three trials which were drawn from the pre-test stimuli to be used as fillers, the results of which were not counted into the final analysis. The pre-test stimuli (Appendix 3) made use of the onsets and rimes that had been included in the post-test stimuli, which were or were not familiar to the English speakers. In the final data analysis, the fillers would not be taken into account.

The tests were conducted on a Dell DCNE desktop computer with DMDX software in a psycholinguistics laboratory. All stimuli were programmed into DMDX software, which recorded the correctness and the reaction time of each judgment.

5.4 Procedure

5.4.1 Pre-test

The pre-test was used to test the initial perceptual abilities of the participants in both the experiment group and the British control group. The results are comparable only when the initial abilities of the two groups are equal; as otherwise, the improvement can be due to the difference in initial abilities. The Chinese control group also completed the pre-test only to maintain consistency between the tasks.

At the beginning of the experiment, the participants were asked to sign consent forms (Appendix 4). The consent form contained information on the title of the study, the purpose of the study, participants' task, the incentive information and data protection. A short questionnaire was attached at the end, which investigated the participants' handedness, ages, genders, language backgrounds, music training backgrounds, and hearing conditions.

The participants were explicitly informed of the features of the four Mandarin tones after signing the consent form. As in the pilot study, the experimenter used *ā* [Ą] 55, *á* [Ą] 35, *ǎ* [Ą] 214 and *à* [Ą] 51 as examples to illustrate Mandarin lexical tones, then compared the lexical tones to musical notes. Despite the fact that there was specific description of participants' task in the consent form, the experimenter repeated the task to make sure that the participants fully understood the task. They would hear two syllables consecutively and their task was an AX discrimination task (Culter and Chen 1997; Zeng 2008), in which the participants were expected to judge whether the tones of the two syllables in each trial were the same or not, regardless of the consonants and vowels. The participants made their decision by pressing the right "shift" key on a computer keyboard for the same tones and "N" key for the different tones. "N" key had a negative indication as in "No", which helped participants to remember its purpose. The participants did not get any feedback despite of their judgment. Participants were encouraged to respond as fast and accurately as possible.

The pre-test had 40 trials in total. Each trial was made up of two Mandarin syllables, which were separated by a 500ms interval. Participants had up to 4s after each trial to make a judgment. Before each trial, there was a fixation cross shown on the screen for 500ms to help the participants concentrate. The design of right "shift" key took the placement of participants' left and right hands into consideration. After each button pressing by the participants or the 4s automatic expiry period, there was a 1s interval before the next trial was played.

5.4.2 Post-test

The post-test took place immediately after the pre-test when the experimenter was sure that the participants completely and correctly understood the task. The post-test was the same with the pre-test in terms of the participants' task. They differed with respect to the total number of the trials and the feedback.

In the post-test, there are 172 trials in total for each participant. 63 consecutive trials made up a block, in which the first three were fillers. Between each block, there was a chance for the participants to have a short break to avoid problems caused by the limit of concentration span.

For the experiment group, feedback was given on incorrect judgments while no feedback was given to the control group. The feedback, in the form of a word "Wrong" on the computer screen, appeared for 500ms after the button pressing of an incorrect judgment or the 4s expiry period. After pressing the button for a correct judgment, or receiving a feedback for an incorrect judgment, there was a 500ms interval before the next trial. For the control group, who did not receive any feedback, the procedures were the same with the pre-test.

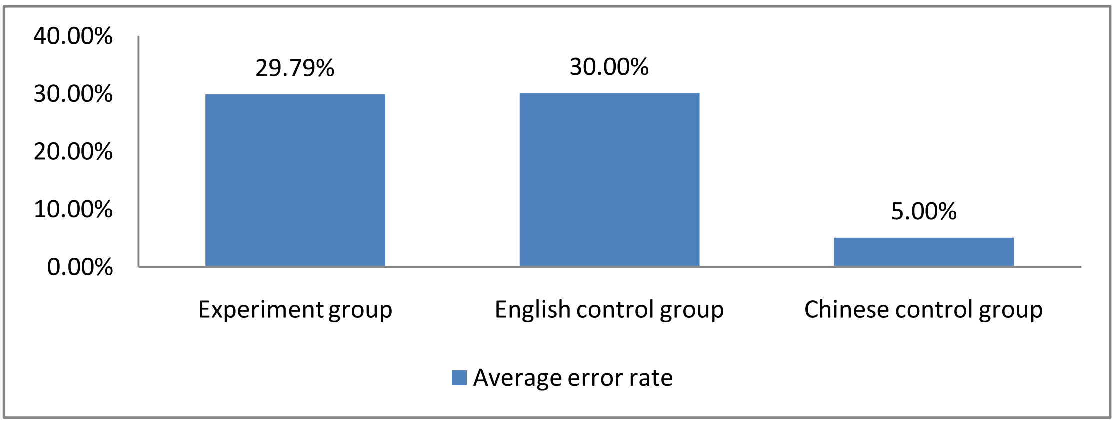
5.5 Results

5.5.1 Pre-test

5.5.1.1 Error rates

The average error rates of both British groups were on the same level. The average error rate of the experiment group was 29.79% while that of the British control group was 30.00%. The Chinese control group's average error rate was 5.00%. (Figure 5)

Figure 5: Average error rates in the pre-test



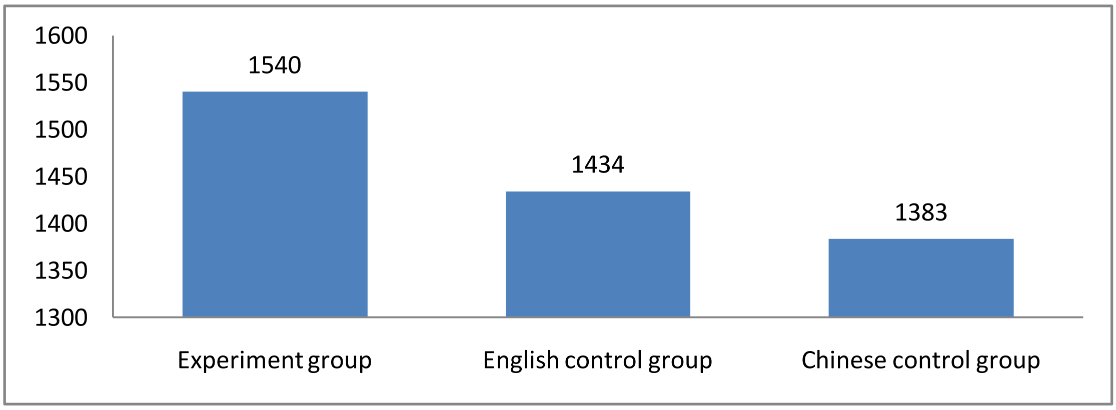
An independent t-test was performed on the error rates of the participants in the Experiment group and the participants in the English control group to test the variances of the error rates of the two groups. The result showed no significant variance, t (22) = -0.49, p= .819, which suggested the error rates of the two groups were equal variances. The initial levels of the two groups in terms of accuracy therefore were on the same level.

5.5.1.2 Average reaction times for correct responses

Judged by the mean scores, the reaction times for correct responses[1] displayed a certain gap between the English groups. On average, the experiment group used longer time to react to each stimulus, which was 1540ms, while the average reaction time of the English control group was 1434ms. In analyzing the final results, the gap was taken into consideration and was found not to have any influence on the analysis. The Chinese control group's average reaction time was 1383ms. (Figure 6)

Figure 6: Average reaction times for correct responses in the pre-test

(in milliseconds)



An independent t-test was also performed on the reaction times of the participants in both English groups to decide whether their levels are equal in terms of reaction speed. The result showed no significant variance, t (22) = .818, p= .282, which meant the reaction times of the two groups were also equal variances. The initial levels of

[1] The reaction times for incorrect responses were excluded due to the different psychological processing mechanism between the correct responses and incorrect responses.

the two groups in terms of reaction speed therefore were on the same level as well.

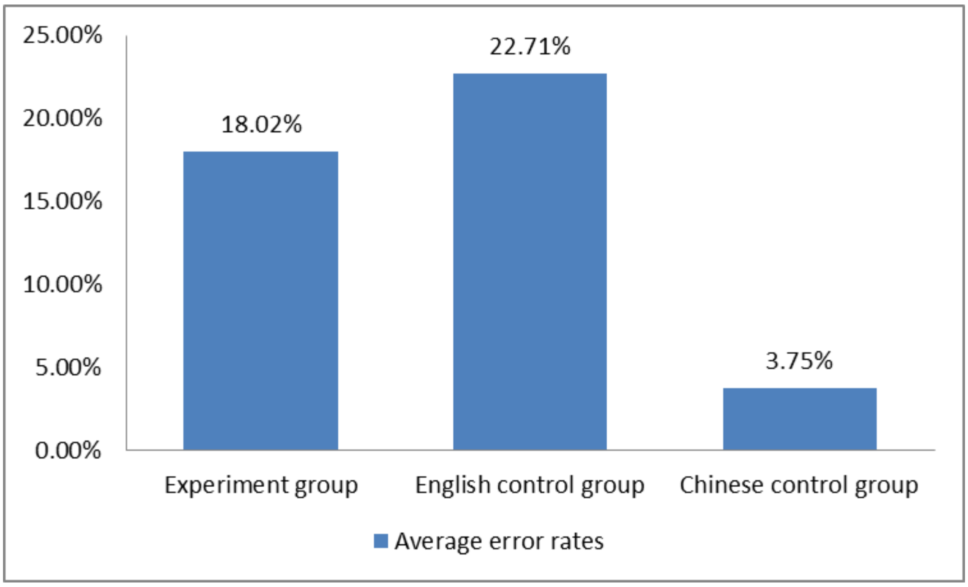
5.5.2 Post-test

5.5.2.1 Descriptive data

5.5.2.1.1 Average error rates

The average error rate of the English control group was higher than the experiment group while the Chinese control group undoubtedly had the lowest average error rate. The average error rate of the English control group and the English experiment group was 22.71% and 18.02%, while that of the Chinese control group, without any surprise, performed much better than the English groups, which had an error rate of 3.75% (Figure 7).

Figure 7: Average error rates in the post-test



With the stimuli divided into four blocks, the average error rates by blocks of the English control group were 23.33%, 24.79%, 22.29% and 20.42% respectively while those of the experiment group were 20.63%, 20.21%, 16.25% and 15.00% respectively. The Chinese control group had much lower error rates in any of the four blocks, which were 2.50%, 5.50%, 3.50% and 3.50%.

Figure 8: Average error rates by blocks (post-test)

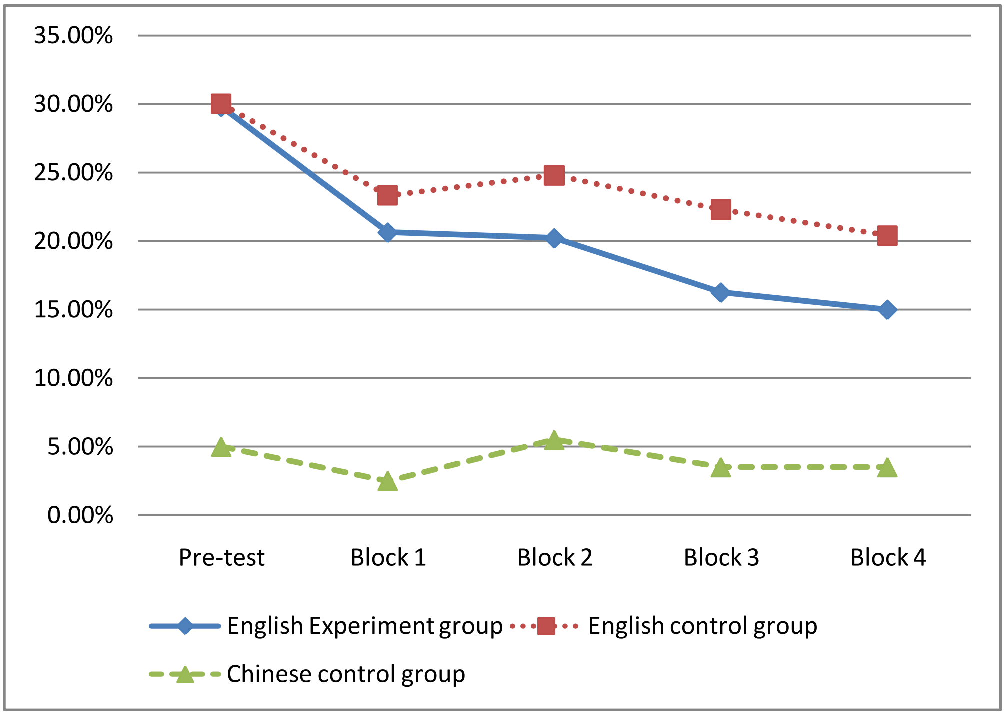
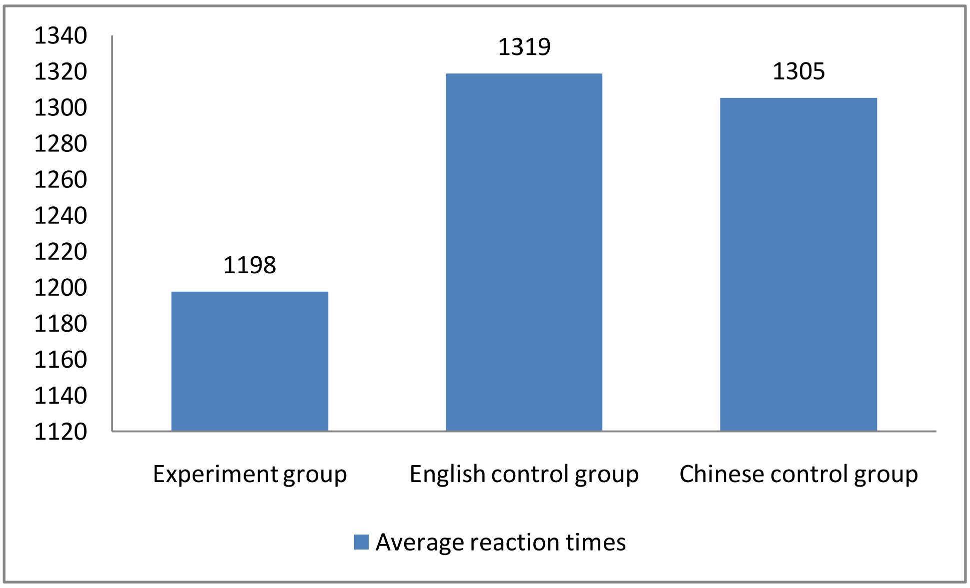


Figure 8 clearly shows the trend of the accuracy of the three groups. The English experiment group and the English control group started at basically the same accuracy level in the pre-test; as the experiment group received more feedback, its error rate showed decrease in every block, while the English control group, which revealed similar pattern as the native Chinese group, had less decrease than the English experiment group.

5.5.2.1.2 Average reaction time for correct responses

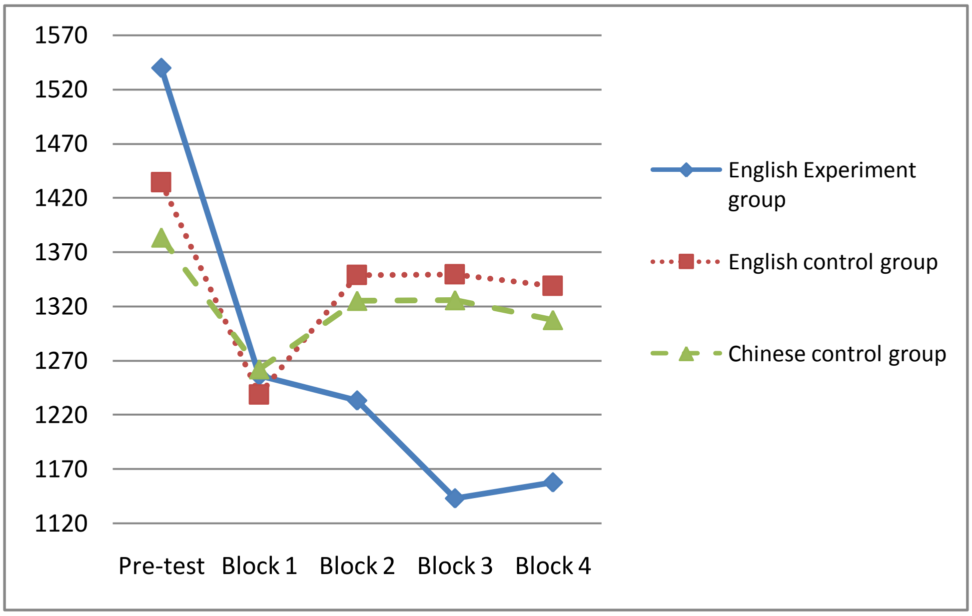
The reaction times for the correct responses of the English control group and the English experiment group was 1319ms and 1198ms, while that of the Chinese control group was 1305ms (Figure 9).

Figure 9: Average reation times for the correct responses in the post-test (in milliseconds)



With the stimuli divided into four blocks, the average reaction times for the correct responses by blocks were 1238ms, 1349ms, 1349ms and 1339ms respectively for the English control group, while those of the experiment group were 1256ms, 1233ms, 1143ms and 1158ms respectively. The Chinese control group had 1262ms, 1325ms, 1326ms and 1307ms as the average reaction times for each block. The data are shown in Figure 10.

Figure 10: Average reaction times for the correct responses by blocks (post-test) (in milliseconds)



In the pre-test, the English experiment group used the longest time to respond. However, at the end of the experiment, the English experiment group responded much faster than the pre-test or the other groups. The English control group again showed a similar pattern as the native Chinese control group but ended up with longer reaction time at the end of the experiment.

5.5.2.2 Statistical data analysis

To assess the significance of the effect of the variables, a 2×2×4 repeated-measures

ANOVA was conducted separately for the error rates and the reaction times for the correct responses. The within-group dependent variables were contexts of the onsets (same, different), blocks (1, 2, 3, 4) and the between-group dependent variable was the group (experiment group, English control group).

In error rate, the ANOVA results only showed a significant Context effect, $F(1, 22) = 4.364$, $p<.05$. The other effects were not significant. The Block effect showed $F(1, 22) = 2.331$, $p=.082$. The between group variable, Feedback effect was $F(1, 22) = 1.537$, $p=.228$. No significant interactional effect was found. This result means that different onset contexts affect the perception of Mandarin lexical tones by English speakers in terms of error rate.

For the reaction times for correct responses, there was no significant main effect or interactional effect. The Block effect was $F(1, 22) = 0.586$, $p=.626$; the Context effect was $F(1, 22) = 0.066$, $P=.800$. The between group variable, Feedback effect, was $F(1, 22) = 2.032$, $p=.168$.

5.6 Discussion

The present experiment is aimed at investigating the effect of immediate and simple feedback on the perception of Mandarin lexical tones by English speakers in terms of accuracy and reaction time. The results show that the given feedback did not affect the perception significantly. However, the descriptive data does show a trend of influence. Besides, different onset contexts had significant effect on the accuracy of perception. This result indicated that the onsets context (same onsets or different onsets) had an influence on the perception of Mandarin lexical tones by English speakers, which suggested that the difficulty levels were significantly different when the onsets were the same as opposed to when they were different. Therefore, the onsets, or in another word consonants, influenced the perception. For future studies, care should be taken on this point when designing the experiment involving different onset conditions.

An unexpected issue from the results was that the native Chinese control group used longer time to respond to the trials than the English experiment group. This result in fact is consistent with previous findings. In Lee and Nusbaum (1993) and Repp and Lin (1990), Chinese speakers also had longer reaction times than the English speakers at tone discrimination tasks. Repp and Lin (1990) attributed this to higher accuracy of the Chinese speakers' judgments. Cutler and Chen (1997) also found in an AX discrimination task on the perception of Cantonese lexical tones, Dutch speakers responded faster than native Cantonese speakers. They considered the reasons as the better understanding of reaction time experiments by the Dutch participants according to their education backgrounds, or that the Dutch speakers were not affected by the process of understanding the meaning of the syllables, etc. Although the reasons for this phenomenon should be further studied, most studies show that non-native speakers' reaction times are shorter than native speakers'.

The other unexpected issue was that only the Context effect in error rates was significant. According to previous studies, feedback was very likely to have an effect on lexical tone perception by non-Mandarin speakers. The pilot experiment also supported the hypothesis. However, the result of the current full-scale study did not show significant effects for the between-group variable. The possible reasons to account for this result will be analyzed in the next chapter.

5.7 Summary

The current full-scale experiment also made use of an AX discrimination task to test the perception of Mandarin lexical tones by 24 native British English speakers, with 12 in the experiment group that received feedback on their judgments and the other 12 as a control group that did not receive any feedback. Five native Mandarin speakers acted as another control group for the experiment. The statistical analysis showed that different onset context would cause significantly different perception accuracy by the experiment group compared with the English control group. No feedback effect or block effect was found from the results of the experiment. However, based on the simple descriptive data analysis, the performance of the experiment group was in accordance with the hypotheses. The average error rates of the experiment group were lower than the English control group after receiving feedback. The average reaction times for the correct responses of the experiment group were also much shorter than the English control group. In the next chapter, a general discussion on the results is provided.

6. Discussion and Conclusion

6.1 Introduction

The present study examined the effect of immediate and simple feedback on the perception of Mandarin lexical tones by English speakers. The result did not show any statistically significant feedback effect; however, in the descriptive data, the average error rates and reaction times of the experiment group that received feedback showed larger decrease than the control group that did not receive feedback. In this section, the results of the current study will be discussed in terms of: 1) the contribution to previous theories and studies; 2) possible factors that affected the effectiveness of the immediate and simple feedback in the experiment. Finally, a conclusion will be drawn and suggestions for future studies will be summarized.

6.2 Feedback

6.2.1 Evidence for previous studies on feedback

Section 3.4.2 was a review of different types of feedback and their effectiveness which were examined by the previous studies. This study tried to provide evidence for the effectiveness of immediate and simple feedback; however, the effect of immediate and simple feedback was not significant in the final results. The reason could be due to the inefficiency of the immediate feedback in comparison with the delayed feedback (e.g. Kulhavy and Anderson 1972; Bardwell 1981, as reviewed in Section 3.4.2.1); or it could be that the simple feedback, i.e. knowledge of response, is not an efficient type of feedback (e.g. Jaehnig and Miller 2007, as reviewed in Section 3.4.2.2). However, the most possible reason could be that the combination of the two forms of feedback is not efficient enough to reveal any significant effect in a short experiment of 20-25 minutes. Detailed analyses are discussed in the next section.

In 3.4.3, the effect of feedback in perceptual learning studies was reviewed, which showed that the effect of feedback varies from study to study (e.g. Shiu and Pashler 1992; Herzog and Fahle 1999; Petrov et al. 2006). The results of the current study contributed to the view that feedback is not always necessary for perceptual learning to occur, converging with many studies (McKee and Westheimer 1978; Ball and Sekuler 1987; Karni and Sagi 1991; Poggio et al. 1992; Fahle and Edelman 1993; Crist et al. 1997; Watanabe et al. 2001; Petrov et al. 2006; Liu et al. 2012).

6.2.2 Possible feedback factors for the absence of feedback effect

6.2.2.1 Complexity of feedback

The current study made use of the simplest form of feedback, namely knowledge of response, which only informed the participants of the incorrectness of their errors. This type of feedback is the only feedback form that does not contain elaborated information such as the repetition of the former knowledge. Most studies have shown that knowledge of response is not an efficient feedback type as reviewed in 3.4.2.2 (e.g. Jaehnig and Miller 2007).

To account for the failure of knowledge of response, the Noticing Hypothesis can provide a reasonable explanation. Schmidt (1990) advanced the Noticing Hypothesis. He believed that only the noticed part of the input is effective; therefore, language features could only be learnt when they are noticed. There are oppositions to the strong version of this hypothesis, such as Truscott (1998) who argued that Noticing Hypothesis has weak cognitive and conceptual foundations, which make the hypothesis hard to interpret or test; nevertheless, many scholars (Ellis 1993;1994; Foto and Ellis 1991; Fotos 1993; 1994; Harley 1993; Larsen-Freeman et al. 1991; Long 1991; Robinson 1995; 1996; Zalewski 1993) supported the weak version of this hypothesis, arguing that noticing is necessary but not sufficient for language acquisition. In the current study, knowledge of response, the simplest form of feedback, resulted in being too simple to draw the English speakers' attention to the correct cue, i.e. f0 contour, so that no significant effect of feedback was found. This result therefore provided perceptual evidence for the Noticing Hypothesis.

Since one of the mechanisms for feedback is consciousness-raising, feedback at different complexity levels (see Section 3.4.2.2 for detailed categorization of feedback) raise the consciousness of the participants to different extents. The more elaborated the feedback is, the more the correct cues will be noticed by the participants. The simplest type of feedback only informs the participants of the correctness of their judgments but does not direct their attention to the information that needs addressing, i.e. f0 contour in the present study. Therefore, knowledge of response, as the simplest feedback type, does not work well in improving the perception of speech, including the perception of

Mandarin lexical tones.

Similar findings about knowledge of response, which indicate the correctness of judgments without providing any elaborated information, can also be found in child language acquisition studies. Brown and Hanlon (1970) observed that parents seldom commented children's utterances as "right" or "wrong"; instead, more corrective feedback that contained more elaborated information were often used. Chomsky (1965; 1980; 1986) ascribed this phenomenon to the lack of necessity of explicit learning. However, whether the phenomenon was due to the lack of necessity or sufficiency is worth to be studied further. The result of the current study suggested that in adult perception, explicit but simple feedback is not sufficient. Since adult non-native language perception and child language acquisition have many aspects in common (such as the awareness level), the studies in child language acquisition that found similar phenomenon to the current study but explained it differently might need more evidence.

6.2.2 Timing of feedback

The current study gave the participants trial-by-trial immediate feedback instead of delayed feedback. According to Norris et al. (2000; 2003), the difference between online (or immediate) and offline (or delayed) feedback lies on the feedforward mechanism. Online feedback not only feed "back" but also feed "forward", which means if a participant receives a feedback after each trial, the feedback does not only help the participant to reflect on the past trial, but also provides information for the future trials.

However, knowledge of response only provides the correctness of the participants' judgment, which could not provide enough feedforward information. Norris et al. (2000) argued that feedback was nothing useful but that it would only cause confusion. Therefore, the combination of the immediate trial-by-trial feedback and knowledge of response might not be an effective combination to improve the perception of Mandarin lexical tones.

In the pilot experiment, although the feedback form was the same with that in the full-scale experiment, the allowed reaction time, nevertheless, was substantially longer than the full-scale study owing to the lack of controlling. Therefore, it was possible for the participants to introspect the feedback for a longer time and transform the feedback information into feedforward information.

6.3 Perceptual learning

6.3.1 Evidence for previous studies on perceptual learning

In Section 3.3, various aspects of perceptual learning were reviewed. Successful perceptual learning means that people could pick up information that "are independent of conscious forms of learning" (Fahle and Poggio 2002, in Section 3.3.1). From the descriptive data of the current study, a decreasing trend can be found in both the accuracy and the reaction time. This result indicates that perceptual learning happened during the process of the experiment, as the participants getting increasingly familiar with the stimuli. Therefore, although most studies used training as a condition for assisting perceptual learning to happen (e.g. Wang et al. 2003; Kraljic and Samuel 2005; Francis et al. 2008; Carcagno and Plack 2011, reviewed in Section 3.3.3), the current study supports the argument that perceptual learning can occur under unsupervised conditions which include little or no training or feedback (Norris et al. 2003; Petrov et al. 2006; Molley et al. 2012, reviewed in Section 3.3.3)

6.3.2 Possible perceptual learning factors for the absence of feedback effect

The inconsistency between the pilot experiment results and the full-scale experiment results indicated possible reasons of individual differences of the participants, for example, the participants' perceptual learning preferences, learning rates, concentration, or motivation.

Different people have different learning rates. Firstly, perceptual learning through receiving immediate feedback within a short period of time requires the participants to have fast learning ability. If the participants could not process the given feedback information at a high speed, the feedback would only become confusing instead of helpful. Therefore, if the participants have different learning rates, the effect for the slower learners may not be present.

In addition, some studies showed that perceptual learning may not happen immediately after receiving training or feedback. Kulhavy and Anderson (1972) forwarded the theory of "Delay-Retention Effect", which indicated that the delayed feedback had robust effects in the retention tests. Some other studies also found significant effects in the retention tests by receiving immediate feedback (e.g. Dihoff et al. 2010). The current study did not involve a retention test; we therefore could not find how much long-term improvement has been created by the perceptual learning. Therefore, we could not exclude the possibility that the feedback was only effective after a period of subliminal processing.

Other than the rate of perceptual learning, the preference for perceptual learning styles is also a possible reason that causes the difference between the pilot experiment results (feedback effect was significant) and that of the full-scale experiment (feedback effect was not significant). In the pilot experiment, the feedback was given in the form of audio, while in the full-scale experiment the feedback was visually presented on the screen. During an audio discrimination task, the audio feedback would undoubtedly arouse the attention of the participants; while the extent of attention that visual feedback could draw is different for participants with different learning styles.

Besides the different rates and style preferences of perceptual learning, some other individual differences, such as concentration or motivation, could also alter the participants' results. Compared with longer-term exposure-based perceptual learning, the feedback-based perceptual learning – especially trial-by-trial immediate and simple feedback given within a short period of time – relies more on the participants' sub-conscious processing of the feedback. Therefore, if the participants did not concentrate enough or did not feel motivated enough, the feedback might not have any effect on their judgments.

6.4 Tone perception

6.4.1 Evidence for previous studies on lexical tone perception

Section 3.2 reviewed the lexical tone perception by native Mandarin speakers, tonal-language speakers, and non-tonal language speakers. It concluded the differences between the perceptual abilities of the speakers with different language backgrounds could be due to Categorical perception (tonal-language speakers perceive lexical tones categorically while non-tonal language speakers do not) and/or Autosegmental Theory (native speakers perceive lexical tones automatically separate from the segments while non-native speakers cannot) as reviewed in 3.3.4.

The result of the current study did not show significant feedback partially due to the reason that the feedback was too simple to provide sufficient information for the English speakers to draw more attention to the same cue as the native speakers used, i.e. f0 contour, as reviewed in 3.2.1 and 3.2.3. Moreover, the unsuccessful use of feedback indicated that the simple form of feedback failed to help the English speakers perceive the lexical tones separately from the vowels. After the experiment, some participants expressed that the changes of vowels caused some confusion on their judgment, which happened to support Autosegmental theory (Goldsmith 1979). Future studies could therefore interview the participants after the experiment to examine the elements, of which the participants are aware, in order to understand the underlying reasons of the studies better.

6.4.2 Possible tone perception factors for the absence of feedback effect

6.4.2.1 Stimuli complexity

The task used in the current study was essentially the same with that in Zeng (2008); however, the error rate results showed a fairly big gap. Zeng (2008), as reviewed in Section 3.2.3.1, found that the English speakers had an average error rate of 49% for the conditions in which the rimes were different and the native Chinese speakers' error rate was 21.5%; while in the current study, the average error rate of the experiment group was 18.02% and 22.71% for the English control group and 3.75% for the Chinese control group. The perception of all participants, including native speakers and non-native speakers, was better in the current study by a large margin. The only difference between the task in Zeng (2008) and the current study lies on the difficulty of the stimuli. In the current study, non-English consonants and vowels were excluded (see Section 5.3). From the differences between the results of the two studies, we can infer that the non-native segments caused difficulty for the English speakers on their perception of Mandarin lexical tones. Future studies may investigate how much non-native segments pose difficulty to the perception of Mandarin by non-Mandarin speakers.

6.4.2.2 Musical experience

Since Mandarin lexical tones are very similar to music pitches, some recent studies have shown significant interaction between music experience and tone perception (e.g. Burnham and Brooker 2002; Alexander et al. 2005; Wong et al. 2007; Cooper and Wang 2012). These studies unanimously showed that musical experience had a positive influence on the perception of lexical tones.

Among all the participants in the current study, 9 out of 12 participants in the experiment group and the same number of participants in the English control group, i.e. 9 out of 12, believed they had “musical experience”, which referred to formal musical training. However, 7 out of 9 participants with musical experience in the experiment group stated they had more than six years of musical training while only 3 in the English control group considered themselves as having more than six years of musical experience. The imbalance might have had an effect on the final results. Future studies on lexical tone perception could draw more attention to the control of musical experience.

6.5 Conclusion

This study has examined the effect of immediate and simple feedback on the perception and showed no significant effect. However, the trend lines of the mean error rates and reaction times showed that the participants in the English experiment group who received immediate feedback decreased more than the English control group, both by blocks and in general. In addition, an onset context effect for the error rate is found in the AX discrimination task, which means that different onset conditions may cause different perceptual difficulties for the non-Mandarin speakers.

The insignificant effect of the feedback on the perception of Mandarin tones by English speakers may lie on the complexity and timing of the feedback. The results could lend support to the Noticing Hypothesis (Schmidt 1990). Other reasons include the individual differences on perceptual learning, task design, and the influence from other factors such as musical experiences.

Some debated issues can find support in the results of the current study. For example, the results supported that lexical tone perception by native Mandarin speakers is categorical and autosegmental; perceptual learning is effective under unsupervised situation; and feedback is not efficient when combining immediate feedback and knowledge of response, etc. (See Section 6.2.1, Section 6.3.1, Section 6.4.1 for detailed discussion).

Possible reasons for the inefficiency of the feedback, as well as the reasons for the differences between the pilot study and the full-scale study, were discussed. Different feedback types have different effects on the perception. The simplest form of feedback, i.e. knowledge of response, may be too simple to make any difference within a short period of time. The feedback timing also influences the final effect that feedback plays on the perception. Immediate feedback not only serves as a reflection of the previous information but also a cue for the future test items. Knowledge of response, nevertheless, is too simple to provide information for the “feedforward” mechanism of the immediate feedback.

Besides feedback, the rates and style preferences of perceptual learning also affect the study results. Individual differences on the speed of adapting to new things influence the research results in the studies with short test time. The perception results might be different after a period of time. A retention test is therefore needed to make sure of the final effect of the immediate feedback on the perception. Some other individual differences such as concentration and motivation also affect the test results.

From the perspective of tone perception, the study results might be due to the influence from the stimuli difficulty and musical experience. The AX discrimination task used in the current study employed simpler stimuli, which excluded the non-English consonants and vowels, and resulted in lower difficulty levels for the English speakers. The results could draw the attention of future studies to the influence of the segment difficulty on Mandarin lexical tone perception. Besides the task design, the participants’ different levels of musical experience may also be a reason as to why the final results in the current study were not in accordance with the hypotheses.

6.6 Future studies

The current study examined the effect of the immediate and simple feedback on the perception of Mandarin lexical tones by English speakers and found no significant feedback effect. According to the results, the simplest form of feedback, i.e. knowledge of response, is not recommended for future experiments on feedback or lexical tone perception. In addition, teaching and learning in classrooms or through computer assisted language learning should also use more elaborated form of feedback to improve the efficiency.

Despite the insignificant feedback effect, an onset context effect was found in the current study. This can serve as a ground for the future studies that segmental features have a significant effect on lexical tone perception. Therefore, more studies could focus on this point.

There are several other suggestions of understudied topics for the future studies mentioned in the discussion of this paper:

1) Why is the reaction time of the non-native speakers' perception on lexical tones shorter than the native speakers'? (Section 5.6)

2) Is the phenomenon that parents seldom use the explicit negative feedback during children's acquisition of their first language due to the lack of necessity or is it due to the lack of sufficiency of the explicit negative feedback? (Section 6.2.1)

3) How much do the non-native segments pose difficulty to the perception of Mandarin by non- Mandarin speakers? (Section 6.4.1)

4) How much does musical experience affect the perception of Mandarin tones? (Section 6.4.2)

Bibliography

Abramson, A. S. 1975. 'The tones of central Thai: Some perceptual experiments'. *Studies in Tai linguistics in honor of William J. Gedney. Bangkok: Central Institute of English Language:* 1-16.

Alexander, J. A., P. C. M. Wong and A. R. Bradlow. 2005. 'Lexical tone perception in musicians and non-musicians'. *9th European Conference on Speech Communication and Technology*. Lisbon, Portugal.

Aljaafreh, A. and J. P. Lantolf. 1994. 'Negative feedback as regulation and second language learning in the zone of proximal development'. *The Modern Language Journal* 78: 465-483.

Annett, J. 1969. *Feedback and human behaviour: The effects of knowledge of results, incentives and reinforcement on learning and performance*. Penguin.

Aoyama, K., J. E. Flege, S. G. Guion, R. Akahane-Yamada and T. Yamada. 2004. 'Perceived phonetic dissimilarity and l2 speech learning: The case of Japanese /r/ and English /l/ and /r/'. *Journal of Phonetics* 32: 233-250.

Ball, K. and R. Sekuler. 1987. 'Direction-specific improvement in motion discrimination'. *Vision research* 27: 953-965.

Bardwell, R. 1981. 'Feedback: How does it function?'. *The Journal of Experimental Educational*: 4-9.

Best, C. T. 1995. 'A direct realist view of cross-language speech perception'. *Speech Perception and Linguistic Experience : Issues in Cross-Language Research*: 171-204.

Best, C. T. and M. D. Tyler. 2007. 'Nonnative and second-language speech perception: Commonalities and complementarities'. *Second language speech learning: The role of language experience in speech perception and production*: 13-34.

Bluhme, H. and R. Burr. 1971. 'An audio-visual display of pitch for teaching chinese tones'. *Studies in Linguistics* 22: 51-57.

Bradlow, A., D. Pisoni, R. AkahansYamada and Y. Tohkura. 1997. 'Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production'. *JourmI of the Acoustical Society of America* 101: 2299-23.

Bradlow, A. R., R. Akahane-Yamada, D. B. Pisoni and Y. Tohkura. 1999. 'Training Japanese listeners to identify English /r/ and /l/: V Long-term retention of learning in perception and production'. *Attention, Perception, & Psychophysics* 61: 977-985.

Bradlow, A. R. and T. Bent. 2008. 'Perceptual adaptation to non-native speech'. *Cognition* 106: 707-729.

Brandl, K. K. 1995. 'Strong and weak students' preferences for error feedback options and responses'. *The Modern Language Journal* 79: 194-211.

Brown, R. and C. Hanlon. 1970. 'Derivational complexity and order of acquisition in child speech'. *First language acquisition: the essential readings*: 155-75.

Bruning, R. H., G. J. Schraw and R. R. Ronning. 1999. *Cognitive psychology and instruction*. Prentice Hall.

Buonomano, D. V. and M. M. Merzenich. 1998. 'Cortical plasticity: From synapses to maps'. *Annual review of neuroscience* 21: 149-186.

Burnham, D. K. and R. Brooker. 2002. *7th International Conference on Spoken Language Processing*. Denver, Colorado, USA.

Carcagno, S. and C. Plack. 2011. 'Subcortical plasticity following perceptual learning in a pitch discrimination task'. *Journal of the Association for Research in Otolaryngology* 12: 89-100.

Carroll, S. and M. Swain. 1993. 'Explicit and implicit negative feedback'. *Studies in Second Language Acquisition* 15: 357-386.

Chandrasekaran, B., P. D. Sampath and P. C. M. Wong. 2010. 'Individual variability in cue-weighting and lexical tone learning'. *The Journal of the Acoustical Society of America* 128: 456-465.

Chao, Y. R. 1930. 'A system of tone letters'. *Le Maitre Phonetique* 45: 24-27.

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

Choi, H. and T. Watanabe. 2012. 'Perceptual learning solely induced by feedback'. *Vision research* 61: 77-82.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1980. *Rules and Representations*. New York: Columbia University Press.

Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Chuang, C. K. and S. Hiki. 1972. 'Acoustical features and perceptual cues of the four tones of standard colloquial Chinese'. *The Journal of the Acoustical Society of America* 52: 146.

Chung, S. T. L., D. M. Levi and B. S. Tjan. 2005. 'Learning letter identification in peripheral vision'. *Vision research* 45: 1399-1412.

Cincotta, C. M. and C. A. Seger. 2007. 'Dissociation between striatal regions while learning to categorize via feedback and via observation'. *Journal of cognitive neuroscience* 19: 249-265.

Clark, K. A. 1993. 'The effects of feedback on the problem solving ability of academically at-risk students', ED362137 [Online]. Available at: http://eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED362137&ERICExtSearch_SearchType_0=no&accno=ED362137.

Clarke, C. M. and M. F. Garrett. 2004. 'Rapid adaptation to foreign-accented English'. *Journal of the Acoustical society of America* 116: 3647-3658.

Cooper, A. and Y. Wang. 2012. 'The influence of linguistic and musical experience on cantonese word learning'. *The Journal of the Acoustical Society of America* 131: 4756-4769.

Crist, R. E., M. K. Kapadia, G. Westheimer and C. D. Gilbert. 1997. 'Perceptual learning of spatial localization: Specificity for orientation, position, and context'. *Journal of Neurophysiology* 78: 2889-2894.

Crist, R. E., W. Li and C. D. Gilbert. 2001. 'Learning to see: Experience and attention in primary visual cortex'. *Nature neuroscience* 4: 519-525.

Cutler, A. and H. C. Chen. 1997. 'Lexical tone in cantonese spoken-word processing'. *Attention, Perception, & Psychophysics* 59: 165-179.

Cutler, A., A. Weber and T. Otake. 2006. 'Asymmetric mapping from phonetic to lexical representations in second-language listening'. *Journal of Phonetics* 34: 269-284.

Czerwinski, M., N. Lightfoot and R. M. Shiffrin. 1992. 'Automatization and training in visual search'. *The American journal of psychology*: 271-315.

Das, A., J. G. Franca, R. Gattass, J. H. Kaas, M. A. L. Nicolelis, C. Timo-Iaria, C. D. Vargas, N. M. Weinberger and E. Volchan. 2001. 'The brain decade in debate: Vi. Sensory and motor maps: Dynamics and plasticity'. *Brazilian Journal of Medical and Biological Research* 34: 1497-1508.

Davis, M. H., I. S. Johnsrude, A. Hervais-Adelman, K. Taylor and C. McGettigan. 2005. 'Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences'. *Journal of Experimental Psychology: General* 134: 222-241.

De Bot, K. 1983. 'Visual feedback of intonation i: Effectiveness and induced practice behavior'. *Language and speech* 26: 331-350.

de Heering, A. and D. Maurer. 2011. 'Perceptual learning of inverted faces across different spatial frequency bands'. *Journal of Vision* 11: 590-590.

Dempsey, J. V., M. P. Driscoll and L. K. Swindell. 1993. 'Text-based feedback', In *Interactive instruction and feedback*. 21-54.

Dihoff, R. E., G. M. Brosvic, M. L. Epstein and M. J. Cook. 2010. 'Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback'. *The Psychological Record* 54: 4.

Ding, H., R. Hoffmann and O. Jokisch. 2011. 'An investigation of tone perception and production in german learners of Mandarin'. *Archives of Acoustics* 36: 509-518.

Dosher, B. A. and Z.-L. Lu. 1998. 'Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting'. *Proceedings of the National Academy of Sciences* 95: 13988-13993.

Duanmu, S. 2007. *The phonology of standard Chinese*. Oxford: Oxford University Press.

Dupoux, E. and K. Green. 1997. 'Perceptual adjustment to highly compressed speech: Effects of talker and rate changes'. *Journal of Experimental Psychology: Human Perception and Performance* 23: 914-927.

Ellis, N. 1994. 'Implicit and explicit language learning: An overview'. *Implicit and explicit learning of languages***:**1-31.

Ellis, R. 1993. 'The structural syllabus and second language acquisition'. *Tesol Quarterly* 27**:** 91-113.

Ellis, R., S. Loewen and R. Erlam. 2006. 'Implicit and explicit corrective feedback and the acquisition of l2 grammar'. *Studies in Second Language Acquisition* 28: 339-368.

Fahle, M. and S. Edelman. 1993. 'Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback'. *Vision research* 33: 397-412.

Fahle, M. and T. Poggio. 2002. *Perceptual learning* . The MIT Press.

Fanselow, J. F. 1977. 'The treatment of error in oral work'. *Foreign Language Annals* 10: 583-593.

Feldhusen, J. F. and A. Birt. 1962. 'A study of nine methods of presentation of programmed learning material'. *The Journal of Educational Research* 55: 461-466.

Fiser, J. 2009. 'The other kind of perceptual learning'. *Learning & Perception* 1: 69-87.

Flege, J. E. 1995. 'Two procedures for training a novel second language phonetic contrast'. *Applied Psycholinguistics* 16: 425-442.

Fotos, S. S. 1993. 'Consciousness raising and noticing through focus on form: Grammar task performance versus formal instruction'. *Applied Linguistics* 14**:**385-407.

Fotos, S. S. 1994. 'Integrating grammar instruction and communicative language use through grammar consciousness-raising tasks'. *Tesol Quarterly 28***:** 323-351.

Fotos, S. and R. Ellis. 1991. 'Communicating about grammar: A task-based approach'. *Tesol Quarterly***:** 605-628.

Francis, A. L., V. Ciocca, L. Ma and K. Fenn. 2008. 'Perceptual learning of cantonese lexical tones by tone and non-tone language speakers'. *Journal of Phonetics* 36: 268-294.

Fu, Q. J. and F. G. Zeng. 2000. 'Identification of temporal envelope cues in chinese tone recognition'. *Asia Pacific Journal of Speech Language and Hearing* 5: 45-58.

Fu, Q. J., F. G. Zeng, R. V. Shannon and S. D. Soli. 1998. 'Importance of tonal envelope cues in Chinese speech recognition'. *The Journal of the Acoustical Society of America* 104: 505-510.

Gårding, E., P. Kratochvil, J.-O. Svantesson and J. Zhang. 1986. 'Tone 4 and tone 3 discrimination in modern standard Chinese'. *Language and speech* 29: 281-293.

Gandour, J. 1983. 'Tone perception in far eastern languages'. *Journal of Phonetics* 11: 149-175.

Gandour, J., D. Wong, L. Hsieh, B. Weinzapfel, D. V. Lancker and G. D. Hutchins. 2000. 'A crosslinguistic pet study of tone perception'. *Journal of Cognitive Neuroscience* 12: 207-222.

Gandour, J., Y. Xu, D. Wong, M. Dzemidzic, M. Lowe, X. Li and Y. Tong. 2003. 'Neural correlates of segmental and tonal information in speech perception'. *Human Brain Mapping* 20: 185-200.

Gandour, J. T. and R. A. Harshman. 1978. 'Crosslanguage differences in tone perception: A multidimensional scaling investigation'. *Language and speech* 21: 1-33.

Gaynor, P. 1981. 'The effect of feedback delay on retention of computer-based mathematical material'. *Journal of Computer-Based Instruction* 8: 28-34.

Gibson, J. and E. J. Gibson. 1955. 'Perceptual learning: differentiation or enrichment?'. *Psychological review* 62: 32-41.

Gilman, D. A. 1969. 'Comparison of several feedback methods for correcting errors by computer-assisted instruction'. *Journal of Educational Psychology; Journal of Educational Psychology* 60: 503-508.

Goldsmith, J. 1979. 'The aims of autosegmental phonology'. *Current approaches to phonological theory*: 202-222.

Goldstone, R. L. 1998. 'Perceptual learning'. *Annual review of psychology* 49: 585-612.

Guion, S. G., J. E. Flege, S. H. Liu and G. H. Yeni-Komshian. 2000. 'Age of learning effects on the duration of sentences produced in a second language'. *Applied Psycholinguistics* 21: 205-228.

Hallé, P. A., Y.-C. Chang and C. T. Best. 2004. 'Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners'. *Journal of Phonetics* 32: 395-421.

Hao, Y. C. 2012. 'Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers'. *Journal of Phonetics* 40: 269-279.

Harley, B. 1993. 'Instructional strategies and SLA in early French immersion'. *Studies in Second Language Acquisition* 15**:** 245-259.

Herzog, M. H., K. R. F. Ewald, F. Hermens and M. Fahle. 2006. 'Reverse feedback induces position and orientation specific changes'. *Vision research* 46: 3761-3770.

Herzog, M. H. and M. Fahle. 1997. 'The role of feedback in learning a vernier discrimination task'. *Vision research* 37: 2133-2141.

Herzog, M. H. and M. Fahle. 1999. 'Effects of biased feedback on learning and deciding in a vernier discrimination task'. *Vision research* 39: 4232-4243.

Hodes, C. L. 1985. 'Relative effectiveness of corrective and noncorrective feedback in computer assisted instruction on learning and achievement'. *Journal of Educational Technology Systems* 13: 249-54.

Holland, J. G. 1960. 'Teaching machines: An application of principles from the laboratory'. *Journal of the experimental analysis of behavior* 3: 275-287.

Hsieh, I. L. G. and H. F. O’Neil Jr. 2002. 'Types of feedback in a computer-based collaborative problem-solving group task'. *Computers in Human Behavior* 18: 699-715.

Huang, L. M. 1992. 'Remarks on the phonological structure of Mandarin Chinese'. *Bul. Nat 1 Taiwan Nor. Univ* 37: 363-383.

Huang, T. 2007. ‘Perception of Mandarin tones by Chinese and English speakers’. *16th International Congress of Phonetic Sciences.* Germany.

Hume, E. and K. Johnson. 2003. *Proceedings of the 15th International Congress of Phonetic Sciences*.

Hussain, Z., A. B. Sekuler and P. J. Bennett. 2009. 'Perceptual learning modifies inversion effects for faces and textures'. *Vision research* 49: 2273-2284.

Iverson, P., P. K. Kuhl, R. Akahane-Yamada, E. Diesch, Y. Tohkura, A. Kettermann and C. Siebert. 2003. 'A perceptual interference account of acquisition difficulties for non-native phonemes'. *Cognition* 87: B47-B57.

Jaehnig, W. and M. L. Miller. 2010. 'Feedback types in programmed instruction: A systematic review'. *The Psychological Record* 57: 4.

Jonassen, D. H. and W. H. Hannum. 1987. 'Research-based principles for designing computer software'. *Educational Technology* 27: 7-14.

Jones, I. 1998. 'The effect of computer-generated spoken feedback on kindergarten students'. *Journal of Computing in Childhood Education* 9: 43-56.

Kaan, E., C. Barkley, M. Bao and R. Wayland. 2008. 'Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: An event-related potentials training study'. *BMC neuroscience* 9: 53.

Kaan, E., R. Wayland, M. Bao and C. M. Barkley. 2007. 'Effects of native language and training on lexical tone perception: An event-related potential study'. *Brain research* 1148: 113-122.

Karni, A. and D. Sagi. 1991. 'Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity'. *Proceedings of the National Academy of Sciences* 88: 4966-4970.

Kasper, G. 1985. 'Repair in foreign language teaching'. *Studies in Second Language Acquisition* 7: 200-215.

Kingston, J. 2003. 'Learning foreign vowels'. *Language and speech* 46: 295-348.

Kimura, M. and A. Graybiel. 1995. 'Role of basal ganglia in sensory motor association learning'. *Functions of the Cortico-Basal Ganglia Loop. Springer, Tokyo*: 2-17.

Kiriloff, C. 1969. 'On the auditory perception of tones in Mandarin'. *Phonetica* 20: 63-67.

Klein, D., R. J. Zatorre, B. Milner and V. Zhao. 2001. 'A cross-linguistic pet study of tone perception in Mandarin Chinese and English speakers'. *Neuroimage* 13: 646-653.

Kong, Y. Y. and F. G. Zeng. 2006. 'Temporal and spectral cues in Mandarin tone recognition'. *The Journal of the Acoustical Society of America* 120: 2830-2840.

Kraljic, T. and A. G. Samuel. 2005. 'Perceptual learning for speech: Is there a return to normal?'. *Cognitive Psychology* 51: 141-178.

Krishnan, A., Y. Xu, J. Gandour and P. Cariani. 2005. 'Encoding of pitch in the human brainstem is sensitive to language experience'. *Cognitive Brain Research* 25: 161-168.

Kuhl, P. K., K. A. Williams, F. Lacerda, K. N. Stevens and B. Lindblom. 1992. 'Linguistic experience alters phonetic perception in infants by 6 months of age'. *Science* 255: 606-608.

Kulhavy, R. and W. Stock. 1989. 'Feedback in written instruction: The place of response certitude'. *Educational Psychology Review* 1: 279-308.

Kulhavy, R. W. 1977. 'Feedback in written instruction'. *Review of Educational Research* 47: 211-232.

Kulhavy, R. W. and R. C. Anderson. 1972. 'Delay-retention effect with multiple-choice tests'. *Journal of Educational Psychology; Journal of Educational Psychology* 63: 505.

Kulik, J. A. and C. L. C. Kulik. 1988. 'Timing of feedback and verbal learning'. *Review of Educational Research* 58: 79-97.

LaBerge, D. 1973. 'Attention and the measurement of perceptual learning'. *Memory & Cognition* 1: 268-276.

Larsen-Freeman, D., M. H. Long and Z. Jiang. 1991. *An introduction to second language acquisition research*. Longman: London.

making'. *Topics in Cognitive Science* 2: 226-238.

Leather, J. 1990. 'Perceptual and productive learning of Chinese lexical tone by dutch and English speakers'. *New sounds* 90: 72-97.

Lee, C. Y. 2009. 'Identifying isolated, multispeaker Mandarin tones from brief acoustic input: A perceptual and acoustic study'. *The Journal of the Acoustical Society of America* 125: 1125.

Lee, L. and H. Nusbaum. 1993. 'Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese'. *Attention, Perception, & Psychophysics* 53: 157-165.

Lee, Y. S., D. A. Vakoch and L. H. Wurm. 1996. 'Tone perception in cantonese and Mandarin: A cross-linguistic comparison'. *Journal of Psycholinguistic Research* 25: 527-542.

Li, C. N. and S. A. Thompson. 1977. 'The acquisition of tone in Mandarin-speaking children'. *Journal of Child Language* 4: 185-199.

Liebscher, G. and J. Dailey-O'Cain. 2003. 'Conversational repair as a role-defining mechanism in classroom interaction'. *The Modern Language Journal* 87: 375-390.

Lightbown, P. and N. Spada. 1990. 'Focus on form and corrective feedback in communicative language teaching'. *Studies in Second Language Acquisition* 12: 429-448.

Liu, J., Z.-L. Lu and B. A. Dosher. 2012. 'Mixed training at high and low accuracy levels leads to perceptual learning without feedback'. *Vision research* 61: 15-24.

Liu, S. and A. G. Samuel. 2004. 'Perception of Mandarin lexical tones when f0 information is neutralized'. *Language and speech* 47: 109-138.

Lively, S. E., J. S. Logan and D. B. Pisoni. 1993. 'Training Japanese listeners to identify English /r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories'. *The Journal of the Acoustical Society of America* 94: 1242-1255.

Lively, S. E. and D. B. Pisoni. 1994. 'Training Japanese listeners to identify English /r/and fl/. III. Long-term retention of new phonetic categories'. *Journal of the Acoustical Society of America* 96: 2076-2087.

Loewen, S. 2011. 'Focus on form'. *Handbook of research in second language teaching and learning:* 576-454.

Logan, J. S., S. E. Lively and D. B. Pisoni. 1991. 'Training Japanese listeners to identify English /r/and /l/: A first report'. *Journal of the Acoustical society of America* 89: 874-886.

Long, M. H. 1991. 'Focus on form: A design feature in language teaching methodology'. *Foreign language research in cross-cultural perspective* 2: 39-52.

Long, M. H., S. Inagaki and L. Ortega. 1998. 'The role of implicit negative feedback in sla: Models and recasts in Japanese and Spanish'. *The Modern Language Journal* 82: 357-371.

Lopez, L. and U. Indiana State. 2009. *Effects of delayed and immediate feedback in the computer-based testing environment*. Indiana State University.

Lyster, R. and L. Ranta. 1997. 'Corrective feedback and learner uptake'. *Studies in Second Language Acquisition* 19: 37-66.

Martin, A. a. S. P. 2011. 'Speech perception and phonology', In Marc van Oostendorp, C. J. E., Elizabeth Hume, and Keren Rice (ed.) *The Blackwell companion to phonology*. Wiley-Blackwell.

Mason, B. J. and R. Bruning. 2001. 'Providing feedback in computer-based instruction: What the research tells us'. *Retrieved February* 15: 2007.

Massaro, D. W. and M. M. Cohen. 1985. 'The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese'. *Journal of Chinese Linguistics* 13: 267-290.

McDevitt, E., B. Bays, A. Rokem, M. Silver and S. Mednick. 2012. 'Men need a nap to show perceptual learning of motion direction discrimination, but women do not'. *Journal of Vision* 12: 1132-1132.

McGarr, N. S. 1983. 'The intelligibility of deaf speech to experienced and inexperienced listeners'. *Journal of Speech and Hearing Research* 26: 451-458.

McKee, S. and G. Westheimer. 1978. 'Improvement in vernier acuity with practice'. *Attention, Perception, & Psychophysics* 24: 258-262.

McKendree, J. 1990. 'Effective feedback content for tutoring complex skills'. *Human-Computer Interaction* 5: 381-413.

Merrill, J. 1987. 'Levels of questioning and forms of feedback: Instructional factors in courseware design'. *Journal of Computer-Based Instruction* 14: 18-22.

Miller, N., P. Resnick and R. Zeckhauser. 2005. 'Eliciting informative feedback: The peer-prediction method'. *Management Science*: 1359-1373.

Miyawaki, K., J. J. Jenkins, W. Strange, A. M. Liberman, R. Verbrugge and O. Fujimura. 1975. 'An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English'. *Attention, Perception, & Psychophysics* 18: 331-340.

Molloy, K., D. R. Moore, E. Sohoglu and S. Amitay. 2012. 'Less is more: Latent learning is maximized by shorter training sessions in auditory perceptual learning'. *PLoS ONE* 7: e36929.

Moore, C. B. and A. Jongman. 1997. 'Speaker normalization in the perception of Mandarin Chinese tones'. *The Journal of the Acoustical Society of America* 102: 1864.

Moore, J. W. and W. I. Smith. 1964. 'Role of knowledge of results in programed instruction'. *Psychological Reports* 14: 407-423.

Morrison, G. R., S. M. Ross, M. Gopalakrishnan and J. Casey. 1995. 'The effects of feedback and incentives on achievement in computer-based instruction'. *Contemporary Educational Psychology* 20: 32-50.

Mory, E. H. 1994. 'Adaptive feedback in computer-based instruction: Effects of response certitude on performance, feedback-study time, and efficiency'. *Journal of Educational Computing Research* 11: 263-290.

Nagata, N. and M. V. Swisher. 1995. 'A study of consciousness-raising by computer: The effect of metalinguistic feedback on second language learning'. *Foreign Language Annals* 28: 337-347.

Nomura, E. M., W. T. Maddox, J. V. Filoteo, A. D. Ing, D. R. Gitelman, T. B. Parrish, M. M. Mesulam and P. J. Reber. 2007. 'Neural correlates of rule-based and information-integration visual category learning'. *Cerebral Cortex* 17: 37-43.

Norris, D., J. M. McQueen and A. Cutler. 2000. 'Merging information in speech recognition: Feedback is never necessary'. *Behavioral and Brain Sciences* 23: 299-325.

Norris, D., J. M. McQueen and A. Cutler. 2003. 'Perceptual learning in speech'. *Cognitive Psychology* 47: 204-238.

Opitz, B., N. K. Ferdinand and A. Mecklinger. 2011. 'Timing matters: The impact of immediate and delayed feedback on artificial language learning'. *Frontiers in Human Neuroscience* 5.

Pallier, C., N. Sebastian-Gallés, E. Dupoux, A. Christophe and J. Mehler. 1998. 'Perceptual adjustment to time-compressed speech: A cross-linguistic study'. *Memory & Cognition* 26: 844-851.

Palmeri, T. J. 1997. 'Exemplar similarity and the development of automaticity'. *Journal of experimental psychology: learning, memory, and cognition* 23: 324-354.

Park, O. and S. S. Gittelman. 1992. 'Selective use of animation and feedback in computer-based instruction'. *Educational Technology Research and Development* 40: 27-38.

Peabody, M. and S. Seneff. 2006. 'Towards automatic tone correction in non-native Mandarin'. *Chinese Spoken Language Processing*: 602-613.

Perfors, A., J. B. Tenenbaum and E. Wonnacott. 2010. 'Variability, negative evidence, and the acquisition of verb argument constructions'. *Journal of Child Language* 37: 607-642.

Petrov, A. A., B. A. Dosher and Z. L. Lu. 2005. 'The dynamics of perceptual learning: An incremental reweighting model'. *Psychological Review; Psychological Review* 112: 715-743.

Petrov, A. A., B. A. Dosher and Z. L. Lu. 2006. 'Perceptual learning without feedback in non-stationary contexts: Data and model'. *Vision research* 46: 3177-3197.

Pisoni, D. B., R. N. Aslin, A. J. Perey and B. L. Hennessy. 1982. 'Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants'. *Journal of Experimental Psychology: Human Perception and Performance* 8: 297-314.

Poggio, T., M. Fahle and S. Edelman. 1992. 'Fast perceptual learning in visual hyperacuity'. *Science* 256: 1018-1021.

Poldrack, R. A., J. Clark, E. J. PareÂ-Blagoev, D. Shohamy, J. C. Moyano, C. Myers and M. A. Gluck. 2001. 'Interactive memory systems in the human brain'. *synthesis* 52: 297-314.

Polka, L. and J. F. Werker. 1994. 'Developmental changes in perception of nonnative vowel contrasts'. *Journal of Experimental Psychology: Human Perception and Performance* 20: 421-435.

Pressey, S. L. 1932. 'A third and fourth contribution toward the coming" industrial revolution" in education'. *School & Society* 36: 668-672.

Pressey, S. L. 1950. 'Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction'. *The Journal of Psychology* 29: 417-447.

Pridemore, D. R. and J. D. Klein. 1991. 'Control of feedback in computer-assisted instruction'. *Educational Technology Research and Development* 39: 27-32.

PujolÀ, J.-T. and Agrave. 2001. 'Did call feedback feed back? Researching learners' use of feedback'. *ReCALL* 13: 79-98.

Renner, K. E. 1964. 'Delay of reinforcement: A historical review'. *Psychological bulletin* 61: 341-361.

Repp, B. H. and H.-B. Lin. 1990. 'Integration of segmental and tonal information in speech perception: A cross-linguistic study'. *The Journal of the Acoustical Society of America* 87: S46-S46.

Reynolds, J. N. J. and J. R. Wickens. 2002. 'Dopamine-dependent plasticity of corticostriatal synapses'. *Neural Networks* 15: 507-521.

Robinson, P. 1995. 'Attention, memory, and the “noticing” hypothesis'. *Language learning* 45**:** 283-331.

Robinson, P. 1996. 'Learning Simple and Complex Second Language Rules Under Implicit, Incidental, Rule-Search, and Instructed Conditions'. *Studies in Second Language Acquisition* 18**:** 27-67.

Roberts, F. C. and O. Park. 1984. 'Feedback strategies and cognitive style in computer-based instruction'. *Journal of Instructional Psychology* 11: 63-74.

Roper, W. 1977. 'Feedback in computer assisted instruction'. *Programmed Learning and Educational Technology* 14: 43-49.

Rosa, E. M. and R. P. Leow. 2004. 'Computerized task - based exposure, explicitness, type of feedback, and spanish l2 development'. *The Modern Language Journal* 88: 192-216.

Rosenstock, E. H., J. W. Moore and W. I. Smith. 1965. 'Effects of several schedules of knowledge of results on mathematics achievement'. *Psychological Reports* 17: 535-541.

Rutherford, W. E. 1987. *Second language grammar: Learning and teaching*. Longman.

Rutherford, W. E. and M. Sharwood. 1985. 'Consciousness-raising and universal grammar'. *Applied linguistics* 6: 274-282.

Sachs, R. and C. Polio. 2007. 'Learners' uses of two types of written feedback on a l2 writing revision task'. *Studies in Second Language Acquisition* 29: 67-100.

Sadakata, M. and J. M. McQueen. 2011. ‘The role of variability in non-native perceptual learning of a Japanese geminate-singleton fricative contrast’. *Twelfth Annual Conference of the International Speech Communication Association*. Florence, Italy.

Sagart, L. 1986. 'On the departing tone'. *Journal of Chinese linguistics* 14: 90-113.

Saito, K. and R. Lyster. 2011. 'Effects of form-focused instruction and corrective feedback on l2 pronunciation development of/ɹ/by Japanese learners of English'. *Language Learning*.

Salasoo, A., R. M. Shiffrin and T. C. Feustel. 1985. 'Building permanent memory codes: Codification and repetition effects in word identification'. *Journal of Experimental Psychology: General* 114: 50-77.

Samuel, A. and T. Kraljic. 2009. 'Perceptual learning for speech'. *Attention, Perception, & Psychophysics* 71: 1207-1218.

Schiltz, C., J. M. Bodart, S. Dubois, S. Dejardin, C. Michel, A. Roucoux, M. Crommelinck and G. A. Orban. 1999. 'Neuronal mechanisms of perceptual learning: Changes in human brain activity with training in orientation discrimination'. *Neuroimage* 9: 46-62.

Schmidt, R. W. 1990. 'The role of consciousness in second language learning'. *Applied linguistics* 11: 129-158.

Schwartz, S., P. Maquet and C. Frith. 2002. 'Neural correlates of perceptual learning: A functional mri study of visual texture discrimination'. *Proceedings of the National Academy of Sciences* 99: 17137-17142.

Seger, C. A. 2008. 'How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback'. *Neuroscience & Biobehavioral Reviews* 32: 265-278.

Seger, C. A., R. A. Poldrack, V. Prabhakaran, M. Zhao, G. H. Glover and J. D. E. Gabrieli. 2000. 'Hemispheric asymmetries and individual differences in visual concept learning as measured by functional mri'. *Neuropsychologia* 38: 1316-1324.

Seitz, A. and T. Watanabe. 2005. 'A unified model for perceptual learning'. *Trends in cognitive sciences* 9: 329-334.

Seitz, A. R., J. E. Nanez Sr, S. Holloway, Y. Tsushima and T. Watanabe. 2006. 'Two cases requiring external reinforcement in perceptual learning'. *Journal of Vision* 6.

Shapiro, P. N. and S. Penrod. 1986. 'Meta-analysis of facial identification studies'. *Psychological Bulletin* 100: 139-156.

Shen, X. S. and M. Lin. 1991. 'A perceptual study of Mandarin tones 2 and 3'. *Language and speech* 34: 145-156.

Shibata, K., N. Yamagishi, S. Ishii and M. Kawato. 2009. 'Boosting perceptual learning by fake feedback'. *Vision research* 49: 2574-2585.

Shiu, L. and H. Pashler. 1992. 'Improvement in line orientation discrimination is retinally local but dependent on cognitive set'. *Attention, Perception, & Psychophysics* 52: 582-588.

Shohamy, D., C. E. Myers, S. Grossman, J. Sage, M. A. Gluck and R. A. Poldrack. 2004. 'Cortico striatal contributions to feedback based learning: Converging data from neuroimaging and neuropsychology'. *Brain* 127: 851-859.

Skinner, B. F. 1954. *The science of learning and the art of teaching*. USA: Cambridge Mass.

So, C. K. and C. T. Best. 2010. 'Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences'. *Language and speech* 53: 273-293.

So, C. K. and C. T. Best. 2011. 'Categorizing Mandarin tones into listeners’ native prosodic categories: The role of phonetic properties'. *PSiCL* 47: 133.

Stagray, J. R. and D. Downs. 1993. 'Differential sensitivity for frequency among speakers of a tone and a nontone language'. *Journal of Chinese Linguistics* 21: 143-63.

Sun, K.-C. and T. Huang. 2012. 'A cross-linguistic study of Taiwanese tone perception by Taiwanese and English listeners'. *Journal of East Asian Linguistics* 21: 305-327.

Treisman, A. and H. Schmidt. 1982. 'Illusory conjunctions in the perception of objects'. *Cognitive Psychology* 14: 107-141.

Triskova, H. 2011. 'The structure of the Mandarin syllable: Why, when and how to teach it'. *Archiv orientální* 79: 99-134.

Truscott, J. 1998. 'Noticing in second language acquisition: A critical review'. *Second Language Research* 14**:** 103-135.

Vogels, R., G. Sary, P. Dupont and G. A. Orban. 2002. 'Human brain regions involved in visual categorization'. *Neuroimage* 16: 401-414.

Waldrop, P. B., J. E. Justen and T. M. Adams. 1986. 'A comparison of three types of feedback in a computer-assisted instruction task'. *Educational Technology* 26: 43-45.

Wang, Y., D. M. Behne, A. Jongman and J. A. Sereno. 2004. 'The role of linguistic experience in the hemispheric processing of lexical tone'. *Applied Psycholinguistics* 25: 449-466.

Wang, Y., A. Jongman and J. A. Sereno. 2003. 'Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training'. *The Journal of the Acoustical Society of America* 113: 1033-1043.

Wang, Y., M. M. Spence, A. Jongman and J. A. Sereno. 1999. 'Training american listeners to perceive Mandarin tones'. *The Journal of the Acoustical Society of America* 106: 3649.

Watanabe, T., J. E. Náñez and Y. Sasaki. 2001. 'Perceptual learning without perception'. *Nature* 413: 844-847.

Wentling, T. L. 1973. 'Mastery versus nonmastery instruction with varying test item feedback treatments'. *Journal of Educational Psychology* 65: 50-58.

Werker, J. F. and R. C. Tees. 2002. 'Cross-language speech perception: Evidence for perceptual reorganization during the first year of life'. *Infant Behavior and Development* 25: 121-133.

Whalen, D. H. and Y. Xu. 1992. 'Information for Mandarin tones in the amplitude contour and in brief segments'. *Phonetica* 49: 25-47.

White, L. 1989. *Universal grammar and second language acquisition*. John Benjamins Pub Co.

Whyte, M. M., D. M. Karolick, M. C. Nielsen, G. D. Elder and W. T. Hawley. 1995. 'Cognitive styles and feedback in computer-assisted instruction'. *Journal of Educational Computing Research* 12: 195-203.

Wong, P., R. G. Schwartz and J. J. Jenkins. 2005. 'Perception and production of lexical tones by 3-year-old, Mandarin-speaking children'. *Journal of Speech, Language, and Hearing Research* 48: 1065-1079.

Wong, P. C. M. 2002. 'Hemispheric specialization of linguistic pitch patterns'. *Brain Research Bulletin* 59: 83-95.

Wong, P. C. M., E. Skoe, N. M. Russo, T. Dees and N. Kraus. 2007. 'Musical experience shapes human brainstem encoding of linguistic pitch patterns'. *Nature neuroscience* 10: 420-422.

Wright, B. A. and Y. Zhang. 2009. 'A review of the generalization of auditory learning'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364: 301-311.

Xu, C. X. and Y. Xu. 2003. 'Effects of consonant aspiration on Mandarin tones'. *Journal of the International Phonetic Association* 33: 165-181.

Xu, Y. 1997. 'Contextual tonal variations in Mandarin'. *Journal of Phonetics* 25: 61-83.

Xu, Y., J. T. Gandour and A. L. Francis. 2006. 'Effects of language experience and stimulus complexity on the categorical perception of pitch direction'. *The Journal of the Acoustical Society of America* 120: 1063-1074.

Yang, W. J., J. C. Lee, Y. C. Chang and H. C. Wang. 1988. 'Hidden markov model for Mandarin lexical tone recognition'. *Acoustics, Speech and Signal Processing, IEEE Transactions on* 36: 988-992.

Yip, M. J. W. 2002. *Tone*. Cambridge: Cambridge University Press.

Zalewski, J. P. 1993. 'Number/person errors in an information-processing perspective: implications for form-focused instruction'. *Tesol Quarterly* 27: 691-703.

Zeng, B. 2008. *Tone Perception and Migration by English speakers* . Unpublished PhD dissertation, Bristol University.

Zeng, B. and S. Mattys. 2011. 'Separability of tones and rhymes in Chinese speech perception: evidence from perceptual migration'. *17th International Congress of Phonetic Sciences*, Hong Kong, China: ICPhS.

Zhang, T., L.-Q. Xiao, S. A. Klein, D. M. Levi and C. Yu. 2010. 'Decoupling location specificity from perceptual learning of orientation discrimination'. *Vision research* 50: 368-374.

Zheng, H.-Y., J. W. Minett, G. Peng and W. S. Y. Wang. 2011. 'The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study'. *Language and Cognitive Processes* 27: 184-209.

Appendices

Appendix 1: Pilot Stimuli List (Order fixed)

No. 1	mo55	mi55
No. 2	gang55	gong214
No. 3	mu214	fa214
No. 4	mo214	ga51
No. 5	ge35	gu214
No. 6	du55	ta55
No. 7	fei51	fen214
No. 8	bo55	ning214
No. 9	po35	pin214
No. 10	du35	da35
No. 11	dou51	hao51
No. 12	nin35	da35
No. 13	du51	di51
No. 14	nan55	mao55
No. 15	hou35	di214
No. 16	du55	dou55
No. 17	ba35	hu214
No. 18	ha214	hou35
No. 19	kou214	hai35
No. 20	ken214	kai214
No. 21	ke55	ka55
No. 22	bi55	tou55
No. 23	heng35	han214
No. 24	pi214	po51
No. 25	ka214	mai214
No. 26	nong35	nao35
No. 27	mo214	ge35
No. 28	ku51	kan51
No. 29	hao214	hu214
No. 30	mo35	ti35
No. 31	mu214	mo35
No. 32	bao51	pa51
No. 33	de55	ti214
No. 34	fo35	fang35
No. 35	fan35	ning214
No. 36	dong51	pao214
No. 37	fang35	ping35
No. 38	bing214	hu35

No. 39	ge214	gen35
No. 40	pin214	pan55
No. 41	hu214	he55
No. 42	tao51	ting51
No. 43	ni35	na214
No. 44	fang55	deng55
No. 45	mai35	mei35
No. 46	he35	ba35
No. 47	pa51	ku51
No. 48	nei214	nin35
No. 49	di214	ge214
No. 50	te51	ti214
No. 51	ba214	bi214
No. 52	gen35	ku214
No. 53	ni214	nan55
No. 54	tu214	pin214
No. 55	hang55	heng55
No. 56	kan214	bei51
No. 57	tu214	na35
No. 58	ba214	gu55
No. 59	fan214	fou214
No. 60	mao55	kai214
No. 61	tan35	ni35
No. 62	na51	fu214
No. 63	bao35	bing214
No. 64	dei214	dan51
No. 65	mei214	bang214
No. 66	pi51	pai51
No. 67	nan51	gen51
No. 68	ti55	ting55
No. 69	pu214	geng51
No. 70	kou51	ke214
No. 71	gei214	dao35
No. 72	nen214	ni214
No. 73	tong214	pin55
No. 74	mi35	meng214
No. 75	hou55	bu55
No. 76	men35	mo35

No. 77	bo55	bi214
No. 78	men51	fu51

No. 79	mo51	ma51
No. 80	fen214	pi35

Appendix 2: Post-test Stimuli List (No fixed order)

na55	ni55
tu35	ti214
bo55	bi214
fan214	fou214
ge55	hu55
na35	min35
mo51	ma51
dong214	mou214
bai214	bo35
hu35	he35
mo51	nei51
gen35	ku214
kang35	kou214
ke214	nin35
te51	ti214
bo55	ning214
tong214	tai35
kong51	tang51
dong51	pao214
bai55	ben55
tong214	pin55
tan35	tou35
fan35	ning214
dao35	da214
mo214	ge35
hao214	hu214
de55	ti214
fang55	deng55
mao214	meng55
ba35	hu214
te51	gu51
tao51	ting51
ting55	ta214
pao214	ke214
feng214	hong51
hou55	bu55
bei51	bi51
gai214	gong51
pi35	fu35

fan55	fu55
gang55	gong214
mo55	mi55
bi55	tou55
fen214	fei35
bi35	beng35
mo35	ti35
ku51	kan51
ke35	po214
heng35	han214
men35	mo35
men51	mang214
ka214	mai214
dao214	de35
beng35	tang214
gao55	gong55
fang35	ping35
nan55	mao55
mao55	kai214
du55	dou55
ha214	hei51
nei214	de55
mu214	fa214
pi214	po214
mo214	ga51
ba55	po55
ke214	mi55
po35	pin214
ni214	nan55
kou214	hai35
pa51	ku51
gu214	gei214
bing214	hu35
na51	fu214
ka214	ke55
ge51	gu51
men51	fu51
nin35	da35
ga35	ge35

ni35	min214
ni35	na214
he51	mi214
he35	ba35
ke55	ka55
ge35	mo214
ha214	hou35
pou55	ting55
du35	da35
dou51	hao51
pi51	pai51
ba214	pu214
hang55	heng55
pu214	geng51
fa55	fu214
mai35	mei35
fen214	pi35
nu214	ni51
ge55	pen55
ding214	dong214
di214	ge214
hou51	hen51
ba214	gu55
tan35	ni35
bao35	bing214
han214	fei55
fo35	fang35
fei35	fu214
ba214	bi214
bao51	pa51
kou51	ke214
po214	pi35
pin214	ming35
nen214	ni214
hai35	bao35
bo51	ba214
dao35	tong214
bai214	hu214
ge35	gu214
di214	fa35
fen55	fou214

tan51	kai214
tu214	na35
fan214	hen214
fa55	hu214
ge214	gen35
pin214	pan55
da51	ni51
pi214	po51
ge51	ben214
pei55	po55
bo214	tou35
hu35	beng35
kan214	bei51
ping35	pa35
nan51	gen51
kang214	ke35
ken214	kai214
pou55	gen214
fei55	di55
ta214	tu214
mu214	mo35
hou35	di214
hu214	he55
mou35	hen35
nong35	nao35
ti55	ting55
ti214	du51
dei214	dan51
tu214	pin214
nan51	nong51
gei214	dao35
mo214	min214
mi35	meng214
he51	fan51
fen51	fu51
nei214	nin35
du55	ta55
fo35	peng214
du51	di51
fei51	fen214
mei214	bang214

Appendix 3: Pre-test Stimuli List (No.1 to No.12 are used as fillers in the post-test)

No. 1	bie1	niu3
No. 2	dui4	puo3
No. 3	tui3	pie1
No. 4	dun1	tuo3
No. 5	liu3	hui4
No. 6	guo2	kun3
No. 7	zhan3	shen2
No. 8	lan2	chun3
No. 9	chan3	lin2
No. 10	sha2	lu3
No. 11	zhuo1	zhi3
No. 12	le4	li3
No. 13	shao3	sheng1
No. 14	sheng1	sha3
No. 15	gun3	gui4
No. 16	tui2	tie3
No. 17	lie3	lun2
No. 18	chang2	chou3
No. 19	zhong3	zhai2
No. 20	lao2	la3
No. 21	she1	shu1
No. 22	la2	lin2
No. 23	chong3	chou3
No. 24	zhu4	zhe4
No. 25	sheng4	shang4
No. 26	chang1	cheng1
No. 27	tie4	gui4
No. 28	puo3	kun3
No. 29	huo1	bie1
No. 30	lan2	zhu2
No. 31	cha1	shu1
No. 32	chan3	luo3
No. 33	miu4	mie4
No. 34	hun2	huo2
No. 35	lin1	lun1
No. 36	chan2	chou2
No. 37	huo3	hui3
No. 38	tie4	tuo4
No. 39	shu4	she4
No. 40	lu1	luo1

Appendix 4: Consent Form

What is this experiment about? What will I be asked to do?

Study. "The effect of immediate feedback on the perception of Mandarin lexical tones by English speakers" conducted by School of Education Communication and Language Sciences and School of Literature, Language and Linguistics, Newcastle University

Purpose. This study examines how native English speakers perceive Mandarin lexical tones.

Your participation. You will sit in front a computer. You will see a "+" on the screen and then hear a pair of syllables consecutively from the headphone. Your task is to decide whether the tones (i.e. the music-like pitches) of the two syllables are the same or not, regardless of the consonants and vowels. If same, please press the right SHIFT key; if different, please press the "N". Please put your hands on the keyboard all the time and try to make your judgment as quickly and accurately as possible during the test.

Before the testing, we will first ask you to tell us your handedness, age, gender, language background and hearing condition. The testing phase will last about 30 minutes. We won't take any visual or auditory recordings.

You will receive £3 for your participation. Your decision to participate in this study is entirely voluntary and you may decide at any time to withdraw from the study. You do not need to explain your decision.

Your responses and any personal data will remain confidential, and your data will be anonymized. Only researchers associated with the project will have access to the data. The results of the study may be presented at scholarly conferences and/or published in professional journals. The data will be securely stored.

Experiment Duration. The duration of this experiment will be around 30 minutes.

Please ask the experimenter for further clarification if there is anything that you do not understand.

Consent

Please sign below to indicate that you have read and understood the nature and purpose of the study and that you agree to participate.

Title of the Study: The effect of immediate feedback on the perception of Mandarin lexical tones by English speakers

Name: ___________

Age: ______

Sex: □Female □Male

Handedness: □Right-handed □Left-handed

Language background

First Language: English □Yes □No

Knowledge of other languages: ____________________________________

Music experience: □above 6 years □ under 6 years □ no experience in music

Participant signature: __________________________________ date:__________

Pre-test file number: _______

Post-test file name: _______

Experimenter signature: __________________________ date:___________

Contact information:

Intentionally cancelled in this dissertation.