Chapter 2
Literature Review

Chapter 1 introduced three kinds of tonal features and explained that the current research will focus on the connection between acoustic and psychological features. We will do so by transferring F_0 to normalized pitch values, by using contextual cues to explore tone perception, and then comparing NNS' perception and production with NS' perception. This chapter explains the justification for the current study from the perspective of non-native speakers' acquisition of tones.

Section 2.1 focuses on studies relevant to the common challenges L2 learners experience in acquiring suprasegmental features. The section also discuss why learning Mandarin tones is so difficult for non-native speakers whose first language is non-tonal.

Section 2.2 reviews the research on tonal acquisition of Mandarin as a second language (L2), and justifies how to use acoustic and contextual cues to explore non-native speakers' perception and production. This section concludes by identifying the research questions which consider the aspects of perception and production that pose the greatest barriers to second language learners of Chinese.

Section 2.2 mainly discusses acquisition order, the relationship between NNS' perception and production vis-à-vis, tone categories, and phonetic details. The section then considers categorical perception, whether acoustic cues can be easily mapped onto perceptual features and whether contextual cues should be added in the current study.

Section 2.3 uses the literature review to analyze the most significant barriers to non-native speakers' acquisition of Mandarin tones. Section 2.4 surveys problems with previous studies relevant to this research. Section 2.5 delineates this work's research questions.

B. Yang, *Perception and Production of Mandarin Tones by Native Speakers and L2 Learners*, DOI 10.1007/978-3-662-44645-4_2

2.1 Why is it Difficult for Non-native Speakers to Learn Mandarin Tones?

The age of a non-native language learner and the learner's non-tonal native language are the two most salient factors that make it difficult for a non-native speaker to acquire the tones associated with Mandarin. Most scholarly discussions of Mandarin tonal acquisition presume an adult L2, even though tonal acquisition is significantly easier and more instinctual for children.

Children have less difficulty learning tones. The accepted reason for this is that adults are farther away from the critical period for acquiring native-like pronunciation. Burnham (2000) argues that there is a resurgence of tone discrimination ability in adulthood; but the idea that L2 adults can fully acquire tones is still controversial. In a 2008 study, Burnham and Mattock concluded that "the resurgence is due to adults" ability to disregard the linguistic constraints that bind children, so that they can perceive tones in all (or most) of their perceptually-salient glory" (Burnham and Mattock 2008: 267).

Burnham's hypothesis contradicts Cutler's findings that prosodic features are learned early and only once (Cutler 1994; Cutler et al. 1992). However, both studies indicate that children's acquisition of prosodic features is more instinctive and less arduous than adults'.

Tonal acquisition is also more difficult for those second language learners whose native language is non-tonal. There are three main language groups: (a) tonal languages, such as Chinese, Thai, and Yoruba, which comprise about half of the world's languages; (b) stress languages, such as English; and (c) mora-timed languages, such as Japanese. Certain languages, such as Swedish and Serbo-Croatian, also have pitch accent or accentual systems (Yip 2002). Accentual languages have only a small number of tonal contrasts compared to tonal languages. While, as Yip (2002: 4) remarks, "There is no absolute division between accent languages and tone languages," there is a vast and real divergence between the two. Tonal languages use tones to establish every syllable's meaning (see Table 1.1). This is in direct contrast to stress languages, in which stress impacts only a relatively tiny minority of lexical meanings, e.g., the English word "record" is a noun when the first syllable is stressed, and a verb if the second syllable is stressed.

Although stress languages also have pitch, pitch contrasts do not mark semantic differences at the syllabic level; rather, they apply it at paralinguistic or sentential levels. Most research on stress languages indicates that declination is the default intonation pattern (Liberman 1975; Sorensen and Cooper 1980; Pierrehumbert 1980, 1981; Cruttenden 1997), in which "the pitch range narrows and drifts downwards over the course of a major phrase." (Pierrehumbert 1981: 987).

English intonation patterns do not cause much semantic differences. Pierrehumbert (1980) illustrated five intonation patterns for the word "Anna" in the first chapter. The first four patterns have a high pitch for the first syllable and a low pitch for the second, indicating a response or statement. These patterns have the potential to exhibit pitch pattern variation. For example, the pitch at the beginning

of the first syllable may rise a little bit to indicate the speaker is incredulous; the pitch rising at the end again indicates that the answer is incomplete. The pitch contour can have a pitch peak at the first half of the intonation and "stop far short of the bottom of the speakers' range" (Pierrehumbert 1980: 7), which shows that somebody is calling out for Anna. The fifth pattern exhibits a low pitch for the first syllable and a high pitch for the second, the typical question cadence. As this discussion makes clear, the rising, falling, and level contours are composed of the high and low pitch, and different intonation patterns in English indicate paralinguistic features, different emotions, sentence types, and so on.

Pitch contours are also related to a tone's position in a sentence. Some researchers (Broselow et al. 1987) investigated pitch contours' influence on learners of Chinese. Their study indicated that the boundary tone does effect tone acquisition. The falling intonation that indicates a statement positively influences T4 perception at the end of a phrase; yet, not on T4 perception in the middle of a phrase.

The studies mentioned above suggest there are some basic pitch contours, such as rising, falling, and level, for English intonations on a phonetic level. We would expect, then, that tone contours would work with other factors, such as a tone's position of a phrase and emotion, to form intonations. While this is the case in English, it is not so in Chinese where pitch contours are related to syllables' semantic meaning.

It is established that the linguistic function of pitch is very different in Mandarin than it is in English, posing an immense barrier for L2 learners of Chinese who must learn how to match pitch with meaning at the syllabic level. In other words, these learners must acquire a new linguistic category that does not exist in their native language. This is what lies at the heart of what makes it so daunting a task for L2 learners of Chinese to enter within the world of this new language.

This does not mean that L2 learners whose first language is tonal acquire Chinese tones more swiftly or completely, than those with a non-tonal native language. On the contrary, some research (Hao 2012) showed that L2 learners with a tonal native language perform no better than those with a non-tonal native language. Just as the transition from non-tonal to tonal language is fraught with struggle, so the transition from one tonal language to another poses its own challenges.

It is hard for tonal native L2 learners to produce tones accurately because the tone categories of their first language may hinder them in acquiring Mandarin tone categories. By the same token, L2 learners whose non-tonal native languages—once they are able to establish the tone categories—actually exhibit more accurate tone production than tonal native learners. In either circumstance, it is not easy for these learners to reach the proficiency with tone categories similar to a native speaker's.

In summary, the difficulty in learning Mandarin tones lies in the factors inherent to individual learners (such as age of L2 learners), and the nature of students' native languages.

2.2 Survey of Research on Tone Perception and Production of Mandarin as L2

Questions about how second language learners acquire Mandarin tones have attracted a large amount of interest on the part of scholars and teachers. In this section, the major strands of that research are reviewed.

2.2.1 Introduction

Over the past several decades, a great many studies from a wide range of perspectives have been conducted on Mandarin tone acquisition by L2 learners of Chinese whose L1 is not a tonal language. Some of this research has focused on the order of the tones which these learners acquire (Chen 1997; Elliot 1991; Kiriloff 1969; Leather 1990; Miracle 1989; Shen 1989; Sun 1997). Others have explored how learners transfer suprasegmental features of their native language to L2 Mandarin (Broselow et al. 1987; Chiang 1979).

The interests of phonetics and phonology have also been represented in research on this topic (Hao 2012; Leather 1990; Read et al. 1986; Shen 1989), as have those of psycholinguistics (Halle et al. 2004; Leather 1983, 1987; Schwanhauber et al. 2003; Stagray and Downs 1993; Yang 2012).

New technology and phonetic experiments have allowed researchers to carry out experiments on neurophysiological processing (Sereno and Wang 2008; Soares 1982; Sussman et al. 1982; Wang et al. 2004; Wuillemin and Richardson 1994).

Studies that are relevant to the current research have been selected for review below.

2.2.2 Order of Acquisition of Tones

The studies summarized in Table 2.1 present the order of tonal acquisition in different studies (Chen 1997; Elliot 1991; Hao 2012; Kiriloff 1969; Leather 1990; Miracle 1989; Shen 1989; Sun 1997). Even though their results and conclusions vary, they do appear to agree on one underlying outcome: The orders of perception and production are not exactly the same.

These studies raise an important question: What is the most demanding aspect of tones to learn? Most studies agreed that perceptual challenges are different from production challenges. The majority of research indicated that T2 is the most difficult tone for learners to perceive. There is no such agreement regarding production, where results were more varied, although many found both T2 and T3 are difficult to produce. We clearly need to investigate the relationship between perception and production to pinpoint NNS' greatest difficulties in acquiring Mandarin tones.

Table 2.1 Relative difficulty of the toes reported in previous studies

Study	Mode	Order
Kiriloff (1969)	Perception	4 < 1 < 3 < 2
Elliot (1991)	Perception	4 < 3 < 1 = 2
	Perception: self	4 < 1 < 3 < 2
Sun (1997)	Perception: TIDT (stimulus)	4 < 1 < 3 < 2
	Perception: TIDT (response)	1 < 4 ≤ 3 < 2
Hao (2012)	Perception	4 < 1 < 3 ≤ 2
Miracle (1989)	Production	1 < 4 < 3 < 2
Shen (1989)	Production	2 < 3 < 1 < 4
Leather (1990)	Production	1 < 4 < 2 = 3
Elliot (1991)	Production	1 < 4 < 2 < 3
Chen (1997)	Perception/production	1 < 4 < 2 < 3
Sun (1997)	Production: REPT&RDGT (stimulus)	1 < 4 < 3 < 2
	Production: REPT (response)	3 ≤ 4 < 1 ≤ 2
	Production: RDGT (response)	4 < 1 ≤ 3 < 2
	Production: TRAT (stimulus)	1 < 2 ≤ 3 < 4
Hao (2012)	Production	1 < 4 < 3 ≤ 2

Note TIDT means "tone identification task"; REPT means "repetition task"; RDGT means "reading aloud task"; and TRAT means "oral translation task"
Source Adapted based on Sun (1997: 196)

2.2.3 The Relationship Between Perception and Production

Several studies investigated the connection between perception and production (Chen 1997; Elliot 1991; Leather 1990; Yang 2012).

Leather (1990) reported that error patterns in production correlate with those in perceptual tests. He also asserted that the perception and production of tones are interrelated.

Elliot (1991) found a more moderate correlation between perception and production. According to his research, the relationship between the two is not close, especially for the third and fourth tones. Elliot assumed that learners use different prototypes to categorize the fourth tone. Sun (1997) criticized Elliot's method of selecting his 33 study participants, noting that about half of the participants are native speakers of tone languages including Vietnamese or other Chinese dialects. She also points out that Elliot does not discuss whether tones in learners' native language would influence their L2 acquisition of tones.

Chen (1997) looked specifically into the relationship between learners' tone perception errors and their tone production errors. This study, like Elliot's (1991) study, has a subject selection problem. Among the six subjects, three had lived in Hong Kong for more than ten months, which suggested Cantonese tones may have effected their acquisition of Mandarin tones. One subject, born in California to a Chinese-American family, had visited Taiwan once for three months, but did not

speak much Chinese. However, a heritage learner who cannot speak Chinese may still comprehend quite a bit of Chinese. Many American-born Chinese can understand Chinese although they cannot speak it, and such subjects probably exhibit high proficiency in tone perception.

Yang (2012) studied American learners of Chinese who had studied Mandarin for one or two years, all of them are true learners, whose native language is English. The results demonstrated a close relationship between tone perception and production, and the subjects' performance in tone production was much better than their tone perception.

Suprasegmental and segmental categories impact tone perception and production in different ways. In Yang's study (2012), perception is influenced by tone categories and syllable level categories (initials and finals), but tone production errors are independent of tone categories. T2 is significantly unlike other tones in terms of perception, with markedly fewer correct responses than for other tones.

These findings lead us to assert T2 is the hardest tone to perceive; yet, when it comes to production, learners do not show any obvious pattern of difficulty across the tones. The study's results illustrated that tones are perceived with relative systematicity, whereas tone production is not. For example, learners make more tone perception errors in T2, while there is no clear tendency to similar errors when learners produce tones.

Yang's study also showed that the variance in tone perception mean between initial 1 (labials) and initial 2 (alveolars) is significant at the 0.05 level, between initial 2 (alveolars) and initial 4 (retroflexes) is significant at the 0.05 level, and between final 1 (mono-vowel finals) and final 2 (multivowel finals) is significant at the 0.05 level. These findings combining with the means make the case that it is hard for NNS to perceive tones carried by syllables which have alveolars and by syllables which have multivowel finals. It is not significantly different between any other finals or initials when learners produce tones. Based upon these results, Yang (2012) concluded that tones are perceived at the phonological level, while they are produced at the phonetic level.

The difference between perception and production may be a function of scoring methods. NNS' production is perceived and scored by NS judges, while NNS' perceptions—based on a still-emergent model—must fall into one of only four standard categories that diverge from those used by NS. The perceptual difficulties NNS encounter may be accounted for by the fact that NNS have not assimilated the NS' tone categories.

In another experiment (Yang and Ankernmann 2007), participants were asked to listen to a target tone they had perceived incorrectly in the first tone identification task. Listeners compared the tone read by a native speaker with the same syllable read in all four tones, and then wrote down the tone type for the target word. For example, if a participant perceived the word "ping2" incorrectly, then s/he was asked to listen to "ping2," followed by "ping1," "ping2," "ping3," "ping4," and then was to write down the tone type of the target syllable. The purpose of this task was to investigate whether external tonal referents (the native tonal system) can facilitate learners' perception. The accuracy of tone perception did improve, suggesting

learners make perceptual errors because they lack a fully developed internal representation of the tonal system that matches the native speaker's system.

Based on the results of the previous studies, it seems important for learners to develop the correct tone categories. The results described above indicate that it is challenging for learners to acquire the L2 tonal system. If these categories are incorrect, learners will make errors during tone perception.

In some situations, production is based on perception (Flege 1986). It is reasonable to suppose an adult can better produce a sound if s/he can also perceive its tones. In this case, sufficient inputs and intake are helpful. This leads us to ask whether, if a learner's perception is improved, it will benefit their production as well.

Some perception training indicated that learners can improve their tone category perception by listening to tonal variations (Leather 1990; Wang et al. 1999, 2003). Leather (1990) carried out a parallel perception and production training study that focused on Dutch students of Chinese who were acquiring the four tones. Wang et al. (1999) worked on perception training with American learners of Chinese. Both studies showed substantial improvements. In other words, well-designed training can improve listeners' perceptual ability of distinguishing tonal contrasts. They do sound a cautious note, as the improvement in tone perception varies from about 5−21 % across studies. Nevertheless, it was encouraging to note that improvement could be retained over 3 months after training (Wang et al. 1999).

The absence of a substantial body of L2 production training studies has lead a number of scholars to investigate whether perceptual training can be of assistance to these learners. Wang et al.'s (2003) perceptual training study demonstrated that perceptual training can improve production (increases of 18 %) even without special production training. In this study, around 80 NS evaluated NNS production and a statistical method was used to calculate the production scores.

It is clear that perceptual training does have an effect on both the perception and production of tones. In other words, production could be improved by developing and strengthening perception.

The perception and production studies mentioned above agreed that learners' production is superior to their perception. However, we should not interpret this to mean that perception cannot benefit production. A well-designed perception training program could advance learners' ability to perceive tone categories while also benefitting their production. The studies about the tonal training revealed the importance of NNS' perception. The studies about the relationship between perception and production also showed that NNS have more perception errors and lack tone categories that are similar to NS'. Therefore, both studies point to the necessity of investigating NS and NNS' perceptual categories.

2.2.4 Categorical Perception of Tones

It is a matter of debate whether or not tones are perceived categorically. Researchers have conducted numerous perceptual experiments (Abramson 1979; Blicher et al.

1990; Chan et al. 1975; Chang and Halle 2000; Fox and Unkefer 1985; Francis et al. 2003; Gandour 1983; Gandour and Harshman 1978; Halle et al. 2004; Leather 1987; Peng et al. 2010; Shen and Lin 1991; Stagray and Downs 1993; Wang 1967; Wu 2005; Xia et al. 2010; Xu et al. 2006; Zheng et al. 2010) to explore the Mandarin tone categories.

An early study of Thai speakers' perception of tones (Abramson 1979) demonstrated there was no categorical effect for tones. Other studies disagreed (Chan et al. 1975; Fox and Unkefer 1985), finding that NS of Mandarin exhibited categorical perception effects, in contrast to the continuous perception effects of NS of American English.

Studies by Leather (1987), Stagray and Downs (1993), and Wang (1967) explored tone categories in beginning Chinese learners who did not have any prior knowledge of tones. A number of studies approached the topic by concentrating on a particular tone category, such as level tone (Stagray and Downs 1993), or a contrastive pair of tones, such as T2 versus T3 (Blicher et al. 1990; Shen and Lin 1991). Others focused on cross-language tone categories (Gandour 1983; Gandour and Harshman 1978), investigating perceptual representations across tonal languages, such as Mandarin and Thai, and non-tonal languages, such as English.

In this section, we review the studies in this field. We compare definitions of categorical perception for tones and propose the term used in the current research; analyze the phonetic features researchers used to conduct the experiments related to this topic; and compare the two major findings of this research.

2.2.4.1 Definition of Categorical Perception of Tones

Categorical perception occurs when a physical continuum is perceived discontinuously, with sharp identification boundaries between categories, and poor or absent discrimination within the categories (Liberman et al. 1957; Studdert-Kennedy et al. 1970). Consonants are perceived categorically (Liberman et al. 1957; Rosen and Howell 1987), while vowels are perceived continuously (Eimas 1963; Fry et al. 1962).

"A stringent definition of categorical perception requires an optimal fit between observed discrimination performance and performance predicted from identification, reflecting the strong claim that discrimination between two sounds is uniquely determined by the probability that they are labeled differently (Schouten and Van Hessen 1992). Clearly, in many cases, such an ideal fit is not obtained (Pisoni and Lazarus 1975; Pisoni and Tash 1974; Wood 1976)" (Halle et al. 2004: 415).

Most research in categorical perception of tones does not adhere strictly to the original definition of categorical perception, preferring a more liberal interpretation of the phenomenon. There are studies that consider how to categorize the phonetic variations upon which phonetic differences are based, and whether phonetic categories match the linguistic categories related to the linguistic system of a specific language (Wang 1967; Gandour 1983; Stagray and Downs 1993).

Other studies scrutinized category boundaries, and were closer to experiments in the categorical perception of segments (Abramson 1979; Halle et al. 2004; Wu 2005). "'Categorical boundary effects' cannot be explained entirely by psychophysical responses. In other words, the increased sensitivity to differences between the members of the pairs that straddle category boundaries at least partially reflects phonetic coding into linguistic categories that may, …, be language-specific" (Wood 1976; Halle et al. 2004: 415). Categorical perception is related to linguistic category, phonetic category, and psychophysical responses, and it connects physical properties to real-world segmental or suprasegmental categories.

The current research uses acoustic features to explore the boundaries of phonetic categories that match linguistic categories, and to further define the pitch ranges of each category according to the boundaries. Phonetic categories are relevant to phonetic properties, such as the low falling tone, while linguistic categories are related to the linguistic system of a specific language. In this case, it is Mandarin, which uses tones to distinguish meanings.

Section 2.2.3 noted that studies regarding the relationship between perception and production demonstrated that the perceptual categories of tones are important. On the other hand, the results also showed that tone perception is parallel to production although production is better than perception (Yang 2012). However, perception can benefit production during perception training (Wang et al. 2003).

In the current study, we also observe tone categories during production to see how tones produced by NNS align with NS' perceptual categories. For this reason, we employ the term "perceptual category" rather than the traditional "categorical perception" to investigate NS' and NNS' tone perception and production.

2.2.4.2 Acoustic Features Explored in Studies About Tone Perception

Early research about tone perception raised the controversial issue of whether NNS are sensitive to small phonetic differences while NS are not. Wang (1967) and Stagray and Downs (1993) found that NNS have different categories than NS, and that NNS are sensitive to some small acoustic differences which NS ignore. Leather (1987) argued against this view, suggesting that some NNS may have similar categories of NS' and that NNS are less sensitive to small pitch differences. However, all the studies on this topic do agree errors come from interrelated contrastive pairs related to a specific phonetic feature.

Divergent research methodologies make it a challenge to aggregate tone category study results. Stagray and Downs (1993) used pitch height to explore the level tone. Blicher et al. (1990) used duration to compare T2 with T3. Shen and Lin (1991) found that the turning point of the pitch contour is the salient acoustic feature in distinguishing T2 and T3. Gandour and Harshman (1978) and Gandour (1983) used height and direction, similar to register and contour, to characterize a multidimensional model for the perceptual representation of tones. In these studies, subjects were observed to make similar judgments among the standard tones. Gandour and Harshman (1978) and Gandour (1983) used the dimensions they had identified, to

verify that the resultant perceptual space which has a direct relationship to acoustic features which is sufficient to capture the whole system of Mandarin tones.

In some perception studies, scholars chose different acoustic features to synthesize stimuli. In Halle et al.'s (2004) experiments, "intermediate contours were obtained via interpolation between end points." (Halle et al. 2004: 401). Wu based her 2005 study on Shen and Lin's (1991) results, which assumed that the turning points of total contour for T2 or T3 provided the most important cue to distinguishing these two tones. Wu's experiments, therefore, concentrated on changing the turning point to a different time spot.

T3's tonal features pose another problem to researchers studying Mandarin tones acoustically. Most acoustic studies, for example, Zhu's study (2012), and phonological studies, such as Chen's study (2000), specify that T3 is a low tone in most contexts. However, T3 is perceptually contrastive with T2, a rising tone; and also contrasts with T4, a high falling tone. This is the evidence that there must be several cues, including both acoustic and contextual factors, which influence the perception of tone categories. The issue is that, by and large, researchers of different tone perception only employ selected phonetic properties used to contrast T2 with T3, while ignoring other cues. This problem needs to be further explored.

In addition to T2 and T3, we also need to use the same acoustic parameter(s) to explore the tonal system during perception and production.

2.2.4.3 Comparing NNS Perception with NS Perception

At the core of these investigations is the question of whether, how, and why NNS have different tone categories from NS. Recently research has focused on categorical perception from two contrasting methodological foundations.

Halle et al. (2004) showed that NNS whose L1 is not tonal do not process tones linguistically and that their perceptual judgments are based on a general psychophysical function. This research produced evidence that NS perceive tones in a quasi-categorical way, whereas NNS use a different strategy. Halle's results were similar to Burnham and Jones (2002).

In the 2004 study, Halle et al. used the elicitation phrase *yi ge X zi* (one + classifier + X + Chinese character). Here, X works as a single syllable, and is not linked with either the previous or following syllables to compose a meaningful, multisyllable word or phrase. The problem with this approach is that it is not a natural context, thus making it impossible to judge tones based on meanings. Further, there is just one carrier sentence, providing a single context: The end point of the previous tone is low and the start point of the following tone is high. These results, then, still proceed from a context-free environment.

Wu's (2005) study examined two contrastive pairs: T1 and T4; and T2 and T3. The study found that T1 and T4 perceptions are categorical, while T2 and T3 perceptions are not. The study concludes that this is why it is hard for learners to internalize the contrast between T2 and T3.

Halle et al. and Wu's studies showcased the two explanations for why NNS and NS have different tone categories. One is that NNS use tone perception strategies that are linked to their native language (Halle et al. 2004). The other attributes the difference to L2's tonal system (Wu 2005).

Stagray and Down's earlier research (1993) supports the first explanation. It is certainly possible that L2 tonal system influences NNS' perception, as the second explanation maintains. Wu's study was also based on a context-free environment, and did not provide the whole system of tones for the study. As noted above, T3 is much closer to T4 contextually because the final rise in T3 is usually absent in context. Sometimes, T3 and T4 are falling tones, whose only difference is height.

In conclusion, the studies discussed confined their efforts to investigating specific contrastive pairs rather than investigating them all. All were conducted using stimuli presented without a real context. In addition, these "context-free" experiments synthesized tones in a variety of ways. It is eminently reasonable for individual research teams to prefer tone synthesis methods that match the variation(s) of context a given study is considering. Therefore, it remains necessary to explore perceptual categories in NS and NNS perception and production systematically and in context.

2.2.5 Speaker and Rate Normalization

The term "normalization" refers to the idea that listeners normalize speaker pitch range and speaking rate in tone perception. Fox and Qi's study (1990) is typical of those that focused on neighboring tones' contextual effects. Others, among them, Moore's work (1995) examined whether speaker information is relevant to tone identification.

Moore's study (1995) dealt only with T2 and T3. The study concluded that tone identification is influenced by changes in speaker identity. It showed that pitch range influences the recognition of T2 and T3. Moore (1995: 130) found that "it is necessary for us to further clarify the relationship between acoustic variability and normalization." In the current study, we normalize F_0 to avoid the variations caused by individual's pitch range (see Chap. 3).

Moore (1995) also examined rate normalization. These data suggested that "speaker F_0 range may be more important to tone identification than rate information" (Moore 1995: 173). These results convinced us that the current study should not focus on speech rate. Therefore, speech rate was not manipulated during our analysis yet it is controlled during the experiments.

2.3 The Largest Difficulties in NNS Tone Acquisition

The studies reviewed above have shown that the difficulties for perception are different from those for production. The majority of researchers agree T2 is the most difficult tone to perceive. No specific tone stands out as the most difficult to

produce, although some studies suggest that T2 and T3 prove most troublesome to NNS while other suggest T4 and T1.

Studies of acquisition order of tones have revealed the leading reason that research on tone production has yielded heterogeneous results. One type of research used acoustic features such as the pitch height to explain tone errors, while the second type relied upon other acoustic features, like pitch contours.

Sun (1997) stated that "the other [difficulty lies] in holding in the mind the acoustic-phonetic information of tones long enough to be able to interpret the information phonemically" (Sun 1997: 201). Stagray and Downs' (1993) early categorical perception research demonstrated that NNS are sensitive to small differences in acoustic features while NS are not. It is obvious that the greatest difficulty for NNS learners lies not at the phonetic level, but at the phonological level, making it hard for learners to categorize tones. It is a formidable task for NNS without any previous tonal experience to acquire an entire tonal system with clear categories. Studies of the relationship between perception and production, such as Yang and Ankenmann's (Yang and Ankernmann 2007) experiment, also demonstrated that accessing NS' tonal system can help NNS improve their tone perception.

The gap between the phonetic and phonological levels indicates that acoustic cues do not simply map onto perceptual features. NS do not discern all the subtle nuances of the acoustic cues. NNS may respond to these acoustic differences and categorize tones differently. Some studies, such as Yang (2011), found that NS do perceive two tone categories with the same acoustic features appearing in different contexts as belonging to the same phonetic tone. These findings inevitably lead us to the conclusion that it is NNS' lack of a tonal system similar to NS' that makes tone acquisition arduous and tricky for non-tonal L1 learners of Chinese. This is why it is vital that the perceptual space of tone categories is essential in our study.

2.4 Problems with NNS Tone Acquisition Research

Previous NNS tonal acquisition research has fallen prey to one or more of five central problems: (1) attempting to study tones without real contexts; (2) avoiding the fact that the methods used to study how learners acquire differential features of Mandarin are different; (3) overlooking that speakers acquire tones as a system of contrasts, not as individual items; (4) not dealing effectively with the empirical measurement used in production experiments; (5) and, finally, not adjusting for the discrepancy between NNS study subjects with better perception as compared with those with better production.

2.4.1 Lack of Context

Contexts are central in connected speech and contextual cues play an important role in the perception of tones. A tone has different variations when it is in context or in isolation. For example, the citation contour of T3 first falls, then rises

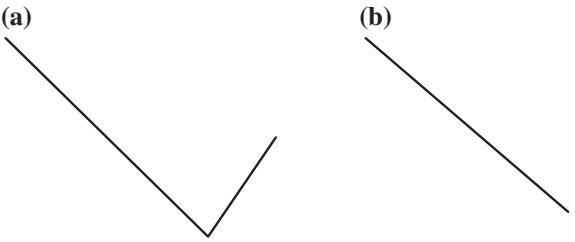

Fig. 2.1 T3 contours, **a** without a context **b** in context

without a context; however, in context, the tone has a falling contour (Fig. 2.1). Investigated outside of context, T3 is regarded as contrastive to T2 according to the tones' contours. In context, T3 can contrast with T4, which has the same falling contour but a different height. In context, contrast is height-dependent.

Each tone category also has unique tonal variations in different contexts. Yip (2002) also provides an example with regards to T3. She states that the "T3 that has the tone value 21(4)[1] is low falling that is represented by the numerical number 21 in non-final position, but acquires a rise, shown by (4) that follows 21, at the end of a phrase" (Yip 2002: 181). Yet, the question remains how these tone variations are processed during perception and production. This issue has not been completed resolved for NS speech, although some research has considered contexts (Broselow et al. 1987; Fox and Qi 1990; Leather 1983; Xu 1997).

2.4.2 Method of Acquisition Research

Acquisition order study findings (Sect. 2.2.2) have not yet conclusively identified which tone offers the greatest resistance to non-tonal L2 learners' efforts at mastery. Section 2.3 illustrates that at the phonological level, L2 learners cannot match acoustic and contextual cues with perceptual features. Chapter 1's review of tonal representation in phonology elucidated two descriptive features: contour and register. While there is no register difference in Mandarin, the language does exhibit a height difference between two tones. For example, T1 is at the high level while T3 is at the low level. Because of this, this study employs the term "register" to distinguish high level from low level tones.

White (2003: 139) stated that "L2 learners can acquire feature strength which differs from the L1, as well as features which are not instantiated in the L1". Research by Leung (2001) and Robertson (2000) showed that L2 learners acquire new categories together with those categories associated features. At the suprasegmental level, it is worth exploring whether learners of Chinese acquire tone categories together with the contour feature and the height feature (register).

[1] 21(4) denotes the pitch value of T3. The value 4 is optional.

There are researchers (e.g., Sun 1997) who maintain that Mandarin does not have register contrasts of tone categories. Shen's (1990) results can be critiqued from this position, since it accounts for the existence of height difference between some Mandarin tones. This study deems it correct to use the feature "register" to refer to "height" as Mandarin has no register difference.

Researchers such as Chen (2000), Xu (1997), and Yip (2002) considered T3's falling contour when in context, thereby we could establish its contrast with T4 (which also has a falling contour) as one of height. Unfortunately, this does not directly address L2 learners of Chinese troubles in distinguishing T2 from T3, even though T3's acoustic features are closer to T4's. This is probably because T2's start point is low, and T3 is also perceived as a dipping tone, despite the reality that it is a low tone (Zhu 2012).

2.4.3 Tones as a System of Contrasts

Speakers should acquire tones as a system of contrasts, not as individual items, a fact which directs us to study this tonal system as an integrated whole. This is in direct contrast to the vast majority of existing studies of NNS' tonal acquisition which consider exclusively one of the four citation tones in isolation, or a subset of contrastive pairs. This narrow focus creates inherent limitations in studies of categorical perception of tones, something that is evident in Halle et al. (2004), where the researchers attended to three pairs of contrastive tones: T1 and T2; T2 and T3; T3 and T4.

2.4.4 Empirical Measurements

Often NNS' pronunciation is not clear enough for NS to judge their accuracy (see Sect. 2.2.3 above), and this has important implications for the development of empirical measures of NNS' speech. Early research on acquisition order generally recruited two to ten NS to judge the accuracy of NNS tones, and then used inter-rater reliability checks on the consistency of the scores, such as Chen's study (1997). Others, such as Shen's study (1989), used acoustic features to measure the accuracy of NNS tones. Shen's study counted errors in NNS' use of register.

It is still an open question whether Mandarin has phonemic contrasts that are conditioned by register or pitch height. Acoustic features are not equal to perceptual features, as they occur at the phonetic level, while listeners perceive tones at the phonological level. Some researchers who carry out production training recognized this problem and use statistical methods to assess NNS production. For example, Wang et al. (2003) recruited 80 NS to evaluate learners' tones. However, regardless of whether a study uses inter-rater reliability or statistical calculations, researchers need a significant number of NS to participate in scoring. The more NS judges there are, the more accurate the evaluation of tone production will be.

It is impossible for even NS to judge NNS' tones as absolutely correct or absolutely wrong. The only thing NS can determine is whether the tones are similar to or far from those NS would produce, as they are perceiving the NNS' production. Developing a model of NS' perception would make it easier for us to measure to what extent NNS' tone production approximates NS' perceptual categories.

2.4.5 Differentiating Between Tonal L1 and Non-tonal L1 NNS

Most studies have compared NS' perception with NNS' or NS' production with NNS' production. NNS in these studies are either from a tonal L1 or a non-tonal L1, with a focus on learners' backgrounds. The subjects' issue in some studies was criticized by some scholars, because they used mixed subjects including both true learners and heritage learners who were born in Chinese families (see Sect. 2.2.3).

The current study is interested in both perception and production, and thus will be best served by two sets of learners: those with stronger perception and those with better production. This will allow us to compare both kinds of learners' perception and production with NS' perception.

Heritage learners (HL) generally tend to perceive tones better than they produce them. Most true learners (TL), lacking any background in Mandarin, actually tend to produce more accurately than they perceive tones. In the USA, heritage learners of Chinese are unique. Their parents speak Chinese at home while HL mainly speak English. This causes HL' perception to be far superior to their production. Most heritage learners of Chinese are not fluent in the language; yet, their comprehension far outstrips that of TL who are at a similar level in terms of production.

In order to compare learners' perception and production with NS' perception, we will explore heritage learners' perception and production, as well as true learners'. Comparing these two kinds of learners is worthwhile, as it presents a twofold model of perceptual and production development: one in which perception has priority, and the other in which production exceeds perception.

The research on NNS tonal acquisition makes clear that further study of the NS' tonal system is required, particularly with regard to NS' perception of tones. In response, this study will establish NS perceptual categories and compare them with NNS' tonal perception and production.

2.5 Conclusions and Research Questions

The literature review brings to light two issues in need of further investigation: (a) the establishment of a perceptual tone system for Mandarin Chinese; (b) a closing of the gap between acoustic features and perceptual results. Addressing these two issues will permit us to compare NS perception with TL' and HL' perception and

production. Given these issues, we need a study that (a) sets up NS perceptual categories for Mandarin tones systematically and in context; (b) explores the acoustic and contextual cues that could contrast tones in NS perception; (c) generates a model against which we can evaluate TL' and HL' tonal perception and production.

Chapter 3 presents the argument for the selection of appropriate parameters for this study. It outlines the need to observe NS' tone production, and how perceptual experiments could be designed based on the features of that production.

The following research questions will be used to set up a model of perceptual categories for Mandarin tones, and then compare it with NNS' perception and production, and further explore the connection between acoustic features and psychological features of NNS' tones.

1. What is the production space for native speakers?
2. What are the common features of the NS perceptual maps across different contexts? What are the features of the perceptual maps according to the different contexts? What differences exist between NS' tone perception and NS' tone production? How do acoustic cues and contextual features influence NS' perception?
3. What are the features of the perceptual categories of T1, T2, T3, and T4 for TL and HL? How do acoustic cues and contextual features influence TL and HL strengths and weaknesses of perception in comparison with NS' perceptual categories?
4. Is the TL' perceptual map similar to the TL' production map? If not, what are the features of TL' production? How do acoustic or contextual features in TL' production influence NS' perception? Is the HL' perceptual map similar to HL' production map? If not, what are the features of HL' production? How do acoustic or contextual features in HL' production influence NS' perception?

References

Abramson, A. 1979. The co-articulation of tones: An acoustic study of Thai. In *Studies in Thai and Mon-Khmer phonetics and phonology: In honour of Eugenie Henderson*, ed. T.L. Thongkum, V. Panupong, P. Kullavanijaya, and M.R.K. Tingsabadh, 1–16. Bangkok: Chulalongkorn University Press.

Blicher, D.L., R.L. Diehl, and L.B. Cohen. 1990. Effects of syllable duration on the perception of the Mandarin tone2/tone3 distinction: Evidence of auditory enhancement. *Journal of Phonetics* 18: 37–49.

Broselow, E., R. Hurtig, and C. Ringen. 1987. The perception of second language prosody. In *Interlanguage phonology: The acquisition of a second language sound system*, ed. G. Ioup, and S. Weinberger, 350–361. Cambridge, MA: Newbury House.

Burnham, D. 2000. Excavations in language development: Cross-linguistic studies of consonants and tone perception. In *Interdisciplinary approaches to language processing: The international conference on human and machine processing of language and speech*, ed. D. Burnham, S. Luksanneeyanawin, C. Davis, and M. Lafourcade, 44–69. Bangkok: Chulalongkorn University Press.

Burnham, D., and C. Jones. 2002. Categorical perception of lexical tone by tonal and non-tonal language speakers. Paper presented at the 9th International Conference on Speech Science and Technology, December 2–5, in Melbourne.

Burnham, D., and K. Mattock. 2008. The perception of tones and phones. In *Language experience in second language speech learning: In honor of James Emil Flege*, ed. O.-S. Bohn, and M.J. Munro, 259–280., Language Learning and Language Teaching Amsterdam: John Benjamins.

Chan, S.W., C.K. Chuang, and S.-Y. Wang. 1975. Cross-linguistic study of categorical perception for lexical tone. *Journal of the Acoustical Society of America* 58: S119.

Chang, Y.-C., and P. Halle. 2000. Taiwan Huayu shengdiao fanchou ganzhi. *Hsing Hua Journal of Chinese Studies, new series, XXX* 1: 51–65.

Chen, Q.-H. 1997. Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers' Association* 32(1): 21–39.

Chen, M. 2000. *Tone sandhi: Patterns across the Chinese dialects*. Cambridge: Cambridge University Press.

Chiang, T. 1979. Some interferences of English intonations with Chinese tones. *International Review of Applied Linguistics* 17(3): 245–250.

Cruttenden, A. 1997. *Intonation*. England: Cambridge University Press.

Cutler, A. 1994. Segmentation problems, rhythmic solutions. *Lingua* 92: 81–104.

Cutler, A., J. Mehler, D. Norris, and J. Segui. 1992. The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology* 24: 381–410.

Eimas, P.D. 1963. The relationship between identification and discrimination along speech and nonspeech continua. *Language and Speech* 6: 206–217.

Elliot, C.E. 1991. The relationship between the perception and production of Mandarin tones: An exploratory study. *University of Hawai'i Working Papers in ESL* 10(2): 177–204.

Flege, J.E. 1986. The production and perception of foreign language speech sounds. In *Human Communication and Its Disorders*, vol. 2, ed. H. Winitz, 224–401. Norwood, NJ: Ablex.

Fox, R. and J. Unkefer. 1985. The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics* 13: 69–90.

Fox, R., and Y.-Y. Qi. 1990. Context effects in the perception of lexical tone. *Journal of Chinese Linguistics* 18: 261–283.

Francis, A., V. Ciocca, and B. Ng. 2003. On the (non)categorical perception of lexical tones. *Perception and Psychophysics* 65(7): 1029–1044.

Fry, D., A.S. Abramson, P.D. Eimas, and A.M. Liberman. 1962. The identification and discrimination of synthetic vowels. *Language and Speech* 5: 171–189.

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

Gandour, J., and R.A. Harshman. 1978. Cross-language differences in tone perception: A multidimensional scaling investigation. *Language and Speech* 21: 1–33.

Halle, P., Y. Chang, and C. 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(2): 269–279.

Kiriloff, C. 1969. On the auditory perception of tones in Mandarin. *Phonetica* 20: 63–64.

Leather, J. 1983. Speaker normalization in perception of lexical tone. *Journal of Phonetics* 11: 373–382.

Leather, J. 1987. The interrelation of perceptual and productive learning in the initial acquisition of second-language tone. In *Second-language speech: Structure and process*, ed. A. James, and J. Leather, 75–101. Berlin: Mouton de Gruyter.

Leather, J. 1990. Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In *New sounds 90: Proceedings of the 1990 Amsterdam symposium on the acquisition of second language speech*, ed. J. Leather, and A. James, 72–97. Amsterdam: University of Amsterdam.

Leung, Y.-K.I. 2001. The initial state of L3A: Full transfer and failed features? In *The past, present and future of second language research: Selected proceedings of the 2000 second language research forum*, ed. X. Bonch-Bruevich, W. Crawford, J. Hellerman, C. Higgins, and H. Nguyen, 55–75. Somerville, MA: Cascadilla Press.

Liberman, A.M. 1975. The intonation system of English. Ph.D. dissertation, Massachusetts Institute of Technology.

Liberman, A., K. Harris, H. Hoffman, and B. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54: 358–368.
Miracle, W.C. 1989. Tone production of American students of Chinese: A preliminary acoustic study. *Journal of the Chinese Language Teachers' Association* 24(3): 49–65.
Moore, C.B. (1995). *Speaker and rate normalization in the perception of lexical tone by Mandarin and English listens*. Unpublished Ph.D. dissertation, Cornell University.
Peng, G., H. Zheng, T. Gong, R. Yang, J. Kong, and W. Wang. 2010. The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics* 38: 616–624.
Pierrehumbert, J. 1980. The phonology and phonetics of English intonation. MIT dissertation.
Pierrehumbert, J. 1981. Synthesizing intonation. *Journal of the Acoustical Society of America* 70: 985–995.
Pisoni, D., and J. Lazarus. 1975. Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America* 55: 328–333.
Pisoni, D., and J. Tash. 1974. Reaction times to comparisons within and across phonetic boundaries. *Perception and Psychophysics* 15: 285–290.
Read, C., Y. Zhang, H. Nie, and B. Ding. 1986. The ability of manipulate speech sounds depends on knowing alphabetic writing. *Cognition* 24: 31–44.
Robertson, D. 2000. Variability in the use of the English article system by Chinese learners of English. *Second Language Research* 16: 135–172.
Rosen, S., and P. Howell. 1987. Auditory, articulatory and learning explanations of categorical perception of speech. In *Categorical perception: The groundwork of cognition*, ed. S. Harnad, 113–160. Cambridge, UK: Cambridge University Press.
Schouten, M., and A. Van Hessen. 1992. Modeling phoneme perception I: Categorical perception. *Journal of the Acoustical Society of America* 92: 1841–1855.
Schwanhauber, B., D. Burnham, and C. Jones. 2003. Categorical perception of tone by tonal and non-tonal language speakers. Paper presented at the 14th Australian language and speech conference, Brisbane, Australia.
Sereno, J.A., and Y. Wang. 2008. Behavioral and cortical effects of learning a second language: The acquisition of tone. In *Language experience in second language speech learning: In honor of Manes Emil Flege*, ed. O. Bhon, and M. Munro, 239–258. Amsterdam: John Benjamins.
Shen, X.-N.S. 1989. Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers' Association* 24(3): 27–47.
Shen, X.-N.S. 1990. Tonal coarticulation in Mandarin. *Journal of Phonetics* 8: 281–295.
Shen, X.-N.S., and M. Lin. 1991. A perceptual study of Mandarin tones 2 and 3. *Language and Speech* 34: 145–156.
Soares, C. 1982. Converging evidence for left hemisphere language lateralization in bilinguals. *Neuropsychologia* 20: 653–659.
Sorensen, J., and W. Cooper. 1980. Syntactic coding of fundamental frequency in speech production. In *Perception and production of fluent speech*, ed. R.A. Cole, 399–440. Hillsdale: Lawrence Erlbaum.
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(1): 143–163.
Studdert-Kennedy, M., A. Liberman, K. Harris, and F. Cooper. 1970. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review* 77: 234–249.
Sun, S.H. 1997. *The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese*. Honolulu: University of Hawai'i Press.
Sussman, H., P. Franklin, and T. Simon. 1982. Bilingual speech: Bilingual control? *Brain and Language* 15: 125–142.
Wang, S.-Y.W. 1967. Phonological features of tone. *International Journal of American Linguistics* 33: 93–105.
Wang, Y., M. Spence, A. Jongman, and J. Sereno. 1999. Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America* 106(6): 3649–3658.
Wang, Y., A. Jongman, and J. Sereno. 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America* 113(2): 1033–1043.

Wang, Y., D. Behne, A. Jongman, and J. Sereno. 2004. The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics* 25: 449–466.
White, L. 2003. *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.
Wood, C. 1976. Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America* 60: 1381–1389.
Wu, X. 2005. (Non-)categorical perception of Mandarin tones: A comparison between speakers of tone and non-tone languages. M.A. thesis, Victoria: University of Victoria.
Wuillemin, D., and B. Richardson. 1994. Right hemisphere involvement in processing later-learned languages in multilinguals. *Brain and Language* 46: 620–636.
Xia, J., L. Zhang, H. Shua, Y. Zhang, and P. Lie. 2010. Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Cognitive, Behavioral, and Systems Neuroscience* 170(29): 223–231.
Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61–83.
Xu, Y. 1999. Effects of tone and focus on the formation and alignment of f_0 contours. *Journal of Phonetics* 27: 55–105.
Xu, Y., J. Gandour, and A. Francis. 2006. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of Acoustical Society of America* 120(2): 1063–1074.
Yang, B., and R. Ankernmann. 2007. The gap between the perception and production of tones by American learners of Mandarin. Presented at American Association for Applied Linguistics annual conference. Costa Mesa, CA.
Yang, B. 2011. Perceptual categories of Mandarin tones. In *Proceeding of Psycholinguistic Representation of Tone Conference 2011*, Hong Kong.
Yang, B. 2012. The gap between the perception and production of tones by American Learners of Mandarin: An intralingual perspective. *Chinese as a Second Language Research* 1(1): 31–52.
Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.
Zheng, H.-Y., J. Minetta, G. Peng, and W. Wang. 2010. The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study. *Language and Cognitive Processes* 27(2): 184–209.
Zhu, X. 2012. Jiangdiao de Zhonglei. *Yuyan Yanjiu* 32(2): 1–16.

http://www.springer.com/978-3-662-44644-7

Perception and Production of Mandarin Tones by Native Speakers and L2 Learners

Yang, B.

2015, X, 162 p. 39 illus., 25 illus. in color., Hardcover
ISBN: 978-3-662-44644-7