LANGUAGE-SPECIFICITY IN AUDITORY PERCEPTION OF CHINESE TONES

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the Graduate

School of The Ohio State University

By

Tsan Huang, B.A., M.A.

* * * * *

The Ohio State University

2004

Dissertation Committee:

Dr. Keith Johnson, Adviser

Dr. Elizabeth Hume

Dr. Marjorie K.M. Chan

Approved by

____

Adviser
Department of Linguistics

ABSTRACT

This dissertation investigates the phenomenon of language-specificity in the auditory perception of Chinese tones. Chinese and American English (AE) listeners participated in a series of perception experiments, which involved short ISIs (300ms in Experiment 1 and 100ms elsewhere) and an AX discrimination (limited set in Experiments 2 and 3, speeded response in Experiments BJ, RG and YT) or AX degree-of-difference rating (Experiment 4) task. All experiments used natural speech monosyllabic tone stimuli, except Experiment 2, which used sinewave simulations of Putonghua (Beijing Mandarin) tones. AE listeners showed psychoacoustic listening in all experiments, paying much attention to onset and offset pitch. Chinese listeners showed language-specific patterns in all experiments to various degrees. The most robust language-specific effects of Putonghua were found in Experiments 1, 3 and 4, where the T214 (as well as T35) neutralization rule shortened the perceptual distance between T35 and T214 (or that between T55 and T35) for Chinese listeners. Cross-dialectal as well as age differences were observed among Chinese listeners in Experiments BJ, RG and YT using natural speech stimuli from Putonghua, Rugao (a Jianghuai Mandarin dialect, Jiangsu Province) and Yantai (a Northern Mandarin dialect, Shandong Province), respectively. Listeners generally showed native advantage in perceiving tones in their own dialects. Cross-dialectal tone category correspondences (R44 to T51 and Y55 to T51) caused more confusion for older Rugao and Yantai listeners between the relevant tones. Furthermore, Yantai older listeners, with more sandhi rules in their dialect, showed different perceptual patterns from other listeners, including Yantai young listeners. Since these experiments employed procedures hypothesized to tap the auditory trace mode (e.g. Pisoni, 1973; Macmillan, 1987), language-specificity found in this dissertation seems to support the proposal of an auditory cortical map (Guenther et al. 1999). But the data also suggest that the model need to be refined to account for different degrees of language-specificity, which are better handled by the lexical distance model advanced by Johnson (2004), although the latter model may be a bit too rigid on how much lexical interference is allowed in low-level auditory perception.

Dedicated to the memory of my grandparents

ACKNOWLEDGMENTS

Many people contributed in various ways to the work reported in this dissertation. Although I will not be able to mention everyone's name here, please know that I will forever be indebted to your help.

Let me start by thanking my advisor Keith Johnson, who has been an inspiration in many ways and an important factor in my survival of graduate school. He contributed a lot to experimental designs as well as ideas discussed in this dissertation. I appreciate his patience for me, which he obviously has more for me than I do for myself! Keith has also been very generous with his time this summer, sacrificing his summer vacation. I am especially grateful for his always being willing to stand behind me even in situations that do not have anything to do with school. Thank you, Keith, for being a mentor, a role model and a friend.

I am grateful to Beth Hume for her guidance, encouragement and tolerance. I thank her for being a passionate teacher. I have always enjoyed her thought-provoking lectures and discussions, both as a student and as a teaching assistant. I also thank her for being a warm motherly (and young and charming!) figure in the lab.

Thanks also go to my other committee member Margie K.M. Chan for her guidance and for introducing me to the various topics in Chinese linguistics. I really enjoyed receiving the "red envelop" around the Chinese New Year's Day and feeling like a kid again, Margie! And despite the mosquito bites, thanks for hosting the many relaxing get-togethers in your backyard!

Coming to this department for my postgraduate education was one of the best decisions I have ever made. I shall forever cherish my experience here. I thank all my teachers for being part of this experience. I especially thank Professor Carl Pollard for convincing me to come to Ohio State and for telling me the story about the senior Professor Pollard's experience as a young professor being assigned to teach a course whose content he was not familiar with. When I have to look into things that I had no previous knowledge of, I often think of the Dean's words ("Pollard, you will know.") and the senior Professor Pollard's courage to take the assignment. A special thank-you to my teacher Mary Beckman, who has a seven-day work schedule yet still found time whenever I requested a meeting with her. With her clear vision of things, Mary always has a way – and lots of patience – to help me see where the problem is. Thanks also go to Professors Brian Joseph, Peter Culicover and Rich Janda for having been so supportive in some rather frustrating situations. Thanks to Professor Michael Broe for the statistic help (and the lecture notes!). I thank the wonderful staff in the department, especially Jim Harmon for technical support and Jane Harper for being so efficient and so kind. I also thank my fellow graduate students for their companionship, help and moral support throughout the years, especially Allison Blodgett, Robin Dautrocourt, Paul Davis, Hope Dawson, Janice Fon, Shelome Gooden, Craig Hilts, Martin Jansche, Ok Joo Lee, Jianguo Li, Matt Makashay, Laurie Maynell, Amanda Miller, Michelle Ramos-Pellicia, Helena Riha, Misun Seo, Giorgos Tserdanelis, Pauline Welby, Steve Winters, Wai-yi Peggy Wong, XU Lei and Kiyoko Yoneyama, to name just a few.

A special thank-you is due to Elizabeth Cooke for freeing me of some unwanted worries during my dissertation writing, and to Professors Burton Rosner and Will Hopkins for suggestions regarding statistical analyses.

The speakers who produced the stimuli recordings for me, Jianguo Li (for Beijing), CAI Xiaoxiang (for Rugao) and ZHANG Conghao (for Yantai); my contacts and consultants, CAO Wen and JIANG Xin (in Beijing), WU Fengshan and LIU Desheng (in Rugao), and CHI Juhua and Prof. ZHANG Lijun (in Yantai), and all my other speakers and listeners deserve a sincere thank-you as well. I also thank Kris Pokorny and Terah Shamberg for running the experiments for me at The Ohio State University, Columbus campus.

I also acknowledge the support I received for part of this research: Alumni Grants for Graduate Research and Scholarship (AGGRS, Spring 2003) from the Graduate School of The Ohio State University) and Grant No. 5 R01 DC04421 from the National Institute on Deafness and Other Communication Disorders (Principle Investigator: Keith Johnson).

Last but not least, I thank everyone in my extended family for their love and care, especially my parents, who had to go through the excruciating moments during my illness and who have been taking care of me during the whole period when I wrote this dissertation. My husband, my siblings, my aunts and uncles, my cousins and my in-laws have all been very supportive. My parents and my parents-in-law had all participated in running the experiments.

VITA

February 16, 1968 …………………………………….. Born - Rugao, Jiangsu Province, China.

July, 1990 ……………………………………………… B.A., Beijing Language Institute, Beijing, China.

1990 – 1997 …………………………………………… Lecturer of English, Capital University of Medical Sciences, Beijing, China

Summer, 1998 ………………………………………… M.A., Department of Linguistics, Ohio State University, Columbus.

1998 – 2003 …………………………………………… Graduate Teaching and Research Associate, Ohio State University.

2003 – 2004 …………………………………………… Visiting Assistant Professor, University at Buffalo, State University of New York.

PUBLICATIONS

1. **Huang, Tsan** (in press). Effects of Preceding Tone and Downstep on Target Tone Realization in Rugao Chinese. In *Proceedings of the Linguistics and Phonetics Conference, 2002*. Urayasu, Japan.

2. Peng, S.-H., M. K.-M. Chan, C.-Y. Tseng, **T. Huang**, O.-J. Lee, and M. E. Beckman, (in press). Towards a Pan-Mandarin System for Prosodic Transcription. Refereed chapter in Jun, S.-N. (ed.), *Prosodic Typology and Transcription: A Unified Approach*. Oxford, UK: Oxford University Press.

3. **Huang, Tsan**, 2001. The interplay of perception and phonology in Tone 3 sandhi in Chinese Putonghua. In Hume, Elizabeth and Keith Johnson (eds.), *Studies on the interplay of speech perception and phonology* (OSU working papers in linguistics, No. 55), pp. 23-42.

4. Muller, J. S., **T. Huang**, and C. Roberts (Eds.), 2000. Varia. The Ohio State University Working Papers in Linguistics (OSUPL), No. 54.

FIELDS OF STUDY

Major Field: Linguistics

Studies in:

Phonetics
Speech Perception

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The interplay between speech perception and phonology should be considered bi-directionally (Hume & Johnson, 2001): in the one direction, how speech perception shapes synchronic phonology and influences diachronic sound change; in the other, how phonology may influence speech perception. There is currently a growing interest in the former of the two (e.g., Hume & Johnson, 1999, 2001; Lindblom, 2000), perhaps as a corrective for decades of interest in the potential influence of speech production on phonology. This dissertation is focused, however, on the latter direction, because the language-specificity of speech perception – if such a thing exists – serves as context and caveat for the whole question of how speech perception influences phonology.

The influence of perception on phonology and historical sound change has been discussed by various researchers from very early on (e.g., Trubetzkoy, 1969, 1939). In Jakobson, Fant, & Halle (1952), perceptual features are treated as primary (see also Jakobson & Halle, 1956). But the generative tradition of phonology since Chomsky & Halle (1968) centers on articulatory definitions of distinctive features (but cf. Liljencrants

& Lindblom, 1972; Ohala, 1981). Now a revival of interest in the interplay of perception and phonology seems to be in progress. In research work done in the past 15 years or so, Kohler (1990), Hura, Lindblom & Diehl (1992), and Steriade (2001) hold that phonological processes such as segmental reduction, deletion, and assimilation are perceptually tolerated articulatory simplification and that the direction of such processes is determined by perception. That is, these processes only take place when the output of such a change is found to be highly confusable with the input perceptually. In particular, Steriade (2001) proposes a universal P-map, which organizes speech sounds in terms of their perceptual salience in different contexts. Johnson (2004) also presents data and analysis in support of the P-map. If a certain contrast is perceptually weak in a certain position, synchronic phonology works to enhance or sacrifice it by way of epenthesis, metathesis, dissimilation, assimilation or deletion (Hume & Johnson, 2001). For example, vowel insertion between sibilants in some English plural noun forms (e.g., buses, bushes, judges; cf. cats, cans), metathesis of /skt/ to [kst] in Faroese and Lithuanian (Hume & Seo, 2004), and manner dissimilation of two consecutive obstruents in Greek (e.g., /kt/ → [xt] or /xθ/ → [xt]; Tserdanelis, 2001) serve to strengthen the syntagmatic contrast between these neighboring segments – although the latter two of these processes also result in paradigmatic contrast neutralization –, while n-lateralization in the /nl/ and /ln/ sequences in Korean and the optional /h/ deletion in Turkish sacrifice the perceptually weak contrasts for ease of articulation, leading to syntagmatic contrast neutralization in both cases (Seo, 2001; Mielke, 2003).

Ohala (1981) suggests that diachronic sound changes may occur due to the listener's misperception and reinterpretation of certain sounds or sound sequences. Instead of correcting distorted phonetic forms based on his knowledge of possible variants of the common underlying forms, the listener-turned-speaker may as well exaggerate the distortion, resulting in historical sound change (p.183). Janson (1983) further hypothesizes a five-stage sound change process involving interaction between perception and production, using as an example the change of /r/ to /R/ in Norwegian (1983: 24). Guion (1998) suggested that the cross-linguistically common sound change of velar palatalization is also perceptually conditioned. Examples of perceptual effects on sound change can also be found in historical tonogenesis and tone developments in the tone languages, where previously redundant pitch differences became contrastive when the conditioning segmental contrast was lost (e.g. Maspero', 1912; Haudricourt, 1954a,b; Matisoff, 1973; Maran, 1973; Hombert, 1978; Hombert, Ohala & Ewan, 1979; Svantesson, 2001).

This dissertation focuses on the other aspect of the phonology-perception interplay, namely how phonology influences perception, and in particular how differences in tonal inventories and neutralization processes may affect tone perception by native listeners, in comparison with non-native listeners speaking a non-tone language or another tone language. It is important to test the influence of phonology on perception, i.e. language-specificity, because much of the work on the other interaction – the influence of perception on phonology – assumes that the perceptibility of speech sounds is uniform across languages and that appeal to a universal perceptual space can be made.

In the sections that follow, I shall first review previous research in this field – mainly on the influence of segmental phonology on speech perception – before moving on to report the results from a series of tone perception experiments that sought to further investigate this phenomenon in the subsequent chapters.

1.1 Influence of Segmental Phonology on Perception

1.1.1 Influence of segmental inventory on perception

Despite claims for a universal P-map (e.g., Steriade, 2001), listeners' perception of speech sounds appears to depend on the inventories of contrastive sounds in their native languages. For instance, Japanese listeners, whose language has only one liquid sound, perceive the /r-l/ distinction differently from American English speakers (see, e.g., Miyawaki, Strange, Verbrugge, Lieberman, Jenkens & Fujimura, 1975). Werker & Tees' (1984) study on the perception of Hindi voiceless unaspirated retroflex versus dental stops and Thompson[1] velar versus uvular ejectives by native English speakers showed that listeners were unable to distinguish these foreign sounds contained in the /ʈa, ta/ and /k'i, q'i/ syllables. Lahiri & Marslen-Wilson (1991), in a gating task, showed that listeners' responses were predictable based on how a phonetic cue, vowel nasalization, is employed in a particular language. Consequently, English listeners were able to guess whether the following sound (not heard in some of the gated stimuli) was a nasal or not based on the presence or absence of nasalization, whereas Bengali listeners, whose language contrasts nasalized vowels with non-nasalized ones, could not predict the

[1] Thomson refers to the languages of one of the First Nations people of British Columbia.

following sound. Beddor & Strange (1982) reported that an oral-nasal vowel series was perceived categorically by the Hindi speakers whose language is similar to Bengali in having an oral-nasal contrast in vowels, while the same series was perceived more continuously by the English speakers whose language lacks such a contrast. Hume, Johnson, Seo, Tserdanelis & Winters (1999) found that while consonant-vowel transition seems to provide more place information for consonant place identification than stop burst for both American English and Korean listeners, the difference between the two kinds of stimuli is greater for Korean listeners. Hume *et al.* suggest that this is related to differences in phonological contrasts in these two languages; that is, Korean listeners with a three-way (tense, lax and aspirated) stop consonant contrast, which is cued in part by the amount/duration of aspiration, seem to pay more attention to the CV transition between the burst and the vowel onset than do the English listeners, who have a two-way (unaspirated versus aspirated) stop contrast.

Hume & Johnson (2003) suggest that four degrees of contrast should be taken into account in predicting the influence of phonology on perception, namely fully contrastive, partially contrastive, allophonic, and non-occurring. They state that the influence of contrast on perception of native sounds "is not all or none". In particular, a neutralization rule will shorten the perceptual difference between two categories that are otherwise contrastive in the language. Harnsberger (2001) and Hall, Hume, Johnson & Reiter (2004) present evidence for the claim that allophony can also affect perception.

There is also a large body of literature on perception of nonnative phonetic categories (see Strange 1995 for a review). In general, studies show that adults have difficulty perceiving nonnative contrasts, but that the degree of difficulty depends on the psychophysical salience of the acoustic parameters differentiating these nonnative contrasts (e.g., nonnative VOT categories might be easier to train subjects on than nonnative place categories, according to Werker & Tees' (1984) findings), similarities or differences between native and nonnative categories, as well as native phonotactics.

Best and colleagues (Best, 1995; Best, McRoberts, & Goodell, 2001) argue for a gradient discrimination performance of non-native contrasts as predicted by the perceptual assimilation model (PAM), which assumes that nonnative categories may be assimilated/mapped to native categories, assimilated as within category tokens differing in "category goodness", uncategorizable but treated as speech sounds, or nonassimilable (treated as nonspeech sounds). Specifically, PAM predicts that discrimination of non-native contrasts should be near-ceiling if perceived as phonologically equivalent to native contrasts, that listeners would discriminate a non-native contrast fairly well if they perceive it as a phonetic distinction between good and poor exemplars of a native contrast, and that a much lower performance is predicted if listeners perceive the non-native contrast as equally good/bad exemplars of a single native category. Similarly, Polka (1991) attributes differences in non-native perceptual performance to phonemic status of the non-native contrast in the listener's native phonology, phonetic experience of the non-native phones by the listener as allophones in his/her native language, and differences in acoustic salience of a particular cue in the listener's native language.

1.1.2 Influence of phonotactics on perception

The phonotactics of a language also has an effect on speech perception. Massaro & Cohen (1983) and Pitt (1998) found that phonotactic constraints biased native listeners' identification toward permissible sound sequences in English when perceiving continua whose two ends consist of a voicing or place contrast. Flege & Wang (1989) found that native language phonotactics determine how listeners would perceive certain positional contrasts in a foreign language. Their data obtained from three groups of Chinese listeners' perception of English word-final /t/-/d/ contrast are rather interesting: Cantonese-speaking listeners whose native language has unreleased word-final stops /p, t, k/ did better than Shanghainese-speaking listeners whose language allows only glottal stop and sonorants in coda position. In turn, Shanghainese-speaking listeners did better than Mandarin-speaking listeners, whose language has only nasal consonant or glide codas. Obviously, having an obstruent (including glottal stop) in coda position in one's language makes one pay more attention to contrast among obstruent in this position.

Prevalent versus restricted occurrences of a sound in different languages may also lead to different perceptual patterns, as has been shown in Mielke's (2003) study of /h/ perception by English, French, Arabic and Turkish speaking listeners. In general, speakers of English and French (two languages with restricted distribution of /h/) showed lower differential sensitivity than speakers of Arabic and Turkish (two languages in which /h/ is a very common sound).

1.1.3 Influence of phonological processes on speech perception

Phonological rules operating in the listener's native language influence his/her perception as well. Fox (1992), in a series of four experiments, found that English listeners fare poorly in identifying or discriminating vowels in the neutralizing context of /hVr(d)/. Fox suggests that knowledge of the phonological rule that neutralizes vowel contrast in this context may have affected the ability of listeners to make perceptual decisions about vowel quality.

1.2 Influence of Phonology on Tone Perception

The studies of segmental perception reviewed above indicate that speech perception ability is language-specific – that phonology influences the perception of speech quite a bit. In this dissertation, I am testing this conclusion in the domain of lexical tone perception in various dialects of Chinese. Therefore, it is important to establish that tone perception, like segment perception, is language-specific, i.e. influenced by the phonology of the listener's native language.

1.2.1 Influence of lexical tone inventory on perception

Gandour (1983, 1984) and Lee, Vakoch & Wurm (1996) showed that differences in lexical tone inventory may play a role in tonal perception. In Gandour's (1983, 1984) study using 19 synthesized f0 stimuli (five levels, four rising, four falling, three falling-rising, three rising-falling), 200 speakers/listeners of Mandarin (Taiwan), Cantonese (Hong Kong), Taiwanese, Thai, and English made dissimilarity judgments on tonal pairs on an 11-point scale (0= no difference; 10 = extreme difference). Results show that the tones were rated significantly differently by tone versus nontone language speakers, by Thai versus Chinese (Mandarin and Taiwanese) speakers, and by Cantonese versus Mandarin and Taiwanese speakers. Such differences were attributed in part to differences in tonal inventories. In particular, tone height seems to be more important for English listeners, while Thai listeners attached most importance to the direction of f0 contour (rising versus falling). When the tone language groups were compared, Cantonese listeners seem to utilize mainly the dimension of tone levels/heights, which is not surprising, given that four of the six Cantonese tones have basically level contours.

Lee et al. (1996) used naturally recorded stimuli of Cantonese and Mandarin tones on word and nonword syllables and Cantonese, Mandarin (Taiwan and Mainland) and English (US) listeners in two "same/different" discrimination experiments. They found that Cantonese and Mandarin listeners were better at discriminating tones in their own dialect and that the tone language speaking listeners did better than the English group.

1.2.2 Influence of Tonal Context on Perception

Using synthetic tokens on high level-rising continua of /ba, da, bi/ syllables embedded in natural speech carrier phrases (spoken by a different speaker and hence of a different pitch range) and identification tests on Chinese and English listeners, Fox & Qi (1982) found a limited effect of contextual/speaker pitch range information. They did not find a significant language effect, leading them to the conclusion that "the contextual effect can involve auditory perceptual mechanisms". However, Lin & Wang (1985) found that native perception of a level contour (set at 115Hz) may be influenced by the f0 onset of a following high-falling contour. The frequency drop was always 40 Hz in the falling contour, but the f0 onset had four different values: 110, 120, 130, and 140. It was found that the higher the f0 onset value of the falling contour, the more the listeners misidentified the preceding level contour as a rising tone.

1.2.3 Influence of lexical status of stimuli on tone perception

Fox and Unkefer (1985) tested the influence of lexical status of stimuli on tone perception. Their stimuli consisted of four continua of synthetic Mandarin T55 (high level) and T35 (rising) (see Chapter 2 for description of Standard Mandarin tones), with one pair having words at both ends of the continuum, the second and the third having word at only one end, and the fourth having nonwords at both ends. Eleven (11) Chinese (Taiwan) subjects and eleven (11) American English subjects listened to these continua and performed a forced-choice identification task; that is, subjects had to identify a stimulus token as having either T55 or T35. The results show that there were relatively more word responses than nonword responses in the nonword/word and the word/nonword pairs for the Chinese group and that such a pattern is significantly different from the word/word pair for the Chinese ($p<.05$). However, this difference was not found in any continuum for the English group. (They did not show the results for the nonword/nonword pair, saying that the patterns were "anomalous"(p.80).)

Lee et al. (1996) also found a lexical effect with the Cantonese listeners who did better at discriminating word-word pairs than word/nonword pairs in the experiment using Cantonese stimuli. This effect was not found with the Mandarin (Mainland) group using Mandarin stimuli. One wonders why there is such a cross dialectal disparity. We are not sure whether the listeners are all monolinguals of Cantonese or Mandarin. In any case, these results regarding the lexical status of the stimuli indicate that in addition to the suprasegmental status of the tones, the Chinese morphemes/words are specified for both segments and tones.

1.2.4 Influence of tone sandhi rules on perception

Gandour (1981; 1983; 1984) suggests that tone sandhi rules may also influence tonal perception. Using INDSCAL (Carroll & Chang, 1970), Gandour (1981) reanalyzed Fok-Chan's (1974) confusion data from native listener identification of naturally produced Cantonese tones. He found that the high falling tone was placed midway between the level and the contour tones. He argues that this is due to the fact that this tone has a high level allotone in Cantonese. Although the allotone was not present in the stimuli, allophony still interfered with listeners' perception. The effect of the same allophonic alternation showed up in Gandour's (1983; 1984) data, where Cantonese listeners perceived a /44/ (high level) contour to be similar to a /53/ (high falling). In the same data, Mandarin listeners perceived the /44/ contour to be similar to /35/ (rising), which, as Gandour argues, is due to the existence of the allophonic rule that turns a rising tone to a high level in Mandarin (e.g., Chao, 1948, 1965, 1968; Cheng, 1973; see also Chapter 2).

In Huang (2001), I have also argued that listeners' native tonology may influence their perception of tones. In particular, the Mandarin T214 sandhi rule shortens the distance between T214 and T35, the latter of which is itself a contrastive Mandarin tone as well as the surface sandhi form of T214 before another T214. The data of that study will be reanalyzed in Chapter 2.

1.3 Modeling Language-Specificity in Speech Perception

As can be seen from the studies cited in §1.1 and §1.2 above, linguistic experience (or the speaker's/listener's interpretation of his/her segmental or tonal inventory as well as the native phonotactics and phonological rules) can lead to language-specific patterns in speech perception. Different category boundaries and discrimination patterns have also been found for listeners of different language backgrounds in perceptual data from studies using synthetic continua simulating changes from one speech sound to another along a certain acoustic dimension at equal steps. These phenomena came to be described as *categorical perception* (Liberman, Harris, Hoffman & Griffith, 1957) and *perceptual magnet effect* (Kuhl, 1991).

As an illustration of categorical perception, Abramson & Lisker (1970) found that for the same synthetic VOT continuum, English-speaking listeners' discrimination data showed one peak on the border of the English /b/-/p/ categories, whereas Thai-speaking listeners' data had two peaks, one on the border of the Thai /b/-/p/ categories and the other on the border of the Thai /p/-/p^h/ categories. Abramson & Lisker (1970) concluded that this resulted from specific linguistic experience of the two groups of listeners. This also seems to be the case with native Hindi speakers' categorical perception of an oral-nasal vowel contrast versus native English speakers' continuous perception of the same series (Beddor & Strange, 1982).

The perceptual magnet effect was first obtained by Kuhl (1991) using synthetic vowel stimuli. Kuhl (1991) found that human adults and human infants as young as 6 months old could not discriminate tokens closer to the prototype (or best exemplar) of the vowel as well as they could tokens farther away from the prototype although the stimuli were equally spaced along certain acoustic dimension(s). In a study on American and Swedish infants in their respective native language settings using tokens of the American English high front unrounded vowel /i/ and the Swedish high front round vowel /y/, Kuhl, Williams, Lacerda, Stevens & Lindblom (1992) found that infants as young as 6 months showed the perceptual magnet effect for native sounds but not for nonnative sounds.

Lexical tone perception by native speakers has also been shown to manifest the pattern of categorical perception (Wang, 1976). In Wang's (1976) study, synthetic tokens of level versus rising toned morphemes were presented to Chinese (Mandarin) and English listeners for identification and discrimination. The Chinese listeners showed more distinct discrimination boundaries than the English listeners, leading Wang to the conclusion that such differences in perception between the two groups of listeners were related to language-specific experiences. Similar results were obtained by Chan, Chuang & Wang (1975). In their study, two (untrained) Chinese listeners and three American English control listeners heard a tone contour continuum on the vowel /i/ with the final pitch fixed at 135Hz and initial pitch varying from 105 to 135Hz at 3Hz intervals (11 contours in total), intended for a transition from the Chinese words /i^{35}/ 'auntie' to /i^{55}/ 'clothes'. The Chinese listeners' identification curve had a sharp category boundary between stimuli 7 (123-135Hz) and 8 (126-135Hz). They also had a peak in a 2-step ABX discrimination task between the same stimuli. The American English listeners showed delayed boundary (at stimulus 9) in the identification test. Their discrimination peak did not match this boundary but was further delayed (between stimuli 9 and 11), suggesting that they were operating on the psychophysical properties of the stimuli.

Different researchers try to account for such language-specific patterns in different ways. For Steriade (2001), the universal P-map does not change for speakers of a particular language. Rather, language-specific patterns arise from different constraint rankings. There is no doubt that general auditory capacities do not differ much among listeners with normal hearing ability, no matter how diverse their linguistic backgrounds are. It is thus reasonable to assume that there exist universal patterns in perception of both speech and nonspeech auditory stimuli by speakers of different languages. But leaving language-specificity to different constraint rankings does not offer an adequate theoretical explanation for the phenomenon, as rankings derived from empirical data only describe the patterns but do not reveal the mechanisms underlying the patterns. Some other force(s) must be assumed to work along with the P-map in determining language-specific patterns.

It might be possible to account for much of the language-specific effects found in previous research on perception by noting that listeners of different languages have different phonetic categories. In tasks that specifically ask listeners to use categorical knowledge (i.e. to tap long-term memory representations), it should be no surprise to find language-specific response patterns. This could be true even if the underlying sensory perceptual map is exactly the same for each listener, regardless of linguistic experience. For example, Carney et al. (1977) contend that categorical perception may co-exist with general psychophysical perception under certain experimental conditions. Patterns not conforming to categorical perception in their discrimination data led Carney et al. (1977) to assert that a distinction should be made between attentional factors in perception and the perceptual capacities of human auditory organisms. Thus, language-specific effects such as those revealed in categorical perception may have resulted from different processing modes, namely a "general auditory" mode and a "phonetic" mode (see also Pisoni, 1973; Johnson, 1988). As a result, none of the data regarding the influence of phonology on perception negates that there is a universal perceptual map.

This is why it is important to test the hypothesis of the existence of an auditory map in the neural models advanced by Guenther and colleagues (Guenther & Gjaja, 1996; Guenther, Husain, Cohen, & Shinn-Cunningham, 1999; and Guenther & Bohland, 2002; Guenther, Nieto-Castanon, Ghosh, & Tourville, 2004) and Bauer, Der & Herrmann (1996). A central component in these neural models of perception is an auditory cortical map, whose formation is influenced by stimulus input and type of training. In particular, Guenther et al. (1999) found that categorical training in psychophysical experiments using nonspeech-like bandpass-filtered acoustic noise in different frequency ranges led to smaller cortical representation of – hence, decreased sensitivity to – stimuli in the training range, while discrimination training led to larger cortical representation – hence, increased sensitivity – in the training range. A recent perceptual study by McGuire (2004, ms) also replicated part of Guenther et al.'s (1999) findings with natural speech stimuli. Functional magnet resonance imaging (fMRI) studies by Guenther & Bohland (2002) and Guenther et al. (2004) provided further supporting evidence for this assertion. Greater temporal lobe activation was recorded when subjects heard non-prototypical examples of

American English /i/ than when they heard prototypical examples of /i/ (Guenther & Bohland, 2002; Guenther et al. 2004). Prototypical examples of /i/ can be seen as stimuli from the "training range", except that the training was not done under laboratory condition but was rather a listener's lifetime experience with his/her native language.

The effect of categorization training in Guenther et al.'s (1999) experiments conforms to the *perceptual magnet effect* (Kuhl, 1991; Kuhl, et al. 1992). Given the fact that humans, including infants, try to form sound categories with reference to a good exemplar, our language acquisition experience is not different from the categorical training received by the participants in Guenther et al.'s (1999) experiment, only a whole lot longer in time. Guenther et al.'s (1999) results are also consistent with research findings that demonstrate the *categorical perception* of speech (Liberman, et al., 1957; Abramson & Lisker, 1970; Beddor & Strange, 1982), as the results from categorical perception studies are exactly what Guenther and colleagues' model would predict: lower discriminability for within-category stimulus tokens – corresponding to the frequencies within the categorization training range in Guenther et al. (1999). The reason why this model accounts for both categorical perception and the perceptual magnet effect equally well is probably that the two phenomena are one and the same: both resulted from "categorization training"; and as a result, both show poorer within-category discrimination (see also Lotto, Kluender & Holt, 1998).

If an auditory warping similar to what Guenther et al.'s (1999) model describes existed, it would certainly serve the linguistic purpose well, as the warping directs neural activities to distinguishing between-category differences and to ignoring irrelevant within-category differences. It would also enable a unified neural model for speech modalities and other sensory and motor modalities (Guenther & Gjaja, 1996). Empirically speaking, one should still have the ability to perceive non-native contrasts and form novel categories in natural language situations; otherwise, adult acquisition of L2 would be more difficult than it is. Thus, auditory warping has to be reversible and the neural map re-arrangeable. This seems to be implicitly allowed in the model, at least under experimental conditions, in Guenther et al. (1999). Under other experimental conditions, certain low memory demand tasks may not produce patterns conforming to patterns of the warped auditory map (Pisoni, 1973; Carney et al. 1977; Werker & Tees, 1984; Goldstone, 1998). Indeed, Guenther et al. (1999) did not find the perceptual magnet effect using experimental procedures intended to promote a sensory-trace auditory processing mode, rather than a context coding mode (e.g. Macmillan, 1987), especially a short inter-stimulus interval (ISI) of 250ms and an AX discrimination task. Wang (1976) also found that extensive practice may shift categorical boundaries of tone-language speaking listeners closer to those of the non-tone language speaking listeners. Guenther et al. (1999) offered no explicit account for these different task-dependent patterns. We can only infer from Guenther & Bohland's (2002) and Guenther et al.'s (2004) fMRI studies that these task-dependent patterns in experimental data not conforming to categorical perception or the perceptual magnet effect may be due to different degrees of activation in different auditory cortical areas in the temporal lobe and supratemporal plane and may be attributed to attentional factors.

On the other hand, Johnson's (2004) lexical distance model, although not explicitly denying the existence of auditory warping, tries to separate the effect of such warping from that of the lexicon. In the lexical distance model, incoming signals are compared with phonetically detailed forms in the lexicon directly. Consequently, the universal perceptual distances assumed need not be altered by linguistic experience to account for language-specific perceptual effects, which simply emerge from the lexicon. The model computes overall perceptual distance (d) from two sources, i.e. inherent auditory/perceptual similarities between two stimuli (d_a), as well as aggregated average difference in lexical activations by the two stimuli (d_l, computed as the difference in the amounts of activation of the lexicon caused by these stimuli, with a constant k gating the influence of this lexical distance on perception under different experimental conditions); or $d = d_a + k \times d_l$. Because the way the overall perceptual distance is computed, it is claimed that the model has the ability to distinguish discrimination performance from categorization performance, the former of which can be found in a minimal uncertainty task of limited stimulus set or speeded AX discrimination with a short ISI (no lexical access, perceptual distance computed almost exclusively from auditory distance) and the latter of which in tasks involving higher memory load such as AXB identification, difference rating (lexical forms consulted for similarity judgments). Johnson's (2004) fricative perception data from a rating task as well as a speeded AX discrimination task by Dutch and American English listeners seem to support this claim.

1.4 Theme of this Dissertation

With a series of tone perception experiments on Chinese tones using native and non-native listeners, this dissertation tests the hypothesis that tonology influences tone perception. Implicitly assumed in this hypothesis is that there exist language-specific effects on perception but not to the exclusion of universal constraints on speech perception such as those proposed in Steriade's (2001) work. That is, phonology may exaggerate or weaken certain perceptual contrasts through language-specific auditory warping but may not alter the universal patterns in a fundamental way. My goal is to try to find out: i) whether there is a perceptual effect due to presence of tonal neutralization rules in the listener's native language; ii) whether tonal perception by listeners of different language background involves different processing strategies, or maybe different processing levels; iii) if different processing levels are involved, whether certain experimental procedures will force listeners to switch from one level to another; and iv) if there is a neutralization effect, whether it will be carried over to a different set of stimuli (from a different tone language/dialect).

Furthermore, I hope to find out whether empirical data from this series of tone perception experiments may fit a speech perception model of auditory cortical warping as suggested in Guenther and colleagues' work (Guenther & Gjaja, 1996; Guenther, et al., 1999; Guenther & Bohland, 2002; Bauer et al., 1996) or one of lexical distance (Johnson, 2004). Neither the neural model nor the lexical distance model explicitly discusses the issue of how neutralization rules may affect discrimination of two contrastive sounds (or tones) that are neutralized in certain phonetic environments. But within Guenther et al.'s neural model, we may imagine a "noisy" training condition under which stimuli categorized into an abstract representation of A (e.g. [x] into /x/ in Greek) may sometimes have to be categorized as A or B (e.g. [x] to /x/ or /k/ before [t] in Greek due to the /kt/ → [xt] neutralization rule mentioned above). As a result of this double-identity status of certain speech sounds, the contrast between the relevant sound categories may be weakened. Within Johnson's lexical distance model, because of the cross-representation of two sounds (e.g., /k/ and /x/ in Greek), a [x] or a [k] input may activate lexical items containing either /x/ or /k/. Consequently, the difference in lexical activation, i.e. the lexical distance, between /x/ and /k/ is predicted to be smaller than if there is no such neutralization rule. If these inferences within the two models about the effect of segmental neutralization rules are correct, we should also be able to extend them to tonal neutralization rules such as the T214 sandhi in Putonghua (or Standard Mandarin; see Chapter 2 for discussion of the sandhi rule), where T35 and T214 substitute /x/ and /k/ in the segmental example from Greek, respectively, and are predicted to have a shortened perceptual distance by both models.

1.5 Outline of the Dissertation

The main body of this dissertation is arranged as follows. Chapter 2 reports results from an AX discrimination task with an inter-stimulus interval (ISI) of 300ms and natural speech Putonghua (Standard Beijing Mandarin) tones, where the T214 sandhi rule (/T214.T214/ → [T35.T214]) was found to contribute to the warping of the Chinese listeners' tone space, leading to longer reaction time (RT) for tone pairs T35-T214 and T214-T35 – hence shorter perceptual distance between the two tones as revealed in the tone space obtained through multidimensional scaling (MDS) using the method of individual differences (INSCAL, Carroll & Chang, 1970).

Chapter 3 compares the results from three experiments, namely a low uncertainty AX discrimination task with synthetic (sinewave) Putonghua tones and an ISI of 100ms (Experiment 2), a low uncertainty AX discrimination task with natural speech monosyllabic Putonghua tones and an ISI of 100ms (Experiment 3), and an AX difference rating task with natural speech monosyllabic Putonghua tones and an ISI of 100ms (Experiment 4). The AX discrimination task using sinewave tones (Experiment 2) reveals mostly psychoacoustic effects and can thus help tease apart the baseline perceptual effect, common in both Chinese and American English (AE) listeners' tone perception due to raw acoustic similarities in the stimuli, from language-specific effects caused by differences in the listeners' native phonology. As will be obvious from the comparisons of the results, data from the AX difference rating task (Experiment 4) show strong language-specific effects due to the T214 and T35 sandhi rules operating in Beijing. These effects also manifest themselves in the low uncertainty AX using natural speech stimuli (Experiment 3), albeit to a lesser degree. Differences in the results obtained in Experiment 1 and the three experiments reported in Chapter 3 are also discussed. A slightly longer ISI of 300ms in Experiment 1 (as opposed to 100ms in Experiments 2, 3 and 4) and a roving presentation of natural speech stimuli induced more linguistic effects than did the low uncertainty task used in Experiments 2 and 3. It also seems that the T35 rule, not taught to second language (L2) learners of Putonghua, is not present in the speech of the Chinese listeners who participated in Experiment 1 and whose first dialect is not Beijing Mandarin.

Chapter 4 introduces two more Mandarin dialects, namely the Jianghuai Mandarin dialect of Rugao (Jiangsu Province; four tones, no neutralization rules of the T214 sort) and the Northern Mandarin dialect of Yantai (Shandong Province; quiet a few neutralization rules). A brief sketch of the synchronic tone system and tone sandhis as well as historical tone development is provided for both dialects. Cross-dialectal tone category correspondences are also outlined for Putonghua (Beijing), Rugao and Yantai.

Chapter 5 tests for order effect in Experiment BJ (using an AX discrimination task, an ISI of 100ms, speeded response and natural speech Beijing/Putonghua tone stimuli), Experiment RG (using an AX discrimination task, an ISI of 100ms, speeded response and natural speech Rugao tone stimuli) and Experiment YT (using an AX discrimination task, an ISI of 100ms, speeded response and natural speech Yantai tone stimuli). These experiments were run with each of AE, Beijing, Rugao (young and older) as well as Yantai (young and older) listeners participating in all three of them in a Latin Square fashion within the same hour. That is, a listener may participate in the three experiments in one of three orders: Beijing-Rugao-Yantai, Yantai-Beijing-Rugao, and Rugao-Yantai-Beijing. No large scale order effect was found. But error data from these experiments turned out to be very interesting and show patterns consistent with those found in the RT data (Chapter 6).

Chapter 6 reports analytic results for Experiments BJ, RG and YT. Language-specific effects were found across language/dialectal groups as well as across age groups.

In particular, the AE listeners again showed psychoacoustic listening. Yantai tone sandhi effects were revealed in the Yantai older listeners' RT (and error data). Perception of Putonghua tones by the older listeners from Rugao and Yantai in Experiment BJ was also affected by the cross-dialectal tone category correspondences between Putonghua and their respective native dialect. The Yantai young listeners with a higher degree of L2 proficiency in Putonghua showed perceptual patterns more similar to Beijing listeners.

Finally, Chapter 7 summarizes the main findings from the seven experiments and evaluates the neural model of auditory cortical map (Guenther & Gjaja, 1996; Guenther et al. 1999) and the lexical distance model (Johnson, 2004) against the findings reported in this dissertation. The results seem to support the proposal of an auditory cortical map, which establishes a neurophysiological basis for the phenomenon. But the model cannot account for the different degrees of language-specific effects found in the various experimental tasks in our experiment. On the other hand, the lexical distance model (Johnson, 2004) correctly predicts that simple experimental task such as AX discrimination with a limited stimulus set (Experiments 2 and 3) induces no lexical activation – hence, no language-specific effects. But our results indicate that this model may also need some refinement to allow more lexical interference in tasks such as AX discrimination (roving) assumed to tap mostly auditory processing of the stimuli.

CHAPTER 2

PERCEPTION OF MANDARIN TONES BY CHINESE- AND ENGLISH-SPEAKING LISTENERS[1]

This chapter reports on an experiment that tested the hypothesis that native phonology may influence speech perception using natural speech tokens of Standard Mainland Mandarin (or Putonghua) tones.

2.1 Tones and tone sandhis in Chinese Putonghua

Putonghua, or Standard (Mainland) Beijing Mandarin, has four lexical tones. Chao (1935, 1965, 1968) describes them as high level [55, ˥], mid-rising [35, ˧˥], low falling-rising [214, ˨˩˦][2], and high falling [51, ˥˩]. The numbers in the square brackets indicate the idealized pitch values of these tones on a five-level scale. The drawings next to the numbers are graphic time-pitch representations of the tones, termed Chao's tone letters (for a detailed discussion of the tone transcription system, see Chao 1930). Based on my small recorded database of ten (10) Beijing speakers, the tones are more likely /44, 24, 212, 51/. The contours are nevertheless similar to Chao's description. The tones are

[1] This chapter is a reanalysis of Huang (2001). The text also includes fragments from that paper.

[2] In Cheng (1973), Tone 3 is described as having the value [315]. In my own recorded database of Beijing speakers, the final rise never gets very high even in prepausal positions.

usually referred to as Tones 1, 2, 3, and 4, respectively, in the Chinese linguistic tradition. For notational purposes in this dissertation, I shall refer to them as T55, T35, T214, and T51, as these are well-established labels. These notations will also refer to syllables bearing the respective tones where segmental makeup is not crucial to discussion. I shall transcribe tones using raised numbers along with segmental makeup (when necessary), as in /li^{51}.ru^{35}/ 'for example'[3].

(2.1) The Four Tones in Chinese Putonghua

Tone 1	high level	/55, ˥/
Tone 2	mid-rising	/35, ˧˥/
Tone 3	low(-dipping)	/214, ˨˩˦/
Tone 4	high falling	/51, ˥˩/

Sometimes, people also talk about a "fifth" tone, namely the neutral tone, whose pitch varies dependent on its preceding tone. There are a few monosyllabic morphemes that are probably inherently neutral-toned, or toneless, such as the tense-aspect marker /le/. In my transcription, a raised "0", as in /hao^{214}.le^{0}/ 'it's done' and [jie^{214}.jie^{0}] 'sister' (underlyingly, /jie^{214}.jie^{214}/), denotes an inherent or surface neutral tone.

Just as in many other Chinese dialects, underlying full tones may be modified under the influence of their tonal environment in Putonghua (see, e.g., Chao 1965, 1968; Kratochvil, 1968; Cheng, 1973; Chen, 2000; Duanmu, 2000). This phenomenon is known as tone sandhi.

[3] I shall adopt Hanyu Pinyin, the official Chinese Romanization system, as my transcription system, unless a situation arises where it is necessary to use the International Phonetic Alphabet (IPA, 1999). The dot [.] separates two syllables.

As described by Chao (1965, 1968), in the third-tone sandhi, T214 of Putonghua becomes T35 when immediately followed by another T214.

(2.2) The T214 Sandhi Rule

/T214.T214/ → [T35.T214]

It is claimed by many that morphological and syntactic boundaries are irrelevant here. (But see Shen 1994, who suggests that syntactic structure works with speech rhythm/prosody in determining whether the sandhi rule gets applied or not. But he also allows ambiguous cases where the sandhi may or may not apply.) Some examples are provided in (2.3) below:

(2.3) Examples of the T214 sandhi

a. /hao^{214}.mi^{214}/ → [hao^{35}.mi^{214}] 'good rice'
| |
modifier head noun

(Cf. / hao^{35}.mi^{214}/ → [hao^{35}.mi^{214}] 'milimeter')

b. /mi^{214}.hao^{214}/ → [mi^{35}.hao^{214}] 'The rice is good.'
| |
subject predicate

c. /mai^{214}.mi^{214}/ → [mai^{35}.mi^{214}] 'to buy rice'
| |
verb object

Since an underlying /T35.T214/ sequence is also realized as [T35.T214], the paradigmatic contrast between T35 and T214 is lost before a following T214, creating many homophonous surface pairs. Thus, /hao^{214}.mi^{214}/ 'good rice' is not distinguishable from /hao^{35}.mi^{214}/ 'millimeter', as both surface as [hao^{35}.mi^{214}]. Nor is /fen^{214}.chang^{214}/ 'flour factory' from /fen^{35}.chang^{214}/ 'graveyard'. (See Chao 1965, 1968 and Cheng 1973 for more examples of this sort.)

There has been some debate on whether the neutralization of T214 and T35 is complete or not. I shall follow Chao in assuming that it is, as there is sufficient experimental evidence for such a view. In an identification task using 130 disyllabic tonal minimal pairs involving T35 and T214 before another T214 (e.g. /qi^{35}.ma^{214}/ "to ride a horse' and /qi^{214}.ma^{214}/ 'at least'), Wang & Li (1967) found that their 14 Chinese listeners' percentage correct responses were below chance (ranging from 49% to 54%). They concluded that the contrast between T35 and T214 was neutralized in this environment.

With a forced-choice identification test between /T35-T214/ and /T214-T214/ for disyllabic sequences, Peng (1996) found that the overall F0 of sandhi tone (178.33Hz) was only slightly lower than the underlying T35 (180.66Hz) in the production and that the derived [T35] was perceptually indistinguishable to Mandarin speakers. Although she used Taiwan Mandarin speakers, the finding can probably be extended to Mainland Mandarin (or Putonghua). From my own experience with Mainland speakers, people sometimes have to actually point out explicitly, as a caveat to their conversation, that they are using a T214 morpheme (before another T214) not a T35 morpheme if the context does not disambiguate the two, as in /wo^{214} shuo^{55}-de^{0} zhe^{51}-wei^{51} lao^{214}-shi^{55} xing^{51} [Xu^{214}], bu^{35} xing^{51} [Xu^{35}]/ "The surname of the teacher I'm talking about is [Xu^{214}], not [Xu^{35}]." The underlying form of the teacher's surname has to be supplied because before the title /Lao^{214}-shi^{55}/ "Teacher", both /Xu^{214}/ and /Xu^{35}/ surface as [Xu^{35}]. Interestingly, Peng (1996) also found in a concept formation experiment that most Mandarin listeners categorized derived [T35] as /T214/ rather than /T35/, whether or not the derived [T35] form has an underlying /T35/ homophone (e.g., the surname /Xu^{214}/ after undergoing the sandhi process in the above example becomes homophonous with underlying /Xu^{35}/, whereas the morpheme /$guan^{214}$/ in the word /$guan^{214}.li^{214}$/ "to manage" does not have a homophonous underlying */$guan^{35}$/). When the listeners were trained to categorize the sandhi [T35] forms as underlying /T35/ forms, they made quite a few mistakes, especially in the first half of the perceptual test following the categorization training. Peng speculates that this may have to do with the influence from the transcription system taught to them at school: despite the obligatory application of the T214 sandhi rule, morphemes with an underlying T214 tone was transcribed as T214 (instead of the sandhi tone T35).

Motivation for T214 sandhi might be attributed to ease of articulation, given the phonetic shape of the citation contour. That is, a low falling-rising-falling-rising contour formed by two consecutive underlying T214 syllables would be complicated to pronounce and is thus simplified to rising-falling-rising in a dissimilation process (Mei, 1977). After deletion of the initial fall of the first T214, the remaining low-rising is re-categorized with the existent mid-rising T35. Cheng (1968) investigated the application of the T214 sandhi rule in code-mixed speech of Chinese-English bilinguals. He found that the sandhi rule applied, especially in fast speech, in phrases such as /hao^{214} pro'fessor/ 'good professor' but not in /hao^{214} 'student/ 'good student'.[4] That is, when a T214 Chinese morpheme precedes an English word with an initial reduced stress

---

[4] This is a simplified account of Cheng's findings. In that study, Cheng differentiated four levels of stress in English, namely primary, secondary, tertiary and reduced. Sandhi was only found to occur before reduced stress. It happened more often before English words with fewer syllables.

(perceived as having a low pitch by the Chinese listeners), the sandhi occurs. This suggests that it is the feature "low" (or [-high] in Cheng's analysis) in T214 that triggers the sandhi application. In Huang (2001), I speculated that such a change might also be perceptually conditioned. That is, T214 changes to T35 in this environment because these two tones are confusable in the first place. But more research is needed to trace the true origin of the sandhi, which had been documented as early as the 16th century (Mei, 1977; Chan 1985) and which might even existed in the 14th century (Mei, 1977; Zhou, 1324 ?).

Traditional analyses state that other phonetic variants of T214 in normal stress positions include [21] and [214], the first of which appears before all full tones except T214 and the second of which appears in prepausal positions, especially sentence-final position (Chao 1965, 1968). Some researchers treat [21] as the underlying form of T214, as this is the most common surface shape. In fact, in the variety of Mandarin spoken in Taiwan, [21] surfaces in prepausal positions as well. Sometimes, it may even surface in prepausal positions in Putonghua.

Chao (1965, 1968) also discusses another sandhi rule, where T35 becomes T55 when following a T55 or T35 and preceding a full-toned syllable. Chao considers this rule to be "of minor importance" (1965: 35).

(2.4) The T35 Sandhi Rule

/T55.T35.Tx/ → [T55.T55.Tx],

or

/T35.T35.Tx/ → [T35.T55.Tx],

where Tx is any non-neutral tone. According to Chao, the middle position of a three-syllable phrase is relatively weak.[5] As a result, the underlying low pitch onset of the sandhi affected T35 gets deleted and the pitch contour is simplified to [55]. Note that, interestingly, the affected T35 does not have to be an underlying /T35/. For instance, a sequence of /T214.T214.T214/ may first be affected by the T214 sandhi rule (applied twice linearly from left to right), supposedly resulting in an invisible middle stage of (T35.T35.T214), which is then further affected by the T35 sandhi rule, yielding the final surface sequence [T35.T55.T214]. Two familiar examples are: /cong^{55}.you^{35}.bing^{214}/ $\rightarrow$ [cong^{55}.you^{55}.bing^{214}] '(Chinese) onion oil pancakes', and /hao^{214}.ji^{214}.zhong^{214}/ $\rightarrow$ [hao^{35}.ji^{55}.zhong^{214}] 'quite a few kinds' Chao (1965: 36). Thus, this sandhi also leads to paradigmatic tonal neutralization: the contrast between T55 and T35 is lost.

Yet another sandhi rule exists in Putonghua, which changes morphemes such as /yi^{55}, qi^{55}, ba^{55}/ 'one, seven, eight' to [yi^{35}, qi^{35}, ba^{35}] before a T51 syllable/morpheme. To my knowledge, no one offered an explanation for the cause of the sandhi. It seems to me that such a change may be auditorily/perceptually based. It is well known that the pitch onset of T51 is higher than that of T55. In my small recorded database of ten (10) Beijing speakers, this difference between the two high pitch targets ranges from 10 to 30 Hz, which is greater than the negligible pitch change found in Wang (1976). Thus, in a [T55.T51] sequence, there will be a perceivable rise during the transition from the first syllable to the second (Lin & Wang 1985). As the numerals co-occur frequently with the Chinese generic classifier /ge^{51}/, phonologization of such a rise at a certain historical

[5] Margie Chan reminds me that examples such as /kan^{51}.dian^{51}.shi^{51}/ $\rightarrow$ [kan^{51}.dian^{35}.shi^{51}] 'to watch television', observable in older native Beijing speakers' speech, may provide supporting evidence for such an analysis.

point seems very likely. Although taught as a sandhi rule to L2 Putonghua speakers, unlike the T214 sandhi, it is not always obligatory, especially for numerals /qi^{55}, ba^{55}/ 'seven, eight'.

There is another special rule that changes the tone of the negation word /bu^{51}/ to [T35] before another /T51/ morpheme, as in /bu^{51}.dui^{51}/ → [bu^{35}.dui^{51}] 'not correct'. This is an obligatory rule. Note that the rule only applies to this morpheme and that a homophonous word such as /bu^{51}.dui^{51}/ 'troops' is realized as [bu^{51}.dui^{51}]. Thus, this and the rules involving the numerals mentioned above may also be seen as morphologically conditioned processes.

It is also well-known that in non-final positions there exist low level "phonetic" tonal realization rules such as deletion of a high (H) target at the end of T35 as in /$hong^{35}$.hua^{55}/ 'red flower' and /$hong^{35}$.dan^{51}/ '(good luck) red egg', as well as undershooting of a low (L) target at the end of T51 as in /bu^{51}.dui^{51}/ 'troops' and /bu^{51}.$bing^{55}$/ 'infantry troops', especially before tones T55 and T51 (see, e.g., Shih, 1988; Xu, 1997, 2001).

As can be seen from the descriptions above, Putonghua has some rather unique tonal processes. The experiment to be reported in this chapter was design to find out whether such processes, especially the T214 sandhi rule, has an impact on tone perception by native Chinese listeners.

2.2 Participants in Experiment 1

Ten Chinese listeners (6 female, 4 male, average age 27.9) and thirteen American English listeners (7 female, 6 male, average age 21.8) were recruited from the Columbus campus of the Ohio State University (OSU). The Chinese listeners were graduate students (or their spouses) at OSU. Although a couple of them are not from the geographical regions where Mandarin is spoken, they were all fluent in the standard language due to their educational background: they all received at least college education, and Putonghua is usually the language of instruction in most classrooms in mainland China. The English listeners were undergraduate students taking an introductory linguistics course at OSU. They were all native speakers of Ohio English. The Chinese were paid for their participation in the experiment, whereas the Americans received course credits.

The American English listeners were included as a control group to see (i) if T214 and T35 of Chinese Putonghua in the sandhi environment share some property that makes them confusable for non-native listeners as well, (ii) whether the sandhi rule affects Chinese listeners' tonal perception, and (iii) whether the two groups of listeners perceive tones in a similar way. It is assumed that, if there is no effect of the listener's native phonology on perception, phonetic universality and human auditory capability should allow everybody to act the same.

Previous studies have shown that it is feasible to include "non-native" listeners. Although English is a non-tonal language, its stress-based prosodic system does utilize pitch as one way to distinguish stress accents, which may be realized as high, low, rising or falling contours and which can thus be very similar to the Mandarin tone contours, although these are not lexical (see, e.g., Beckman, 1984). In addition, if Werker & Tees (1984) are right in that ontogenetic decline that results in reduced ability to discriminate non-native sounds involves change in processing strategy rather than sensory-neural loss, then utilization of pitch contrasts in the pitch accent system and in intonation may have given the English listeners enough experience to maintain their inherent ability to distinguish pitch differences in Chinese. They may not know what to call the stimulus tones, but a paired comparison task does not require that listeners have names/labels for the items being compared. Indeed, Kiriloff (1969) found that, when asked to ignore the segmental element of the syllable and focus on the tones, non-native speakers' performance was quite good (an average of 17.5 correct identifications out of 20 stimuli, or 87.5%). Gottfried & Suiter (1997) did find some degree of performance difference in native versus non-native listeners in a more difficult identification of edited partial syllables (namely "initial only", "center only", "silent center" and "final only"). Lee et al. (1996) also found a small native speaker advantage.

On the other hand, without lexical tone categories in their lexicon, English listeners may have a psychoacoustic advantage, i.e. they may be able to detect subtle pitch differences, which may be missed by the Chinese listeners. Wang (1976) found that Mandarin Chinese speaking listeners perceived synthesized stimuli along a level to rising contour continuum categorically, dismissing rises smaller than 9 Hz as negligible within-category variations. Similarly, Stagray & Downs (1993) reported that Mandarin speaking listeners had significantly larger difference limens for frequency (DLFs) than English-speaking listeners around 1000Hz as well as around 125Hz, the latter of which approximates pitch utilized in natural speech. The Mandarin listeners also had poorly-shaped psychometric functions close to chance response level, as opposed to nice ogive-shaped functions in the English listeners' data. Stagray & Downs concluded that the Mandarin listeners had poorer differential sensitivity for frequency because frequency variations in their stimulus tones were perceived "as being within the same pitch range of a learned, level tone-phoneme category" (1993: 156).

2.3 Stimuli

The stimuli used in this experiment were constructed from recordings produced by a female Putonghua speaker in disyllabic nonsense sequences with 15 tonal combinations of any two tones of /T55, T35, T214, T51/ – that is, all possible pairs except T214-T214 which does not occur in natural speech. The segmental makeup of these recorded sequences was kept constant as /bao.fang/.[6] The typical stress pattern for a disyllabic full-toned sequence was used to get the appropriate pitch contours of the four tones in the environment where T214 sandhi occurs.[7] Ten (10) randomized lists of these sequences were recorded. The original recordings were done in a sound-proof booth in the phonetics laboratory at the OSU Linguistics Department. The speaker read from the afore-mentioned 10 randomized lists and was recorded with a head-mounted microphone (Shure SM10A model) and a DAT recorder.

The recordings were transferred to a UNIX workstation at 22,050Hz with 16-bit samples and edited with Xwaves (Entropic Research Lab). The first syllable (i.e. /bao/) was cut from each of these sequences. The seven (7) best productions of the syllables for each of the four tones (as determined subjectively by the author) were chosen to splice the test stimulus pairs, while three (3) other productions were used in the training session.

---

[6] I shall use hyphen [-] between an ordered pair of monosyllables, and period [.] between two syllables produced as a sequence/word. A notation with a back slash [/] between two monosyllables, as in T55/T35, covers both T55-T35 and T35-T55. The hyphen [-] may be omitted. Thus, T55T35 = T55-T35.

[7] Yip (1980) and Zhang (1988) mention that the T214 sandhi is conditioned by the metrical pattern of the utterance and that the T214 that undergoes the sandhi has to be in the weak branch of the stress matrix, i.e. the syllable bearing the sandhi tone must not be linked to a node at the highest/primary stress level. Thus, it is predicted that the first T214 in /xiao3jie3/ 'miss' (with a weak-strong pattern) would undergo the T214 sandhi and surface as [xiao^{35}.jie^{0}], whereas that in /jie^{214}.jie^{214}/ 'older sister' (with a strong-weak pattern) would not, yielding the surface form [jie^{21}jie^{0}]. But see Shih (1997), where she holds that stress does not play a role in the sandhi processes. We chose not to commit ourselves to any particular phonological framework here and tried to take into consideration all possible conditions for this sandhi process.

Figures 2.1 shows pitch tracks of these stimulus tones. Note that only the first "half" of the T214 tonal contour is realized, which is typical of T214 in this non-final position.

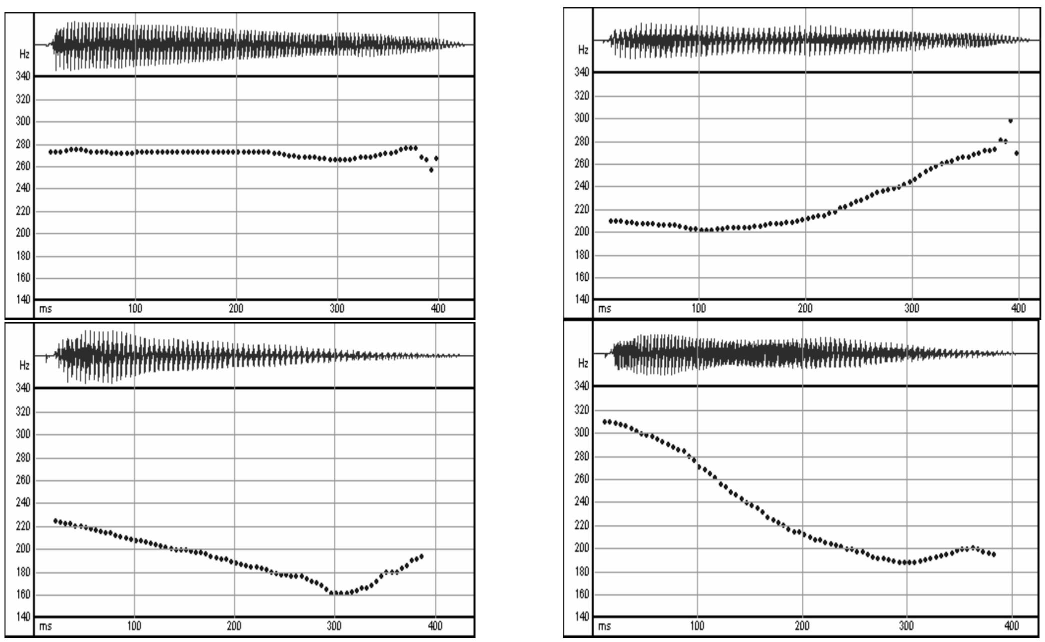


Figure 2.1 Pitch tracks of the stimulus tones T55 (upper left panel), T35 (upper right), T214 (lower left), and T51 (lower right).

The test session consisted of 7 sections, each of which contained 20 stimulus pairs. Thus, all participants listened to 20 × 7 = 140 pairs of the form /bao-bao/. The 20 pairs in each section included 12 different pairs (see the checked boxes, marked with x, in Table 1 below) and 8 identical pairs (i.e., each of the 4 identical pairs in the empty boxes in Table 1 was repeated twice in any of the test sections). Each identical pair contains two repetitions of the same .wav sound file. Only the results of different pairs were analyzed. The identical pairs were included as fillers.

	T55	T35	T214	T51
T55		1	3	5
T35	2		7	9
T214	4	8		11
T51	6	10	12	

Table 2.1 Tonal combinations to be tested. Test pairs are the numbered ones.

2.4 Method

An AX discrimination task was used. Participants were tested in front of a computer one at a time in a sound-proof booth. The stimuli were presented to them through headphones, using the Micro Experimental Lab (MEL) program installed on a PC. While each stimulus pair was played (at a 300ms inter-stimulus interval and a 2000ms inter-pair/trial interval) through the headphones, the words "same" and "different" were also displayed visually on the left and right sides of the computer screen, respectively. The participant input responses by pressing the "same" or "different" buttons on a button-box connected to the PC. Participants were asked to use their left and right index fingers to press the "same" and "different" buttons, respectively. Instructions, both given orally by the experimenter during the training session and displayed visually on the PC screen during the test session, asked the participant to respond as accurately and as quickly as possible. After each correct "same"/"different" judgment was made, the reaction time (RT) would appear on the screen as feedback; otherwise, the screen would display the words "wrong response". This made it clear to the subjects what a good performance was: one with shorter reaction time and fewer errors. Both the "same-different" judgment accuracy and RT were recorded as experiment results. The measurement for RT started from the onset of the second syllable of the stimulus pair. The mean duration measurements for all stimulus syllables are: T55=375.9ms, T35=414ms, T214=389.5ms, and T51=387.8ms. Such differences do not seem to be big enough to affect the RT measurements, as we shall see that tone pairs with T35 as the second tone do not always have longer RTs than pairs with any other tone as the second stimulus tone.

It is predicted that, if T35 and T214 are more confusable, i.e. closer to each other in the perceptual space, then (i) people would make more mistakes when asked to tell whether they are the same or different, and (ii) people would take longer to make the judgment, that is, the shorter the perceptual distance, the longer the reaction time (RT) (see, e.g. Shepard, 1978; Shepard, Kilpatric & Cunningham, 1975; Takane & Sergent, 1983; Nosofsky 1992).

2.5 Results and Analysis

The results show that T35 and T214 are indeed perceptually more confusable than any other tone pairs. In terms of the mistakes that listeners made, there was no statistically significant difference between the tone pairs, as error rates were very low in the responses of both the Chinese and English groups. But the pairs T35-T214 and T214-

T35 did attract more errors than other pairs. Table 2.2 shows group median RT values of correct "different" responses and error rate in percentage for each non-identical tone pair. (As the distribution of the RT measurements is skewed to the right, median RT values were chosen to represent the centrality of the data instead of mean RT values.)

As can be seen from Table 2.2, the Chinese listeners scored 62 correct responses out of all 70 T35-T214 stimulus pairs (= 7 sections × 10 participants) with an error rate of 11% and 65 correct responses out of all 70 T214-T35 stimuli with an error rate of 7%, which is slightly better than the English listeners who scored 76 correct responses out of all 91 T35-T214 stimulus pairs (7 sections ×13 participants) with an error rate of 16% and 79 correct responses out of all 91 T214-T35 stimulus pairs with an error rate of 13%.

Tone Pair	Chinese	English
T55/T35	537 (5.5%)	585 (1%)
T55-T35	542 (4%)	526 (1%)
T35-T55	532 (7%)	643 (1%)
T55/T214	550 (4.5%)	538 (2%)
T55-T214	551 (3%)	513 (2%)
T214-T55	548 (6%)	563 (2%)
T55/T51	556 (4%)	593 (3.5%)
T55-T51	560 (4%)	605 (5%)
T51-T55	551 (4%)	581 (2%)
T35/T214	661 (9%)	690 (14.5%)
T35-T214	664 (11%)	722 (16%)
T214-T35	658 (7%)	657 (13%)
T35/T51	524 (2%)	584 (7%)
T35-T51	497 (0%)	615 (11%)
T51-T35	551 (4%)	552 (3%)
T214/T51	527 (2%)	593 (3.5%)
T214-T51	515 (0%)	584 (5%)
T51-T214	538 (4%)	602 (2%)

Table 2.2 Mean RTs (in milliseconds) for correct "different" responses and percentage of errors. These RT values were computed from each listener's median RT for each tone pair.

Although error rates were too low to be significant, the RT data turned out to be very informative. The graphic representation in Figure 2.2 may help us see clearly what the RT values for the T35/T214 pairs are like compared to other tone pairs. The points on the X-axis represent the non-identical pair types, and the numbers along the Y-axis show reaction time in milliseconds. The solid line represents the Chinese listeners' data, while the dotted line the English listeners'.

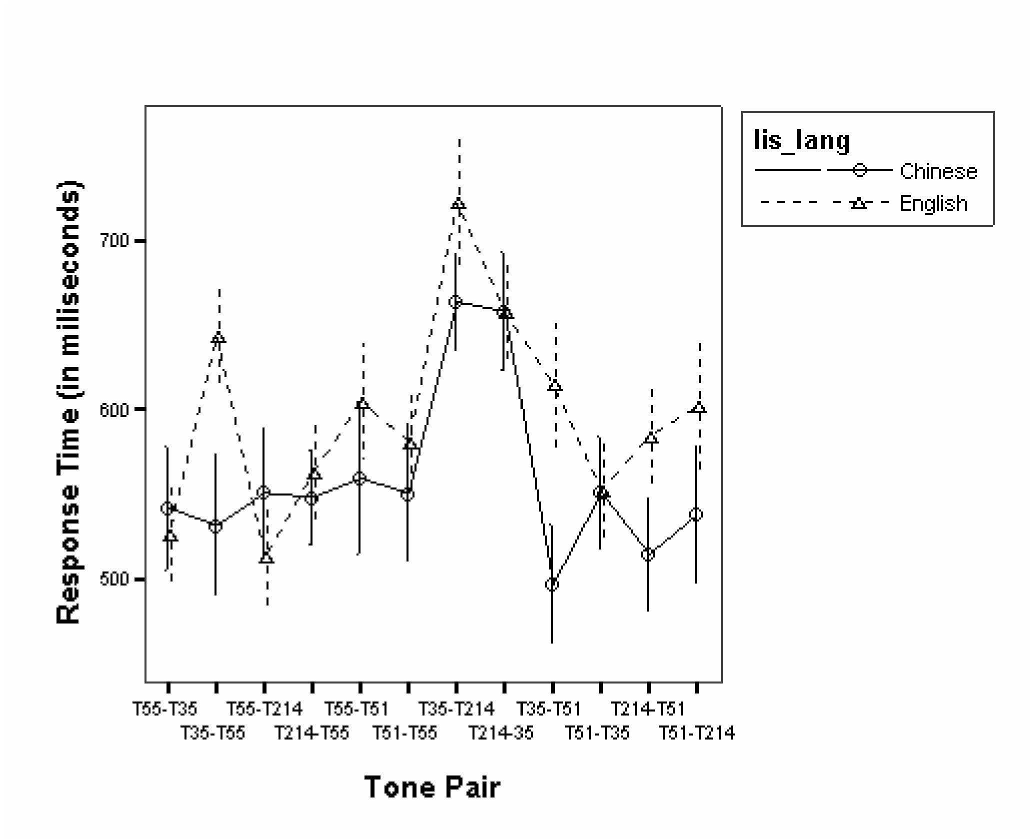


Figure 2.2 Reaction time (RT, in milliseconds) for the correct "different" tone pair responses. Although the repeated measures ANOVA did not find significant between-subject language effect, pairs T35-T55 and T35-T51 turned out to be significantly different in the T test. Error bars show one standard error. The same RT measurements are reported in Table 2.2.

In general, the Chinese listeners did better, having shorter RTs and making fewer mistakes. And as we expected, the slowest RTs were found with the T35/T214 pairs (the two peaks in the solid line in Figure 2.2). However, we also see a similar picture with the American listeners, which implies that phonetically there exists some universal perceptual similarities between these tones for both the native and non-native listeners.

2.5.1 Repeated Measures Analysis of Variance (General Linear Model)

A repeated measures analysis of variance (ANOVA) was performed on the RT data of the listeners' correct "different" responses, with all 12 non-identical tone pairs (i.e. T55/T35, T55/T214, T55/T51, T35/T214, T35/T51, and T214/T51) as the within-subject variable (12 levels), and language (i.e., Chinese and English) as the between-subject variable (2 levels). The median RT value for each non-identical pair was entered for every subject. No significant result was found between listener-language groups, [$F(1, 21) = .76$, $p = .393$]. But there was a significant effect with tone pair types, sig.[$F(7.487, 157.221) = 13.382$, $p < .001$, partial $\eta^2 = .389$]. There was also a significant effect with the interaction of language and pair, sig.[$F(7.487, 157.221) = 3.295$, $p = .002$, partial $\eta^2 = .136$].

Within-group pairwise comparison of the RT data for each language group showed that the Chinese listeners may have processed the tones differently from the English listeners. For the Chinese listeners, tone pairs T35/T214 are the most confusable and are significantly different from all other pairs ($p < .05$). For this group of listeners, pair T35-T51 is the least confusable and significantly different from all other pairs except for pairs T35-T55 and T214-T55. While the English group also found pairs T35/T214 to be the most confusable, pair T214-T35 is not significantly different from pairs T35-55, T35-T51, or T51-T214, which were also found to be very confusable. The AE listeners also found three tone pairs to be the least confusable and significantly different from most other pairs ($p < .05$), namely T55-T35, T55-T214, and T51-T35, which do not stand out in the Chinese listeners' data at all (see Figure 2.2).

2.5.2 Planned between-group comparisons: Independent Samples T Test

One other thing to note about the plot in Figure 2.2 is that, while there are similarities between the two RT curves, the RTs for pairs T35-T55 and T35-T51 seem to be quite different for the two listener groups. In fact, as reported above from the repeated measures analysis, pair T35-T51 was the least confusable pair for the Chinese listeners but one of the more confusable pairs for the English listeners. The Chinese listeners made no mistake at all when discriminating this pair, whereas the English listeners missed it 11% of the time.

Indeed, an independent samples T-test on RT data shows that these between-group differences are significant: $t = -2.136$, $p = .045$, and $\eta^2 = 0.178$ for tone pair T35-T55, and $t = -2.254$, $p =.035$, and $\eta^2 = 0.195$ for tone pair T35-T51. What is special about these tone pairs is that the pitch offset of the first tone (T35 in both cases) and the pitch onset of the second tone (T55 or T51) are very similar in pitch height. This seems to affect the English-speaking listeners' perception, but not the Chinese listeners'. In fact, the RT curve for the Chinese listeners is pretty flat, except for the T35-T214 and T214-T35 pairs, while that for the English listeners has more obvious maxima and minima, some of which are attributable to this factor (e.g., T35-T55, T55-T214, T55-T51 and T51-T214).

A possible interpretation of this difference would be that the English listeners, with no lexical tone categories in their lexicon, used the pitch onsets and offsets as phonetic cues to discriminate the tones (see also Gandour & Harshman, 1978a,b; Fox & Unkefer, 1985). The more similar these points are for a tone pair, the more confusable the pair is for them. Such is the case for tone pairs T35-T55 and T35-T51. On the other hand, the Chinese listeners, with lexical tone categories, seem to perceive the f0 contour on a monosyllable as an indivisible unit and thus ignore such phonetic details of the contour to a certain extent. This is consistent with the experimental data reported in Chan, Chuang and Wang (1975; see §1.3).

Interestingly, these different processing strategies are not always to the advantage of either group of listeners in our experiment: for tone pairs T55-T35 and T55-T214, the English listeners were able to make use the phonetic cues more efficiently and scored shorter RTs, although the difference is not significant. But the Chinese listeners were able to make good use of contour information as in pairs T35-T55 and T35-T51. This difference in strategies is actually a very telling one, because it suggests that the long RTs for the T35/T214 pairs may have resulted from different factors for these two groups of listeners. That is, the Chinese listeners were probably influenced by the T214 sandhi in their native phonology, in addition to the phonetic similarity between the tones that affected the English listeners.

2.5.3 Multidimensional Scaling (MDS): Individual Differences (weighted Euclidean distance) Model

As it may have already been evident from the discussion in the previous sections, tonal perception can be influenced by multiple factors, among which the ones that pertain to the characteristics of the stimuli. Multidimensional scaling techniques (see, e.g., Torgerson, 1952; Shepard, 1962; Carroll & Chang, 1970; Takane, Young & deLeeuw 1977), which position the stimuli in a visible perceptual space, provide a very useful method to uncover these factors. Therefore, an individual differences (weighted Euclidean distance) multidimensional scaling (INDSCAL) analysis (Carroll & Chang 1970) was also performed in order to reveal the factors affecting the Chinese and English listeners' perception of the four tones in our experiment. With this model, we can obtain a stimulus space for each group of subjects. Listeners within the same group are of course assumed to agree on the stimulus attributes important to perception, represented by different dimensions of the space, with the first dimension correlating with the tonal feature accounting for the most variance in the data and each succeeding dimension accounting for less and less variance. In our case, previous studies (Gandour & Harshman, 1978a,b; Gandour, 1981, 1983, 1984; Massaro, Cohen & Tseng, 1985; Fox & Unkefer, 1985; Lin & Repp, 1989) have revealed that such attributes as overall pitch height, starting pitch height, end pitch height, contour shape, and contour direction seem important to listeners in general. The INSCAL analysis also accounts for deviations of individual subjects away from the group space, by way of the subject weights and the "weirdness" indices.

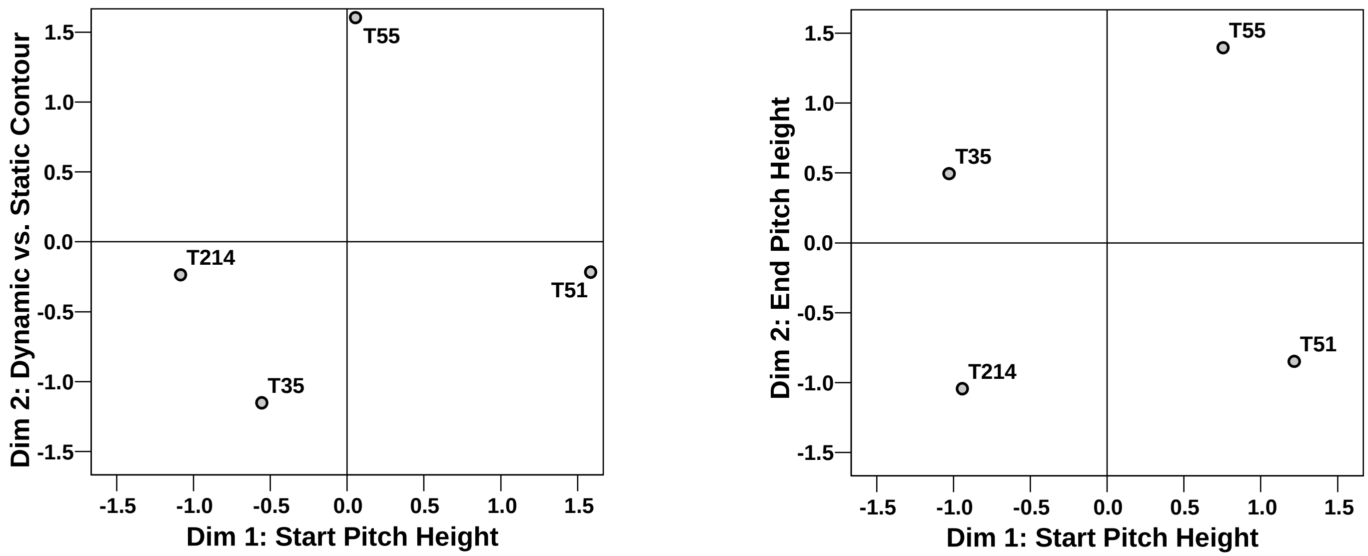
In a sense, the RT data obtained reflect degrees of similarity between the tones: RT values increase as the tones get more similar. Intuitively, perceptual distances are dissimilarities. Thus, one may assume that the closer two "objects" are in the perceptual space, the longer it takes for people to tell them apart (see, for example, Shepard et al. 1975; Shepard, 1978; Takane & Sergent, 1983; Nosofsky, 1992).

How exactly RT reflects perceptual or physical distance is still a question begging to be answered. In our case here, we would probably also need to take into account the influence of phonology on perception as well as the characteristics of the stimuli (i.e., the phonetic characteristics of the tones). Nevertheless, several approaches have been proposed to convert RTs into distances. Curtis, Paulos & Rule (1973), Shepard et al. (1975), and Shepard (1978) advocate for the reciprocal function: distance = 1/RT. Their argument for this approach is, with correct "different" judgments, reaction time values have been found to be nearly reciprocal of distance values. Takane & Sergent (1983) and Nosofsky (1992) suggest the log normal function. Takane & Sergent's (1983) reason for choosing the log normal function over the reciprocal function is that it is not the case that the RTs for correct "same" judgment are reciprocal to distances. As only the RTs of correct "different" judgments were of interest in the present study, the choice of the reciprocal approach seems to be justified. In addition, this approach is well-supported by previous research.

In fact, we did try the log normal approach and found the scaling results to be very similar to the reciprocal approach. In addition to the reciprocal and the log normal functions, which turn linearly related RTs into a non-linear distribution of distances, we also tried a linear approach suggested by Michael Broe (personal communication). RTs were rescaled using the formula (Observed RT/Observed Maximal RT) so that they now distribute along a scale of 0~1. Then, the 0~1 RT scale were turned into a 0~1 distance scale by subtracting the new "RT" values from 1 (i.e. distance = 1 – (Observed RT/Observed Maximal RT)). Again, the scaling results are surprisingly similar to the reciprocal 1/RT approach.

The median RT of each tone pair were converted into perceptual distance for each listener using the reciprocal function 1/RT, resulting in one square (asymmetrical) distance matrix for each listener. Next, using the MDS model in SPSS 12.0 for windows, we scaled the data matrices with four defined variables (i.e. the four tones), or the four points to be put on a perceptual map. These matrices were analyzed as asymmetrical, because distance between, e.g., T55-T35, may not be the same as that between T35-T55. But the directional difference is usually not significant. As a result, distances for these two pairs need to be, and can be, averaged to obtain one distance value for the two tones T55 and T35. We also analyzed the data at the measurement level "ratio", as the distance between two repetitions of the same stimulus is presumably zero (0). The scaling algorithm used is INSCAL (implemented with Young's ALSCAL program; Takane, Young & de Leeuw, 1977), because we had one matrix for each individual subject. Negative subject weights were not allowed, as a negative weight may mean that the distances are not Euclidean (Carroll & Chang, 1970). The distance values are compared "unconditionally", i.e., across matrices (Takane, Young & de Leeuw, 1977). Since we have only four tones and since higher dimensional configurations are usually hard to interpret, we chose to use two-dimensional scaling, which would help to narrow down what two characteristics about the tones affect tone perception most in our experiment.

With this scaling analysis, we uncovered one perceptual space for each group of listeners. As shown in Figure 2.3a, the Chinese listeners' space has a stress of 0.189 (Kruskal's stress formula 1 value) and 0.89 of variance accounted for (or RSQ), while the English listeners' (Figure 2.3b) has a stress level of 0.169 and RSQ of 0.91.



a. b.

Figure 2.3 The Chinese (panel a; stress = 0.189, RSQ = 0.89) and English (panel b; stress = 0.169, RSQ = 0.91) listeners' perceptual spaces of the four tones as revealed by the INSCAL analysis.

We may interpret dimension 1 in both the Chinese and the English listeners' tone spaces as "onset pitch height", as the tones on one end have higher starting pitch than the ones on the other end of the dimension, although the Chinese listeners' space seem to have been tilted a bit.

Dimension 2 in the English listeners' space corresponds rather nicely to "offset pitch height", as the tones on the lower end of dimension 2 (i.e., T214 and T51; recall that T214 is basically [21] in our stimulus set) end low while tones at the other end of the dimension (i.e., T35 and T55) end high. We may further speculate that, if it was not for the dominating similarities between T35 and T214, which pulled these tones close to each other, T35 might have been placed further up along dimension 2 and closer to T55. We note an interesting switch of position between tones T214 and T35 in the two spaces. With this switch, we cannot apply the same label to dimension 2 in the Chinese listeners' space; instead, here this dimension seems to correspond to the tonal characteristic of "static vs. dynamic contours", to use Abramson's (1962) terminology, as the level tone T55 is separated from the three contour tones.

The fact that dimension 2 corresponds to contour shape, instead of "offset pitch height", in the Chinese listeners' tone space shows again that the Chinese listeners may have employed different processing strategies and that they may have tried to predict the contour based on pitch onset information alone. This may be seen as a piece of supporting evidence for Wang's (1967) proposal of such tone features as "contour", "rising" and "falling".

In addition to the differences already pointed out above, the (relative) distance between T35 and T214 appears to be smaller for the Chinese listeners' space. This is not surprising. As we already noticed in Figure 2.2, the Chinese listeners' RT curve is fairly flat except for the T35/T214 pairs, which have the longest RTs, making them stand out among all tone pairs, whereas the English listeners' RT curve has additional identifiable maxima, especially at the RT point for pair T35-T55, reducing the magnitude of difference between the T35/T214 pairs and the other tone pairs.

One other difference that is not obvious directly from the two configurations is that the two groups of listeners attached different degrees of importance to the dimensions. The first dimension in the English listeners' space accounts for 55.01% of the RT variance (RSQ), and the second dimension accounts for 36.02% of the variance, with a ratio of 1.527 : 1. In the Chinese listeners' space, dimension 1 correlates with 47.85% of the RSQ, while dimension 2 accounts for 41.14%, with a ratio of 1.163 : 1. The subject weirdness indices show that both groups reached these ratios rather uniformly across individual subjects, with only one person in each group (highest weirdness scores are .25 and .35 for the Chinese and English groups, respectively) whose dimensional weights are slightly off from the group averages.

Recall that the Chinese group had a couple of listeners whose first languages were non-Mandarin or Southern Mandarin Chinese dialects. As it turned out, MDS seemed to be able to differentiate people from different dialectal groups: in general, these people contributed less to the group stimulus configuration, having lower weights on both dimensions. When their data matrices were analyzed individually, the spaces tend to be well-dispersed; that is, no T214 effect was found. The data of the listener contributing the least to the group MDS solution was excluded from the MDS analysis reported above.

2.6 Summary of Discussion

From our analyses above, we can see that T35 and T214 seem to share some intrinsic phonetic property which affects perception of these tones by both the Chinese and English listeners. It is obvious that the English listeners were dealing with the phonetic characteristics of the tones. They perceived T35 and T214 as being similar for the same reason they did T35 and T55: the starting and/or ending pitch of the first syllable match the starting pitch value of the second syllable in a pair. The fact that the English listeners were paying more attention to these phonetic details in the tonal contours than the Chinese listeners is not inconsistent with the findings reported in Wang (1976) and Stagray & Downs (1993) that the English listeners may have higher sensitivity for frequency changes.

The perceived distance between T35 and T214 is smaller relative to the other tone pairs in the Chinese listeners' space, while the inter-pair difference in perceived distances for the English listeners is less pronounced. However, the absolute T35/T214 distance for the Chinese listeners is longer than that of the English listeners (see Table 2.2).

We should note that there were quite a few previous studies on the confusability of T35 and T214 in native and non-native perception using both natural speech and nonspeech synthetic stimuli (e.g., Kiriloff, 1969; Chuang, Hiki, Sone and Nimura, 1972; Blicher, Diehl, & Cohen, 1990; Shen & Lin, 1991). Most of these studies used citation forms of the tones in either natural speech or synthesized stimuli. It was found that both the intrinsic duration of these tones and the inflection points on the contours (i.e., where the rise starts) contribute to the confusability. But neither duration nor inflection point is relevant in our experiment, as the T214 stimuli in our stimuli were produced with a short low-falling contour. Thus, the confusability between T35 and T214 must have come from the similar initial pitch points for the English listeners, who attached much weight to the "start pitch height" dimension in their perceptual space. For the Chinese listeners who appeared to have used initial pitch to predict contours, these two tones should also be confusable. However, given the fact that contour shape is almost as important as start pitch height in their perception, the magnitude of the confusion between T35 (with a rising contour) and T214 (realized only as a low, slightly falling contour in our stimuli) should be reduced, which did not seem to happen. A possible interpretation would be that comparison between the tones by the Chinese listeners was actually made at a higher cognitive level, where they are represented as abstract categories and where tone sandhi and other allophonic information is accessible.

Recall that the within-group pairwise comparisons (using repeated measures ANOVA) also showed that the Chinese listeners treated the T35/T214 pairs as being different from all other tone pairs. This seemingly surprising pattern can be explained if, as Peng (1996) found, some surface [T35] syllables may be linked to both /T35/ and /T214/ morphemes (maybe as part of a compound) in the Chinese listeners' lexicon. On the other hand, Wang's (1995) findings on Taiwanese tone production indicate that both citation and sandhi tones may be stored in the lexicon. It is worth noting that with such a mental representation of the tonal category (or rather categories) of certain morphemes and a one-to-many mapping of surface tone to underlying tone categories, the boundary between the T35 and T214 categories may be blurred and the confusion between these tones may exist beyond just the sandhi environment, which explains why there is not much difference between the RTs for pairs T35-T214 and T214-T35.

The patterns revealed in the T test and INSCAL also provide evidence for the contention that the Chinese listeners treat each tonal contour as an indivisible unit (Wang, 1967; Jansche, 1999 ms.), as the ending pitch target of the contour did not seem to contribute much to the confusability or distinctiveness of the tones. Instead, contour shape, or pitch movement, showed up as the second dimension, accounting for not much less variance than the dimension of onset pitch height. Thus, unlike the English listeners who were using these pitch points as important cues to distinguish the tones, the perception of the Chinese listeners seemed to be independent of these cues to a certain extent. This may sometimes be a setback in their ability to distinguish tones as in the case of the pair T55-T214, where the Chinese listeners seemed to have "suffered" from their phonological knowledge – i.e., failed to use the acoustic cues, especially pitch offsets, as effectively as the English listeners did. Examination of the error data shown in Table 2.2 reveals a similar pattern there: the mistakes that the English listeners made were more auditorily-driven. This may also explain why dimension 2, which clearly corresponds to the end pitch height in the English listeners' space, has to be interpreted differently in the Chinese listeners' perceptual tone space. The same may be said for why the positions of T35 and T214 are switched along dimension 2 for the two groups of listeners.

CHAPTER 3

LOW UNCERTAINTY AX TONE DISCRIMINATION OF NATURAL SPEECH AND NONSPEECH SINEWAVE TONES AND DIFFERENCE RATING BY CHINESE- AND AMERICAN ENGLISH-SPEAKING LISTENERS

Different experimental conditions and procedures have long been known to yield quite different results. For instance, categorization and discrimination tasks seem to direct participants' attention to different properties of the stimuli and to elicit different processing strategies. In testing and modeling the phonetic magnet effect (Kuhl, 1991; Iverson & Kuhl, 1996), Guenther, et al. (1999) found that categorization training led to the effect, while discrimination training did not. Similarly, Werker & Tees (1984) reported better non-native perception in an AX discrimination test than in a category change test. Shorter ISI may also reduce the memory-load (Pisoni, 1973; Macmillan, 1987; Guenther et al., 1999). This chapter further explores the influence of tone sandhis on perception with three more experiments involving different experimental procedures. Experiments 2 and 3 employed a low uncertainty AX discrimination task, with synthesized sinewave tones and natural speech monosyllables, respectively. Experiment 4 used natural speech monosyllables and a subjective AX difference rating task.

3.1 Experiment 2: AX Discrimination of Sinewave Tones (limited stimulus set)

Although there is evidence that Chinese and AE listeners may have employed different processing strategies in Experiment 1, we felt it necessary to try to further tease apart the effects of raw acoustic similarity in the data from linguistic effects. One way to do this is to compare the results from that experiment with the ones obtained from a psychoacoustic task. Thus, in Experiment 2 an AX discrimination with limited stimulus set and nonspeech synthetic tones was used. If we take the results from this low uncertainty task as the perceptual baseline determined by raw acoustic similarity in the stimuli, deviations from these results can then be seen as linguistic effects.

3.1.1 Participants in Experiment 2

Eleven (11) Chinese listeners and thirteen (13) English listeners were recruited from the OSU Columbus campus. Again, the Chinese were paid a small amount of money, whereas the Americans earned course credits.

3.1.2 Stimuli

This experiment employed non-speech stimulus tones. These are simple sinewave simulations of the natural speech tones recorded as monosyllables by a male speaker in his early thirties from Beijing. The stimuli were generated with a sinewave synthesizer adapted by Keith Johnson from c-code generously shared by Alex Francis and Howard Nusbaum at the University of Chicago. Specifically, the frequency of a single time varying sinusoidal wave was modeled on the f0 of each of four recorded monosyllables /ba^{55}, ba^{35}, ba^{214}, ba^{51}/. The amplitude contour of the sinusoid was also modeld on the amplitude contour of these naturally produced syllables (but see below). The overall impression of the synthetic sinusoidal stimuli was that they were like low-pass filtered speech, but with the pure tone quality of a sinewave. I shall continue to use the labels T55, T35, T214, and T51 for the synthetic stimuli.

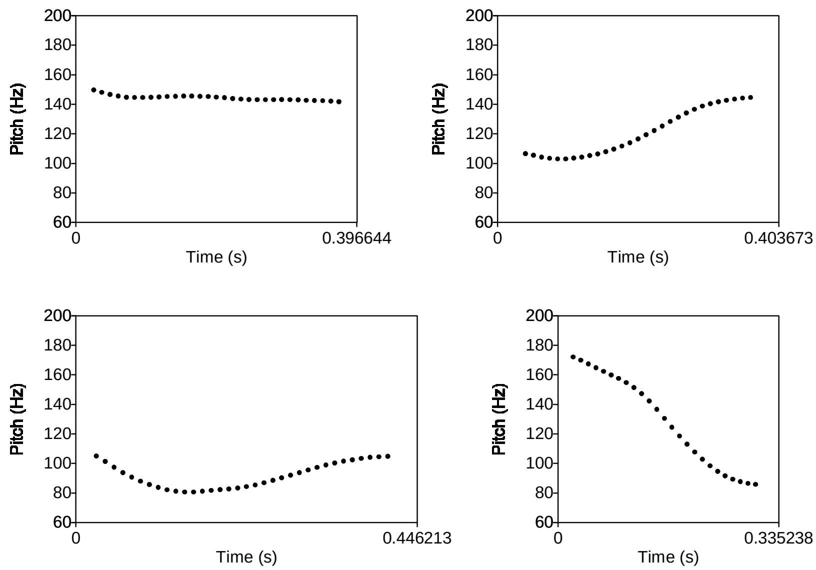


Figure 3.1 F0 traces of tones T55 (upper left panel), T35 (upper right), T214 (lower left), and T51 (lower right) as produced by the male Beijing speaker. The segmental makeup used in these recordings is /ba/. Lengths of the X-axes in these panels try to reflect approximately the relative lengths of the tones. These natural speech tokens were used as templates when generating the synthetic sinewave stimulus tones.

3.1.3 Method: limited stimulus set

Participants were tested in front of a computer (two to four people at a time) in a quiet room, using the E' program (Psychology Software Tools, Inc.) installed on PCs. The stimuli were played through headphones. An AX discrimination task was used. But this was a different task from Experiment 1, in that a limited set of stimuli were presented in each of the six (6) blocks, with each block testing the discrimination of only two tones (e.g., T55 and T35 might be tested in block 1, T55 and T214 in block 2, T55 and T51 in block 3, and so on and so forth). Each of the four possible combinations of the two tones tested was repeated twice in that block (e.g., block 1 might have four pairs, T55-T55, T55-T35, T35-T55, and T35-T35, all of which were repeated twice, yielding 4 x 2 = 8 pairs in total). There were 8 × 6 trials in total. The order of the blocks was randomized for different participants. There was a brief training session, which contained just four pairs of tones, that is, the four possible combinations of two tones without a second repetition.

During the test, each stimulus pair was played at a 100ms inter-stimulus interval. As in Experiment 1, after a listener made a response, a feedback message, detailing his/her RT and percentage correct, was displayed on the screen for 1500ms. Another 2000ms wait period followed. E' then moved on and played the next pair of sounds. As in Experiment 1, both RTs and response accuracy were recorded. Only the RT data of correct "different" responses will be analyzed, and the identical pairs were included as fillers.

3.1.4 Results and Analyses

3.1.4.1 Repeated measures analysis of variance (general linear model)

As can be seen in Figure 3.2 and Table 3.1 below, the overall RTs for the two groups are very similar. No significant between-subject language effect was detected by the repeated measures analysis in the RT data, [$F(1, 22) = .120$, $p = .733$, partial $\eta^2 = .005$]. But there was a significant effect with the within-subject factor of tone pair type, sig.[$F(8.597, 189.124) = 12.992$, $p < .001$, partial $\eta^2 = .371$]. And the tone pair by language interaction has a marginal effect, [$F(8.597, 189.124) = 1.953$, $p = .05$, partial $\eta^2 = .082$].

Tone Pair	Chinese	English
T55/T35	578	597
T55-T35	538	558
T35-T55	617	636
T55/T214	560	528
T55-T214	546	505
T214-T55	573	551
T55/T51	649	591
T55-T51	612	573
T51-T55	685	609
T35/T214	647	629
T35-T214	628	612
T214-T35	666	646
T35/T51	533	566
T35-T51	506	573
T51-T35	559	558
T214/T51	553	541
T214-T51	526	524
T51-T214	579	557

Table 3.1 RTs (in milliseconds) for correct "different" responses from the AX (limited set) discrimination experiment using sine wave stimuli. The RTs are average values computed from each subject's median RTs.

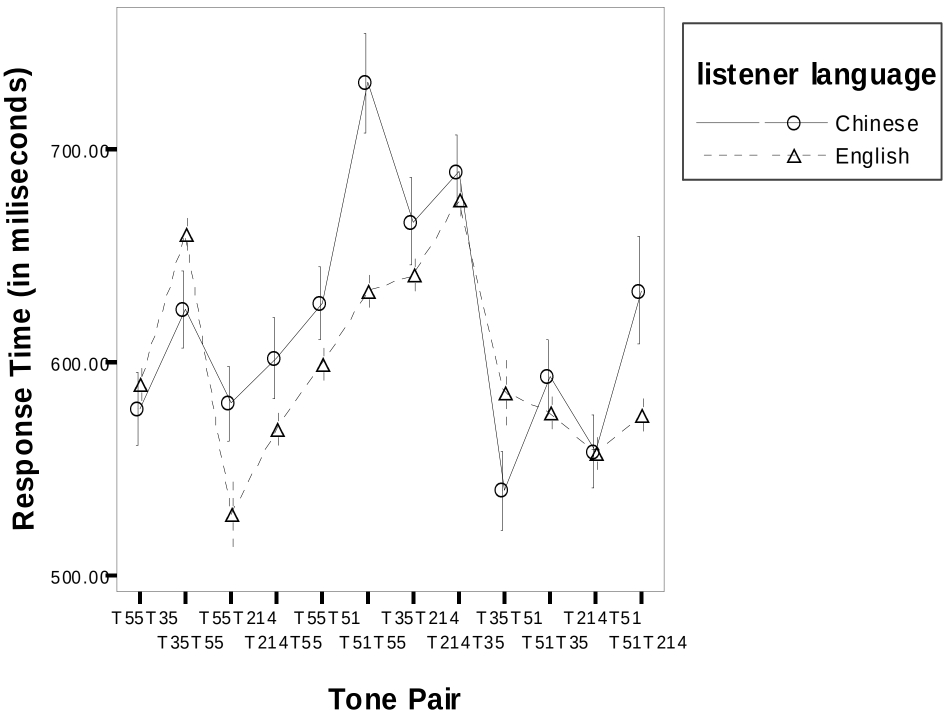


Figure 3.2 Response time plot from the experiment of AX limited stimulus set discrimination of sinewave tones for Chinese and AE listeners. No significant language effect was found in either the repeated measures ANOVA or the T test on the median RT data. Significant or marginal effects did show up in the T test for T55-T51 ($p < .001$), T55-T214 ($p = .032$), T51-T214 ($p = .049$) and T35-T51 ($p = .069$) when all RT observations by individual subjects were included in the analysis.

3.1.4.2 Planned between-group comparisons: Independent Samples T Test

No significant difference was found between the two language groups for any tone pair in the median RT data in the planned coparisons using an Independent Samples T test. But we can see in the plots in Figure 3.2 that for pairs T55-T214, T51-T55 and T35-T51, the two groups of listeners behaved slightly differently. In the first two cases, the English listeners had slightly shorter RTs, whereas in the third they had a slightly longer RT. It should be obvious that these patterns are again due to the fact that AE listeners tend to compare the f0 offset of the first tone with the f0 onset of the second tone, while Chinese listeners do not. Significant or marginal effects did show up in the T test for T51-T55 ($p < .001$), T55-T214 ($p = .032$), T51-T214 ($p = .049$) and T35-T51 ($p = .069$) when all RT observations by individual subjects were included in the analysis.

Tone pair	t	p	η^2
T55-T214	-2.160	.032	.020
T51-T55	-3.660	< .001	.079
T51-T214	-1.984	.049	.022
T35-T51	1.825	.069	

Table 3.2 Results of Independent Samples T Test on all observations of the RT data from the limited stimulus set AX discrimination experiment using synthesized sinewave stimuli.

3.1.4.3 Within-subject pairwise comparisons (ANOVA)

Within-subject pairwise comparison for the American English listeners' RT data showed that pair T55-T214 (with the shortest RT) and pair T214-T35 (with the longest RT) both differed significantly from four other pairs ($p < .05$). The rest of the pairs fell in the middle, with T55-T35, T55-T214, T35-T51, and T214-T51 showing no significant difference from any other tone pair. In the Chinese listeners' data, the pairs are even less well separated due to large variances. But we can roughly derive three groupings: (i) T35/T214 and T55/T51 are the most confusable; (ii) T55-T35, T55-T214, T35-T51 and T214-T51 are the least confusable; and (iii) T35-T55, T214-T55, T51-T35, and T51-T214 are not significantly different from any other pair.

Note that pairs T55/T51 are now among the most confusable for the Chinese listeners, a pattern that is different from Experiment 1. Upon re-examining the sinewave stimulus tones, it was noticed that the falling portion of T51 was somewhat delayed as compared with its original speech template when the pitch traces were aligned relative to vowel onset and the durations normalized (see Figure 3.3 below). (This was only found with the T51 natural speech and sinewave tones. The other three natural speech template and sinewave pairs have good alignments of pitch points.) Such a delay in the falling contour may have made the contour more like T55 for the Chinese listeners. Note that the effect is more obvious with the T51-T55 pair than the other way around. This is so because the f0 onset of the sinewave T51 is about 30Hz higher than that of T55, a difference large enough for the Chinese listeners to make the correct "different" judgment when they compared the onsets of tones in the pair T55-T51. When the order of tones was reversed, however, a difference of 30Hz might not be enough, due to the downstep effect of T51 (caused by the low pitch offset) which lowers high f0 targets following it (see, e.g. Xu, 1997; Huang, 2002). In addition to the discrepancy in the f0 contour, the durations and intensity curves also differ between the sinewave tone and its natural speech template: the sinewave tone was about 1/7 (or 40ms) shorter with a flat intensity line, while the intensity decreases sharply in the last 1/3 in the natural speech monosyllable (total duration = 340ms). Such a disparity in intensity envelops may also contributed to the perceived similarity between T55 and T51.

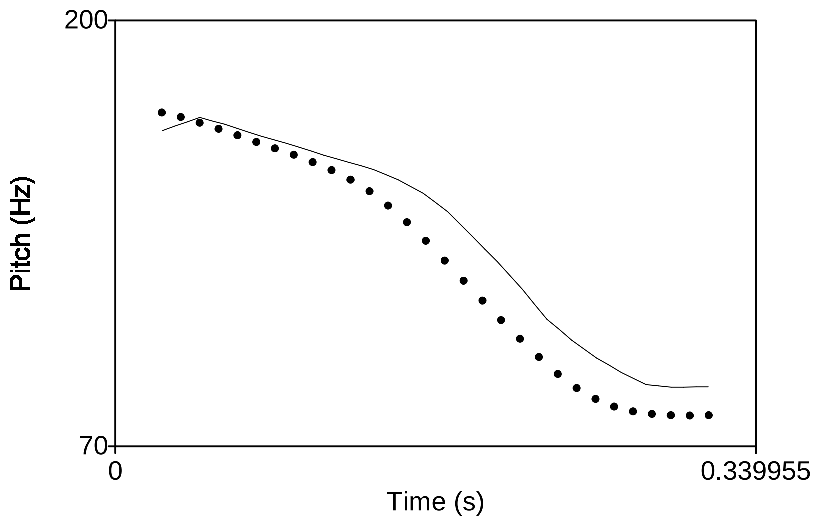
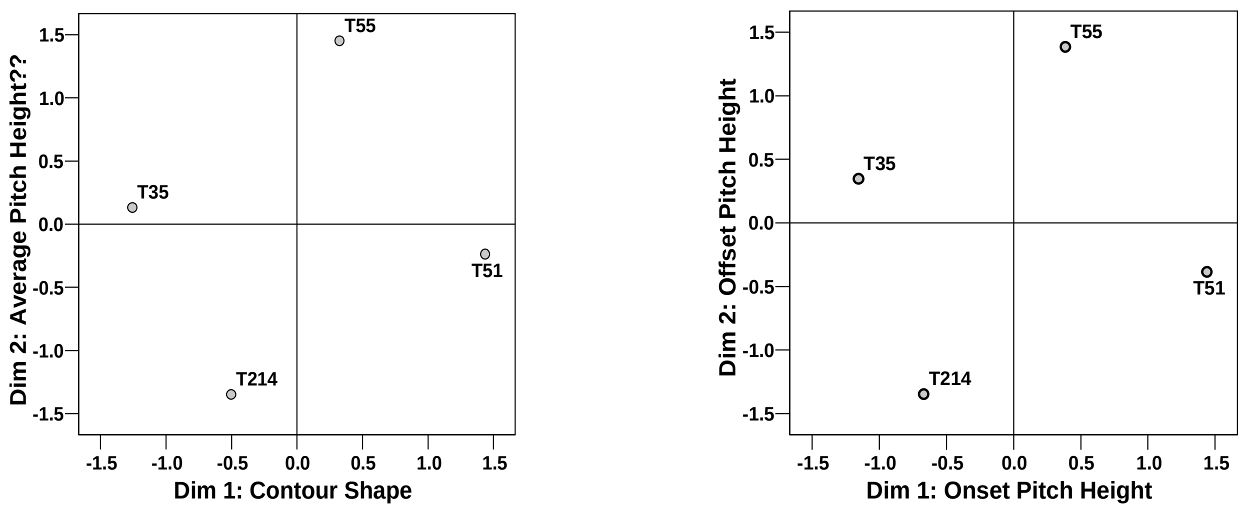


Figure 3.3 F0 traces of the sinewave T51 (solid line) and its original speech template T51 (dotted line). The falling portion was somewhat delayed in the sinewave tone, which may explain the discrepancy in the T55/T51 RTs between Experiments 3 (natural speech) and 4 (sinewave tones). Note that the duration of the sinewave tone was about 300ms and was normalized to the length of its speech template here.

3.1.4.4 MDS

The INSCAL perceptual spaces basically reflect the patterns seen earlier in the RT plots and the groupings obtained through pairwise comparison. In the Chinese listeners' space in Figure 3.4a [stress = .162, RSQ = .915, ratio of weights for dimensions 1 and 2 = .5846 : .3306 (or 1.768 : 1)], T35 and T214 are a bit closer to each other than T55 and T51, which have the next shortest distance. We see a very similar pattern in the English listeners' space in Figure 3.4b [stress = .165, RSQ = .908, ratio of weights for dimensions 1 and 2 = .5258 : .3824 (or 1.375 : 1)], although here T35 and T214 are slightly farther apart than in the Chinese listeners' space. Three listeners' data were excluded from the MDS analysis in the English group, because they had high weirdness numbers. The dimensions seem to correspond to start f0 height and end f0 height, respectively. But with T35 and T51 very close to each other along the Y-axis, it is probably more appropriate to interpret dimension 1 in the Chinese listeners' space as "contour shape". Despite these differences, the two spaces are much more similar than were found in Experiment 1, suggesting a considerable weakening of the language effect.



a.

b.

Figure 3.4 Chinese (panel a; stress = .162, RSQ = .915) and English (panel b; stress = .165; RSQ = .908) listeners' perceptual tone spaces derived from the AX (limited stimulus set) discrimination task using sine wave tones.

3.1.5 Comparison of Results from Experiments 1 and 2

The tonal confusability rankings for the Chinese listeners differ between Experiment 1 (AX discrimination, ISI = 300ms, roving; Chapter 2) and Experiment 2 (AX discrimination, ISI = 100ms, limited stimulus set). Differences were also seen in the MDS spaces, with the relative distance between T35 andT214 in the Chinese listeners' space being much shorter in Experiment 1 than in Experiment 2. Assuming that the patterns reflect raw acoustic similarity in Experiment 2, the patterns found in Experiment 1 cannot be attributed completely to the same cause. The patterns are not very different for the AE listeners in these two experiments, although the characteristic of T51 stimulus in Experiment 2 also seemed to made it more similar to T55 for the nonnative and native listeners alike.

3.2 Experiment 3: AX Discrimination of Natural Speech Tones (limited stimulus set)

As noted above, there are obvious differences between the results of Experiments 1 and 2, although we note that the natural speech templates used to generate sinewave tones were not the same as the stimuli used in Experiment 1, where the stimulus tones were the first syllables extracted from disyllabic natural speech recordings. To make the results more comparable, a third experiment was run right after Experiment 2 with the same procedure (AX discrimination, ISI = 100ms, limited stimulus set) and the same listeners (Chinese and AE), with a short break in-between if when the listener needed it.

3.2.1 Participants

The same Chinese and AE listeners who participated in Experiment 2 later took part in Experiment 3 within the same one hour session.

3.2.2 Stimuli

This experiment used a set of natural speech monosyllabic stimuli recorded in monosyllables by a male Beijing (Putonghua) speaker in his early thirties. (Four of these monosyllables served as templates to generate the synthetic tones used in Experiment 2.) The segmental makeup was kept constant as /ba/. Twenty (20) randomized lists of the monosyllables were recorded in a sound-proof booth in the phonetics laboratory at the OSU Linguistics Department. The speaker read from the afore-mentioned randomized lists and was recorded with a head-mounted microphone (Shure SM10A model) and a DAT recorder. The recordings were later extracted with Xwaves and played to the listeners at 22,050Hz with 16-bit samples. The five (5) best productions for each of the four tones were selected to splice the test stimulus pairs. In determining the best productions, phonetic characteristics usually concomitant with a particular tone, such as pitch height and contour, duration of syllable, and voice quality (e.g., creakiness in T214), were taken into consideration. The most typical productions were selected. The f0 traces of these stimulus tones are shown in Figures 3.1. Note that the T214 tonal contour is fully realized, as it was produced in a prepausal position, although the final rise is still not to the level of "4" as the traditional analysis and label indicate, rendering T214 a rather low tone. Note also that the inflection point is realized no later than 1/3 of the whole contour, not as late as half way through as shown in some earlier studies (see, e.g., Chuang et al., 1972). So, more portion of the contour is a rise, although the rate of change in pitch is very small relative to that in the rising T35.

3.2.3 Method

As in Experiment 2, an AX discrimination task with a short ISI of 100ms, and limited stimulus set.

3.2.4 Results and Analyses

3.2.4.1 Repeated Measures ANOVA

As in previous analyses, median RT values for the tone pairs were selected for each subject for the repeated measures analysis. No significant language group effect was found, [$F(1, 27) = 2.486$, $p = .127$, partial $\eta^2 = .084$]. The within-subject tone pair type effect was significant, sig.[$F(8.864, 239.328) = 6.963$, $p < .001$, partial $\eta^2 = .205$]. There was no significant "tone pair" by "language" interaction effect, [$F(8.864, 239.328) = 1.374$, $p = .202$, partial $\eta^2 = .048$]. Language group profile plots of response time are shown in Figure 3.5 below. The same RT values are reported in Table 3.3. With the low uncertainty design, error rates are even lower than those found in Experiment 1 and negligible, with the most errors occurring in T35-T214 (3.75%) for the Chinese listeners and in T51-T55 and T35-T214 (6.92%) for the English listeners.

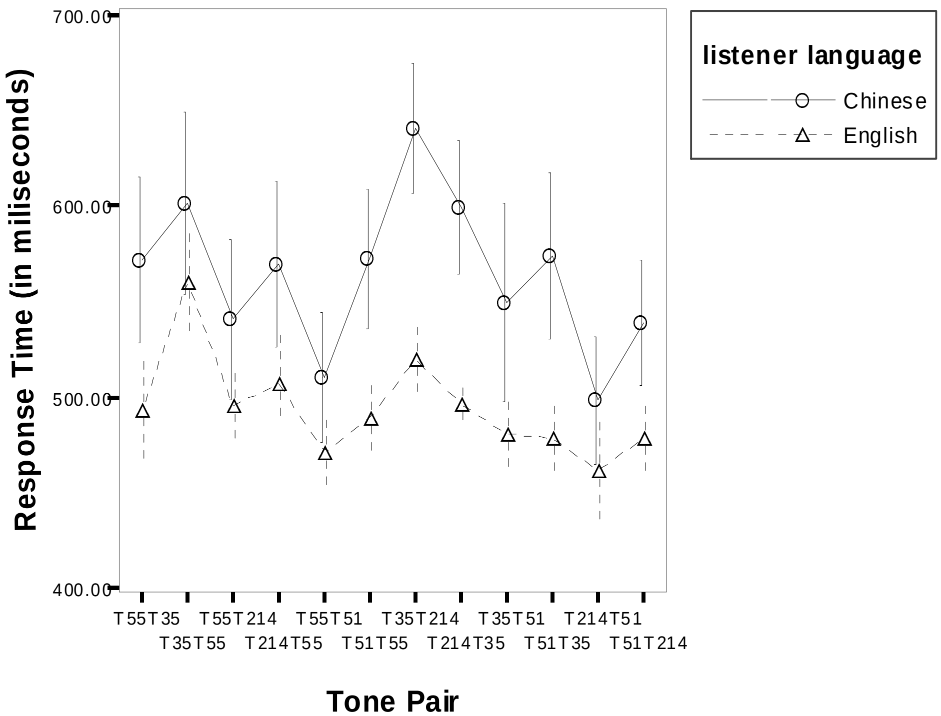


Figure 3.5 RTs (in milliseconds) for the correct "different" responses from the AX limited set discrimination experiment using natural speech stimuli. The same RT measurements are reported in Table 3.5. No language effect was found to be significant in the repeated measures ANOVA. But planned comparisons (T test) revealed significant differences between the two listener groups for pairs T35-T214 and T214-T35.

Tone Pair	Chinese	English
T55/T35	586	527
T55-T35	571	493
T35-T55	601	560
T55/T214	550	515
T55-T214	541	521
T214-T55	569	508
T55/T51	541	480
T55-T51	510	471
T51-T55	572	489
T35/T214	620	508
T35-T214	640	520
T214-T35	599	497
T35/T51	561	480
T35-T51	549	481
T51-T35	573	479
T214/T51	518	471
T214-T51	498	462
T51-T214	539	479

Table 3.3 RTs (in milliseconds) for correct "different" responses from the AX limited stimulus set (fixed block order) discrimination experiment using natural speech stimuli. The RTs are average values computed from each subject's median RTs.

3.2.4.2 Planned between-group comparisons: Independent Samples T Test

Planned between-group comparisons using the independent samples T test showed that the RTs for pairs T35/T214 were significantly different for the two language groups, with η^2 showing over 20% of the variances accounted for in each case (see Table 3.4 below).

Tone pair	*t*	*p*	η^2
T35-T214	2.828	.009	.229
T214-T35	2.648	.015	.301

Table 3.4 Results of Independent Samples T Test on the RT data from the limited stimulus set AX discrimination experiment using natural speech stimuli. (According to Levene's test, equal variance cannot be assumed for the two groups in the case of T214-T35.)

3.2.4.3 Within-subject pairwise comparisons ANOVA

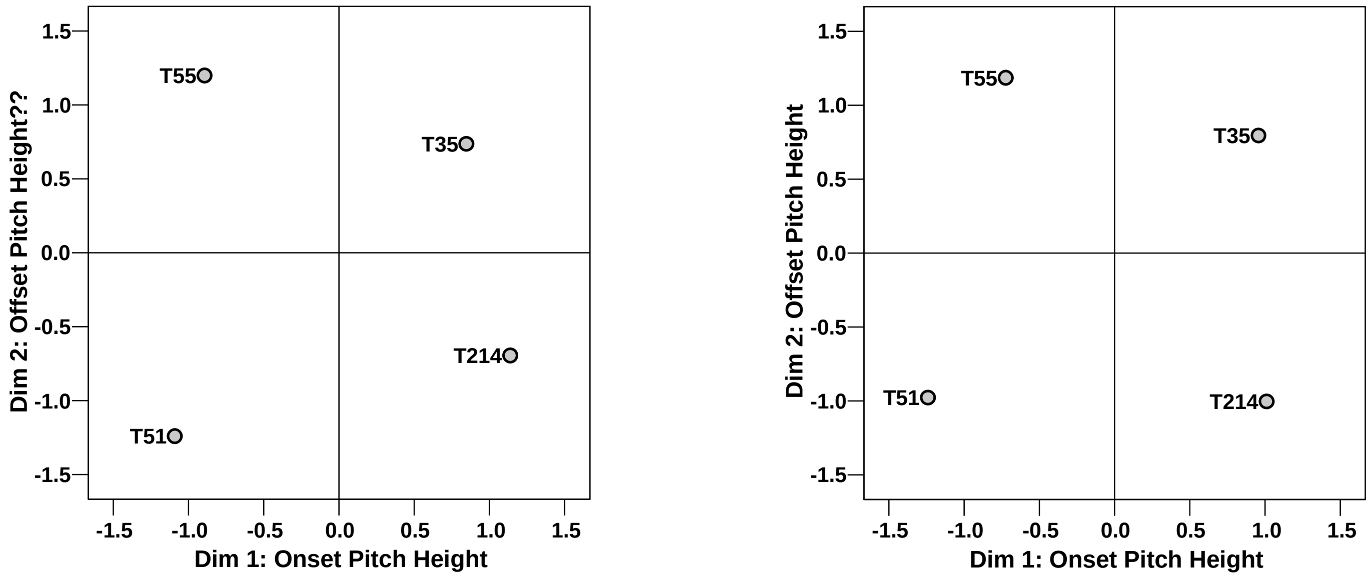
Within-subject Pairwise comparisons on the Chinese listeners' RT data showed pairs T35/T214 as the most confusable, with T35-T214 significantly different ($p < .05$) from six (6) other pairs and T214-T35 from three (3) pairs. On the other hand, pairs T214-T51 and T55-T51 were the least confusable, with T214-T51 significantly different from three

(3) other pairs and T55-T51 from two (2) pairs. Pairs T55/T35, T214-T55, and T35-T51 fell in the middle, none of which was significantly different from any other pair, although pair T35-T55 does show a slightly longer RT (see Figure 3.5).

For the English listeners, although pairs T35-T55 and T35-T214 have relatively longer RTs, no significant difference was found between any two pairs. This means that the significant within-subject main effect of "tone pair type" came from the Chinese listeners.

3.2.4.4 MDS

The perceptual tone spaces from the MDS analyses for the two groups of listeners are somewhat different. Most noticeably in Figure 3.6 below, T35 and T214 are much closer to each other in the Chinese listeners' space than in the English listeners'. The dimensions seem to correspond to "onset f0 height" and "offset f0 height" for both spaces. The label for dimension 2 may be less appropriate for the Chinese listeners' space, as T55 and T35 (i.e., the two tones having high f0 offsets) do not align well. Neither do T51 and T214 (i.e., the two tones having low f0 offsets). Perhaps we should label dimension 2 in the Chinese listeners' tone space as "contour shape": moving from the top to the bottom, we have level (T55), rising (T35), falling-rising (T214) and falling (T51).



a. b.

Figure 3.6 Chinese (panel a; stress = 0.191, RSQ = 0.885) and English (panel b; stress = 0.197, RSQ = 0.871) listeners' perceptual tone spaces computed from the limited stimulus set natural speech discrimination RT data.

3.2.5 Summary if discussion and Comparison of results from Experiments 2 and 3

There are noticeable differences between the results from Experiments 2 and 3. First of all, the overall RT is shorter with the speech stimuli in Experiment 3 for both Chinese and AE listeners, which may be seen as a training effect (Werker & Logan, 1985), as Experiment 3 was run right after Experiment 2. Note that this training effect was more obvious in the AE listeners' RT data (Figures 3.2 and 3.5). Confusability rankings also differ in the two experiments for both Chinese and AE listeners, with the Chinese group showing a pattern more like that found in Experiment 1 (with a shorter relative distance between T35 and T214) and the AE listeners showing no significant difference in RT between any two pairs in Experiment 3. The smaller RT improvement in Experiment 3 (as compared with Experiment 2) for the Chinese listeners might be indicating that with the segmental makeup and a human voice, the stimuli in Experiment 3 prompted more lexical activation (as opposed to very limited lexical activation in Experiment 2).

3.3 Experiment 4: Degree of Difference Rating of Natural Speech Tones

Assuming a positive correlation between RT and confusability, we tried to derive confusability rankings of tones from the RT data in Experiments 1, 2 and 3. These experiments used a simple AX "same"/"different" discrimination task (with a further simplification of limited stimulus set in Experiments 2 and 3) generally assumed to tap auditory processing (Pisoni, 1973; Macmillan, 1987; Johnson, 2004). Thus, the confusability rankings derived in these experiments may reflect mainly auditory similarities among the tones. Experiment 4 further investigated the confusability of Putonghua tones with a difference rating task assumed to tap linguistic processing.

3.3.1 Participants in the experiment

Twenty-one (21) Chinese listeners and thirty (30) American English listeners were recruited from the Columbus campus of the Ohio State University (OSU). The Chinese listeners were OSU graduate or undergraduate students (or their spouses) from the city of Beijing. Three of these Chinese listeners also participated in Experiments 2 and 3. The English listeners were undergraduate students taking linguistics courses at OSU. The Chinese were paid a small amount of money for their participation in the experiment, whereas the Americans earned course credits.

3.3.2 Stimuli

This experiment employed the same natural speech monosyllables as those used in Experiment 3.

3.3.3 Method

An AX difference rating task was used. Participants were tested in front of a computer (two to four people at a time) in a quiet room. The stimuli were presented to them through headphones, using the E' program (Psychology Software Tools, Inc.) installed on PCs. Each stimulus pair was played at a 100ms inter-stimulus interval. After the listener made a response (at any time when s/he was ready, usually within 1500ms), there was a 2000ms wait period before E' moved on to the next stimulus pair. Note that the ISI was again shortened in comparison to Experiment 1 (300ms).

There was a brief training session with three pairs involving two different monosyllables and one identical pair (involving two repetitions of the same monosyllable). The test session consisted of six (6) blocks, each of which contained 32 stimulus pairs. Both the block order and the stimulus pair order within each block were randomized. Thus, all participants listened to $32 \times 6 = 192$ pairs of the form /ba-ba/, with the two monosyllables bearing identical or different tones. The 32 pairs in each block included the 12 x 2 different pairs and 4 x 2 identical pairs (i.e., each of the 16 possible pairings of the four tones was repeated twice in any of the test blocks). Each identical pair contains two repetitions of the same .wav sound file. As with Experiment 1, only the results of different pairs were analyzed, and the identical pairs were just included as fillers.

Written and oral instructions were given in English and Chinese to the two groups of listeners, respectively. The listeners were asked to listen carefully for tonal differences and rate the degree of difference on a "1" to "5" scale subjectively. The scale was described for them in the format shown in Table 3.5. They were especially encouraged to use the full scale, instead of just "1" and "5". They were also asked not to think too much when they rated the differences, lest we would get all "1"s and "5"s if they contemplated for too long. Five different keys on each button box for response input were labeled "1" through "5".

very similar	moderately similar	somewhat different	moderately different	very different
1	2	3	4	5

Table 3.5 Rating scale described for the listeners on the Instruction sheet.

3.3.4 Results and Analyses: Difference Ratings

Despite the instructions for using the whole scale of "1" through "5", some listeners used only "1" and "5". These data had to be discarded. As a result, only twenty-six (26) American English and eighteen (18) Chinese listeners' data were analyzed.

3.3.4.1 Repeated measures ANOVA

The repeated measures ANOVA, with "tone pair type" as the within-subject variable (12 levels) and "listener language" as the between-subject variable, yielded the group profile plot in Figure 3.7. The same group average ratings are also reported in Table 3.6.

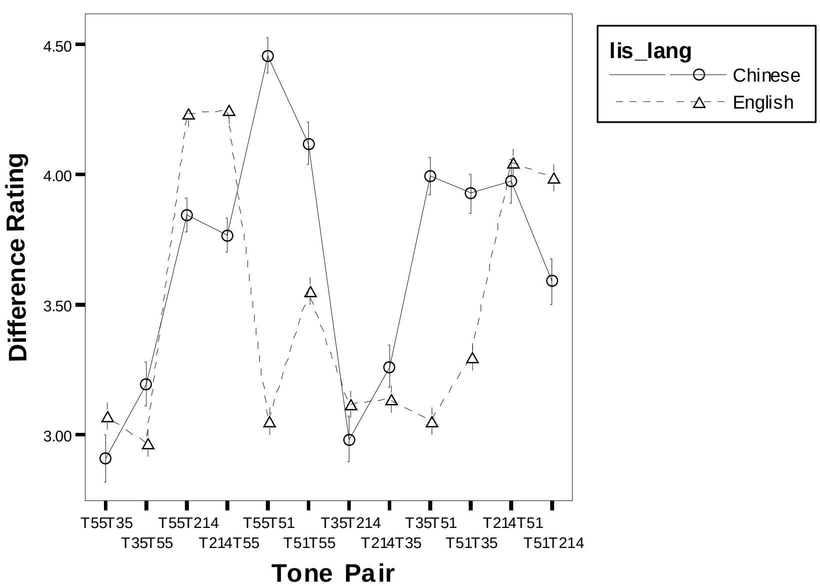


Figure 3.7 Subjective degree-of-difference ratings by the Chinese and English listeners. These group average values were computed from each subject's tone pair median values. The same rating values are also reported in Table 3.2 above. "1" = "very similar", "5" = "very different". Significant between-subject language effect was found in both the repeated measures ANOVA and the T test. Error bars show one standard error.

Tone Pair	Chinese	English
T55/T35	3.08	3.03
T55-T35	2.93	3.07
T35-T55	3.23	2.98
T55/T214	3.82	4.23
T55-T214	3.85	4.21
T214-T55	3.79	4.24
T55/T51	4.28	3.31
T55-T51	4.44	3.04
T51-T55	4.12	3.57
T35/T214	3.13	3.12
T35-T214	2.99	3.11
T214-T35	3.27	3.12
T35/T51	3.98	3.18
T35-T51	4.02	3.05
T51-T35	3.93	3.31
T214/T51	3.80	4.02
T214-T51	4.01	4.05
T51-T214	3.59	3.98

Table 3.6 Degree of difference ratings by the Chinese and English listeners.

There was a significant between-subject language effect, sig.[$F(1, 423) = 13.044$, $p < .001$, partial $\eta^2 = .03$]. That is, how a listener rated the tonal difference was at least partially dependent on his/her native language. There was also a significant effect with tone pair types, sig.[$F(10.025, 4240.384) = 75.929$, $p < .001$, partial $\eta^2 = .152$]. The interaction of language and tone pair types was significant as well, sig.[$F(10.025, 4240.384) = 40.609$, $p < .001$, partial $\eta^2 = .088$].

As can be seen from the plot in Figure 3.7, the two groups of listeners have very different views on how similar or different two tones are. The overall pattern seems to be that for the Chinese listeners, only pairs T55/T35 and T35/T214 were considered to be most similar and that for the English, only pairs T55/T214 and T214/T51 were deemed most dissimilar. They do agree sometimes: pairs T55/T35 and T35/T214 were rated as being "quite similar" by both groups, while pair T214-T51 "quite different" by both. The most obvious differences between the two groups are with pairs T55/T51 and T35/T51. There are also minor differences for pairs T55/T214 and T214-T55.

The Chinese listeners' rating pairs T35/214 as most similar is certainly consistent with the findings in Experiment 1 on tones T214 and T35, the contrast between which is lost due to the T214 sandhi rule that leads to paradigmatic neutralization of these tones before another T214. The pattern with pairs T55/T35 seems different in the two experiments, although recall that we did notice that the Chinese listeners placed these two tones a bit closer along one dimension than the English listeners in the stimulus space derived in Experiment 1. It seems that the "rate difference" task, with less time constraint, brought out a stronger effect of another sandhi rule, namely the T35 rule that neutralizes the contrast between T35 and T55 paradigmatically. But there may be another reason for such a disparity. Unlike the T214 sandhi, the T35 sandhi, noted for Beijing where the participants in Experiment 2 are from, is not as pervasive and may not exist in the Mandarin Putonghua spoken by the participants in Experiment 1. As in Experiment 1, the high versus low pitch contrast seems to be the most salient for the AE listeners.

3.3.4.2 Planned comparisons of English and Chinese listeners' ratings: T test

The differences seen between the two language groups (Figure 3.7 and Table 3.6) and pointed out in the previous section also showed up in planned comparisons of the rating data in the independent samples T test. As reported in Table 3.7 below, the ratings for pairs T55/T214, T55/T51, T35/T51, T51-214 and T35-T55 are significantly different for the two groups of listeners ($p < .05$). The largest disparity lies with pairs T55-T51 ($t = 14.507$, $p< .001$, $\eta^2 = .335$) and T35-T51 ($t = 9.494$, $p < .001$, $\eta^2 = .183$), supporting our analysis that the start/end pitch points of tonal contours are more important for the English listeners than for the Chinese.

Tone pair	*t*	*p*	η^2
T35-T55	2.135	.033	.010
T55-T214	-4.511	< .001	.045
T214-T55	-5.453	< .001	.064
T55-T51	14.507	< .001	.335
T51-T55	5.394	< .001	.063
T35-T51	9.494	< .001	.183
T51-T35	6.45	< .001	.087
T51-T214	-3.732	< .001	.043

Table 3.7 Tone pairs for which English listeners' ratings were significantly different from the Chinese listeners' ratings.

3.3.4.3 Within-subject pairwise comparisons (ANOVA)

Within-subject pairwise comparison revealed significant differences between tone pairs ($p < .05$). For the Chinese group, the most dissimilar pair is T55-T51 and the most similar pairs are T55/T35 and T35/T214, as can be seen in the rating plot in Figure 3.7. All other pairs fall in-between. Thus, here again, we seem to have a pattern very similar to what we found in Experiment 1: the Chinese listeners treated any two tones as being quite different, except when the two tones are involved in a sandhi rule, in which case the two tones were considered very similar.

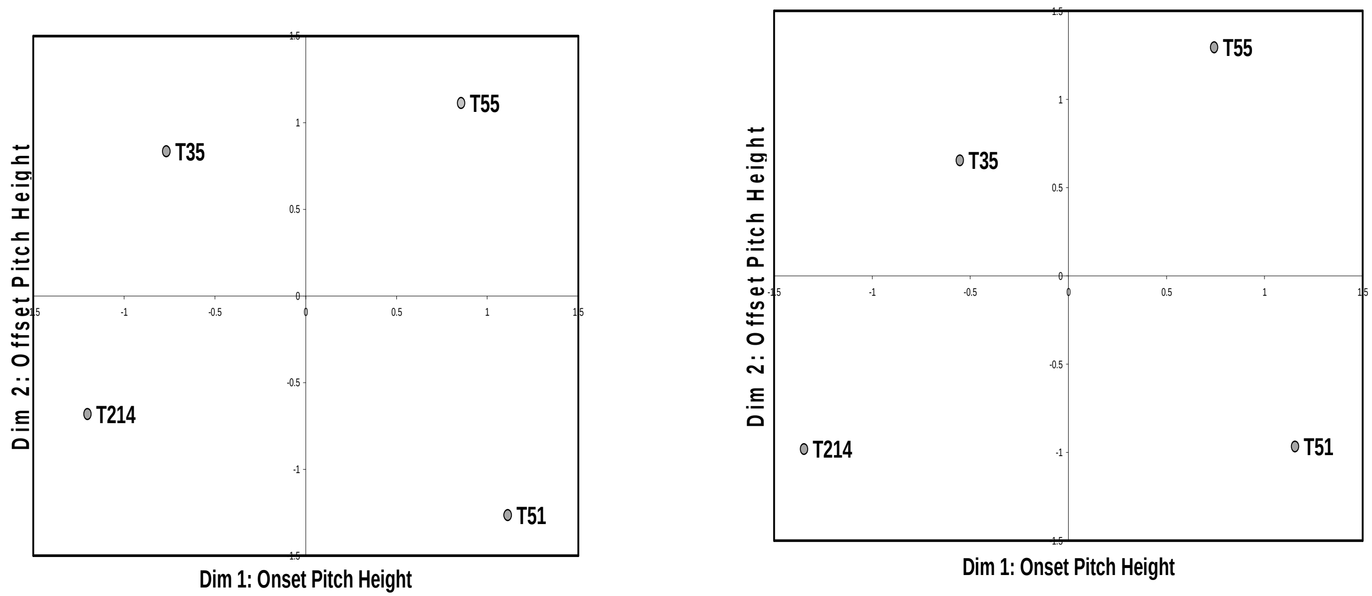
For the English group, at the same significance level, the most dissimilar pairs are T55/T214 and T214/T51, and the most similar pairs are T55/T35, T55-T51, T35/T214, and T35/T51. Pair T51-T55 falls in the middle of the scale but is not significantly different from pair T51-T35, so it may also be grouped with the more similar group.

As found in Experiment 1, the common characteristic shared by the pairs deemed more similar by the English listeners is again matching f0 onset and/or offset values, which explains why pairs T55/T51 and T35/T51 were rated as quite similar by the English listeners. On the other hand, these same pairs were among the more dissimilar to the Chinese listeners, as the contour shapes are very different for the tones involved in these tone pairs (level for T55, rising for T35, and falling for T51). When the pitch onset and/or offset values are different, as in the case of pairs T55/T214, the English listeners perceive the tones as being the most dissimilar, although these pairs do not stand out in the Chinese listeners' data. This again is consistent with the difference found between the two language groups in Experiment 1.

3.3.4.4 MDS

In the INSCAL analysis performed on the rating data, mean rating values of the tone pairs were entered in each subject's matrix. The data were taken as dissimilarities, as the experiment was set up such that the more different the two tones, the higher the rating. We provided a scale of "1" through "5" to be used by all listeners in the rating task, which made it possible to compare the ratings across the matrices unconditionally in the MDS analysis. The stimulus spaces for the two groups of listeners are shown in Figure 3.8. The Chinese listeners' space (Figure 3.8a) has a stress of .174 (Kruskal's stress formula 1 value) and a RSQ of .899, while the English listeners' (Figure 3.8b) stress level is .167 and RSQ is .909.

It should be noted that not all the Chinese listeners' data were included in the MDS analysis reported here, as the analysis with all data included revealed some very high weirdness numbers. Upon checking the individual spaces of these listeners, it was found that four (4) of them probably treated T214 as a low level tone, placing it diagonally from T35 and very close to T55 along one dimension. As a result, "contour shape" is an important dimension in their spaces, while "start pitch height", which was found to be the slightly more important dimension in Experiment 1, did not show up at all. These listeners' data, along with one other dataset that contributed almost exclusively to one dimension, were eliminated from the present analysis.



a. b.

Figure 3.8 The Chinese (panel a; stress = 0.174, RSQ = 0.899) and English (panel b; stress = 0.167, RSQ = 0.909) listeners' MDS tone spaces derived from the difference ratings.

In the Chinese listeners' space, dimension 1 seems to correspond to "onset pitch height", while dimension 2 "end pitch height". We may interpret the American English listeners' space in a similar way. In terms of subject weights, the English listeners' dimensional ratio is .5671: .3423 (or 1.66 : 1). The dimensional weight difference was less pronounced for the Chinese listeners: the first dimension accounts for .4672 of the variance and the second dimension .4319, or a ratio of 1.07 : 1. The English group had very low weirdness numbers, which were "0" or near zero for quite a few listeners.

In general, the patterns in the MDS spaces are similar to those seen in the rating plot (Figure 3.7). T35 and T214 are closer together as are T55 and T35 in both spaces, while T55 and T214 are farther apart as are T51 and T214 in both. The distance between T35 and T51 is longer while that for T214 and T51 is shorter for the Chinese listeners. in addition, the relative distance between T35 and T51 in the Chinese listeners' tone space is noticeably loner than that in the English listeners' space. Thus, the differences between the two groups as revealed in the rating plot and by the T-Test are basically captured as well: the Chinese listeners did not seem to pay as much attention to the transition from the f0 offset of the first tone to the onset of the second tone as did the English listeners; instead, contour shape was more important for the Chinese listeners.

3.4 Summary of Discussion on Experiments 2, 3 and 4

The AX difference rating task (Experiment 4) yielded similar results to the AX discrimination test reported in Chapter 2 (Experiment 1) in that in both experiments the Chinese listeners' tone perception was influenced by the tone sandhi rules in their language (the T214 rule in Experiment 1, and the T214 rule and the T35 rule in

Experiment 4). This influence of phonology on perception seems quite remarkable in strength, because even a simple AX limited stimulus set discrimination using natural speech stimuli (Experiment 2) did not take it away, although there was evidence that the Chinese listeners' attention was directed more toward the acoustic characteristics of the tones than in the previous experiments. They behaved even more like the English listeners in the limited stimulus set AX discrimination test using sine wave tones (Experiment 3). Imaginably, with the segmental makeup taken away and with just four stimulus tone tokens repeated over and over again in Experiment 3, it was easy to focus attention on the acoustic properties of the stimuli, rendering the task mainly a psychoacoustic one. Although some of the Chinese listeners reported that they heard tones in Putonghua in this experiment using sinewave tones, it is rather doubtful that there was lexical activation involved in this task (Johnson, 2004). Nevertheless, even in this experiment using synthetic stimulus tones, there were still some differences in how the two groups reacted to certain tone pairs, namely T55-T214, T51-T55, T51-T214 (shorter RTs for the English than for the Chinese), and T35-T51 (longer RTs for the English). And as the MDS stimulus spaces show, even in this simple task, the relative distance between T35 and T214 is still somewhat shorter in the Chinese space than that in the English listeners' space. The AE listeners' showing similar perceptual patterns in these different tasks suggests that the experiments were mostly psychoacoustic for these nonnative listeners.

CHAPTER 4

TONES AND TONE SANDHI PROCESSES IN THREE MANDARIN DIALECTS

The experimental results reported in Chapters 2 and 3 point to a strong tone sandhi effect on tone perception by native Chinese listeners, in comparison with perceptual performance of the control group of American English listeners. It is well known that tonal inventories and sandi processes differ considerably across the Chinese dialects (Chen, 2000; Duanmu, 2000). Conceivably, we could also find inter-dialectal perceptual differences. Past studies (e.g. Gandour, 1983, 1984) found some differences in perceptual performances by listeners of different dialectal backgrounds, namely Mandarin, Cantonese and Southern Min. However, these studies used non-speech synthetic tones and were not designed to test the sandhi effect directly. Furthermore, as reported in Chapter 3, synthetic stimuli may "de-link" auditory perception from linguistic information.

In terms of number of tones in the inventory, in general Chinese dialects in the south tend to have more tones than those in the north (see, e.g. Lien, 1986). For instance, there are usually six (e.g. Rugao, a Jianghuai Mandarin dialect; Ting, 1966; RGXZ, 1994; Huang 2002) to eight tones in the southern Mandarin and Wu dialects spoken in southern

Jiangsu Province, Shanghai City and Zhejiang Province (e.g. Chao, 1928) and the Min dialects spoken in Fujian & Taiwan (see, e.g., Norman, 1973). One Cantonese dialect was reported to have ten tones (Zong, 1964).[1] On the other hand, the northern dialects usually have four tones. Some dialects in Shandong Province (Qian et al., 2001) have only three underlying tones.

In terms of tone sandhis, in addition to stress related neutralization rules, Putonghua (or Standard Mandarin) has only one obligatory category-changing process involving normal stressed syllables, namely the T214 sandhi, probably two for speakers of Putonghua from Beijing, who also have the T35 sandhi described in Chapter 2. Other dialects may have more such processes, as in the northern Mandarin dialects of Yantai (Shandong Province; Qian, 1982) and Tianjin (Tianjin City; see, e.g. Tan, 1987). Still others may have no category-changing neutralization rules, as in the Jianghuai Mandarin dialect of Rugao (e.g., Ting, 1966; RGXZ, 1994) and the neighboring dialects such as Yangzhou (Wang & Huang, 1996). Sandhi processes in the Wu or Min dialects could be very different in nature (see, e.g. Duanmu, 1997; Chan, 1989).

In order to make the perceptual data more comparable across dialects, we should control the differences in tonal inventories and in the type and number of sandhi rules operating in the dialects. To do that, we need to choose sub-dialects from within a major dialect group, e.g. Mandarin, instead of representative dialects from three major dialectal groups as those used in Gandour's (1983, 1984) studies.

---

[1] In these fairly large inventories, people have counted tones on syllables checked with an obstruent coda as separate categories, although sometimes these tones are similar in contour shapes to and are just shorter in duration than the non-checked tones in the same dialect.

Thus, a series of experiments using natural speech stimuli from three Mandarin dialects, namely Beijing, Rugao and Yantai were designed to test possible language-specificity effects among these dialects. The descriptive sketches for tone inventories and historical tone developments in these dialects, as well as synchronic correspondence among them, are provided below.

4.1 Tones in the three Chinese dialects of Beijing, Rugao and Yantai

4.1.1 Tones in Beijing/Putonghua

The Beijing dialect is the basis for Putonghua and is basically seen as the same system as Putonghua (except for maybe some colloquial lexical items and the more extensive use of the suffix /-r/ to denote diminutive forms). As mentioned in §2.1, it has four tones, namely T55, T35, T214 and T51, which essentially form two major contrasts of high vs. low (i.e. T55 vs. T214) and of rising vs. falling (i.e. T35 vs. T51). Figure 3.1 is repeated here as Figure 4.1 for easier comparison with tones in the other two dialects.

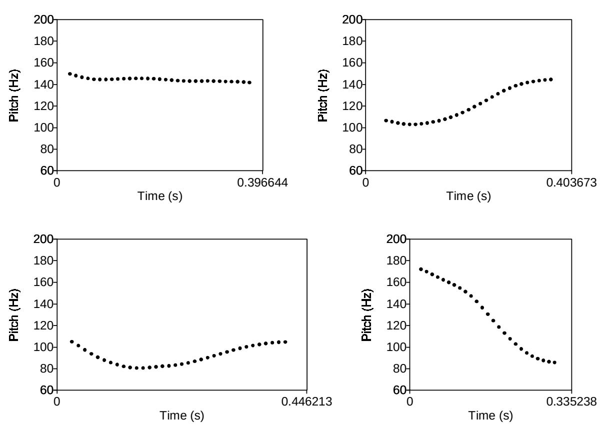


Figure 4.1 F0 traces of the citation tones T55 (upper left panel), T35 (upper right), T214 (lower left), and T51 (lower right) as produced by the male Beijing speaker on the /ba/ syllable. Lengths of the X-axes in these panels reflect approximately the relative lengths of the tones.

4.1.2 Tones in Rugao

Rugao has been classified by most researchers (e.g., Li, 1989a,b) as a Jianghuai Mandarin dialect, also called a Lower Yangtze River dialect, although Ting (1966) suggests that it may also have some Wu dialect elements in it. Previous studies on Rugao posit that the language has either six tonemes (e.g., JSFG, 1960; Ting 丁 1966; RGXZ, 1994; and Wú 吴 , forthcoming), or four tonemes in an analysis that equates the tones in checked syllables with two of the four tones in syllables with sonorant codas based on their distribution patterns (Huang 1999, ms.). As found in Huang (1999, 2002), the checked tones have the exact same tonal contours as two of the unchecked tones but are shorter in duration. Analyses positing six tones for Rugao treated the syllables with a [-ʔ] coda – historically [-p, -t, -k] codas – as having different tones, namely the "entering" (or "checked") tones (入声). As entering tones were excluded from the present study, we shall follow Huang (1999) in positing four tones for this language. Following the Chinese linguistics tradition (see, e.g., Norman 1988), we shall name them Tones 1, 2, 3, and 5 for now. The shapes of these tones are Falling, Rising, Low, and High, respectively. The Low tone (tone 3) may also have a shallow dip in the middle. Descriptions using Chao's (1930) five-level tone transcription system can also be found in the afore-mentioned studies. In Ting's (1966) impressionistic study, the Rugao tones are described as /21, 35, 213, 44/, respectively.

Based on f0 measurements for the tone contours in my small recorded database of 10 speakers from the town of Rugao, the tones seem to be better described as /41, 45, 212 and 44/ (see Figure 4.2). However, due in part to the fact that falling contours need to have larger rate of change, as compared with rising contours, to achieve a certain level of perceived contour prominence (see, e.g., Gandour 1978: 58), the falling Tone 1 sounds fairly low. (This was probably why Ting's (1966) impressionistic study labeled it /21/, or even /11/.) As in Beijing/Putonghua, the rising Tone 2 also has a shallow dip very early in the contour. Tone 3 is fairly low with a very shallow dip. The steady part of the high level Tone 5 occupies the middle to upper of the speaker's pitch range. Using the letter R to stand for Rugao tones, I shall transcribe the tones as follows:

(4.1) The Rugaohua Tones

Falling	Tone 1	/R41/
Rising	Tone 2	/R45/
Low	Tone 3	/R212/
High	Tone 5	/R44/

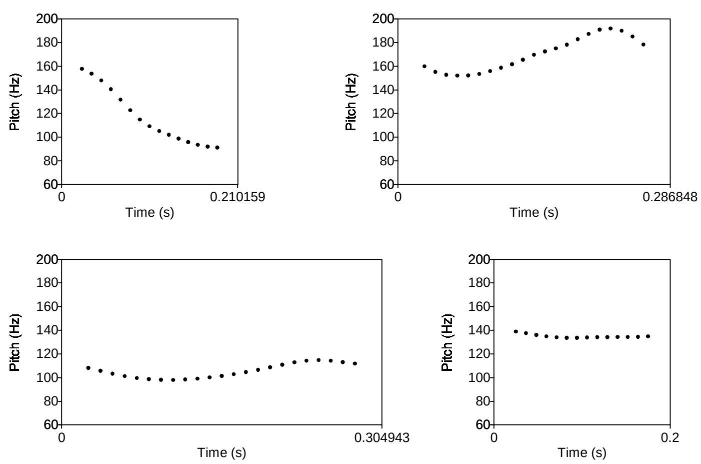


Figure 4.2 F0 traces of the Rugao tones R41 (upper left panel), R45 (upper right), R212 (lower left), and R33 (lower right) as produced by a male Rugao speaker on the /sa/ monosyllable. Lengths of the X-axes in these panels try to reflect approximately the relative durations of the tones.

4.1.3 Tones in Yantai

The Yantai dialect has only three citation tones and four phonetic tones in connected speech (Qian, 1982; Qian, Zhang, & Luo, 2001; Chen, 2000: 99-100). The three underlying tones are described as /31, 214, 55/ (Qian, 1982: 15; see Figure 4.3). Tone [35] is added as a surface tone by rule (i) below. I shall refer to the tones as /Y31, Y214, Y55/, where "Y" stands for tones in Yantai. As can be seen from Figure 4.3, these numerical descriptions are not to be taken at face value. The onset of the falling tone Y31 is actually about 20Hz higher than the so-called high level Y55, which is only about 20Hz higher in overall pitch than the low tone.

(4.2) The Yantaihua Tones

Falling	Tone 1	/Y31/
Low	Tone 2	/Y214/
High	Tone 3	/Y55/

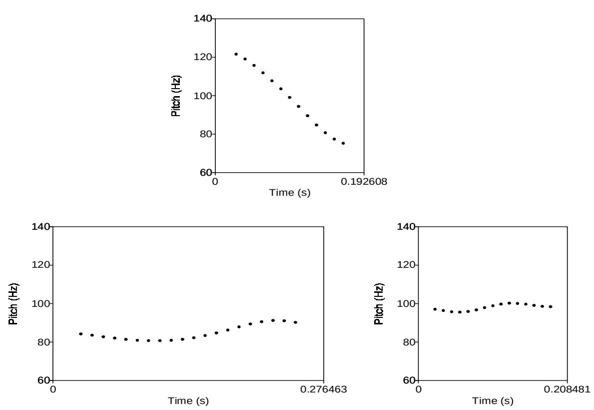


Figure 4.3 F0 traces of the Yantai stimulus tones Y31 (upper panel), Y214 (lower left), and Y55 (lower right) as produced by a male Yantai speaker (in his early twenties). Lengths of the X-axes in these panels try to reflect approximately the relative durations of the tones.

4.2 Historical Development and Synchronic Correspondence of Tones in Beijing (Putonghua), Rugao and Yantai

Since most segmental and tonal distinctions found in modern Chinese dialects, can be traced back to no further than Late Middle Chinese (see Tai & Chan 1999 for a nice summary on periodization of Chinese) around the 10th century A.D., as recorded in the poetic rhyme table tradition, we may assume that Middle Chinese (MC) is ancestral to most Chinese dialects, except for maybe the Min group which shows some distinctive characteristics predating MC (Downer 1963; Norman 1973; Chan, 1985; Ting 1989; Baxter 1992). Thanks to a long tradition of written literature and of Chinese philology, one can say with good confidence that there were four tonal categories in MC (601 A.D., year of publication of Lu Fayan's 陆 法 言 *Qie Yun* 切 韵, to 1278 A. D., end of Song dynasty), namely, Ping 平 "Even/Level", Shang 上 "Rising", Qu 去 "Departing", and Ru 入 "Entering" (on syllables checked with voiceless stop codas).

There is a consensus among scholars working on Chinese historical phonology that MC had a voiced vs. voiceless distinction in onset obstruents and that their later merger conditioned the great Yin (upper) versus Yang (lower) tonal register split across the Chinese dialects (e.g., Maspero, 1912; Haudricourt, 1954a,b; Downer, 1963; Cheng, 1973; Chang, 1975; Chen, 1976; Li, 1980; Wang, 1983; Ho 1994).

We shall follow one convention in the Chinese linguistics tradition (see, e.g., Chan 1989) in using **I**, **II**, **III**, and **IV** to stand for the four MC tone categories, and **a** and **b** for the Yin and Yang registers, respectively. Thus, a Yin register Ping tone will be labeled **Ia**, a Yang register Ping tone **Ib**, and so on and so forth. Some examples are given below to illustrate the correspondence of modern tones to historical tonal categories in Rugao, Yantai and Beijing (Putonghua). MC tone categories are based on *Ju Song Guang Yun* 钜 宋 广 韵 (Shanghai Library, 1980).

The cross-dialectal sound and tone category correspondences between Rugao and Beijing are fairly regular and can be expressed with just a few rules. It suffices for our purpose to state a simplified one-to-one mapping, where /R41, R45, R212, R44/ correspond to /T55, T35, T214, T51/ in Beijing, respectively (see (4.3)).

While cross-dialectal sound correspondence between Yantai and Beijing can be fairly easily established, the tonal correspondence is a bit more complicated due to historical category mergers in Yantai. A simplified – but sufficient for our purposes – account would be to say that /Y31, Y214, Y55/ correspond to /T55, T214, T51/ in Beijing, and that the Yantai category that would have corresponded to Beijing T35 has merged with /Y55/ (see (4.3)).

We should note that because Putonghua (Beijing) is used in the mass media, even people with only a passive knowledge of the standard dialect seem to be aware of these (especially tonal) correspondences between their native dialect and Putonghua. As we shall see in later chapters, such knowledge may also influence their tone perception.

It can also be seen from the f0 traces given in Figures 4.1, 4.2 and 4.3 that there are similarities in the synchronic tone shapes among the three dialects. Beijing and Rugao both employ contrast between high (T55 or R44) and low (T214 or R212) tones as well as between rising (T35 or R45) and falling (T51 or R41) tones. The Yantai tones are very similar to those in Rugao, except that the rising tone is missing due to the historical tone merger mentioned above. There is also a speaker pitch range difference, with the Yantai speaker having the smallest pitch range, which will probably have an effect on tone perception.

(4.3) Reflects of MC categories in modern Rugao, Yantai and Beijing/Putonghua as well as tonal correspondence among the three dialects The Yantai data were checked against Qian's (1982) and Qian et al.'s (2001) descriptions of Yantai.

	Beijing	Rugao	Yantai	Gloss & Character
Ia	**tau^{55}**	**$tɔ^{41}$**	**tao^{31}**	**"knife"** 刀
	t^hau^{55}	**$t^hɔ^{41}$**	**t^hao^{31}**	**"overflow"** 滔
Ib	**t^hau^{35}**	**$t^hɔ^{45}$**	**$t^hɔ^{55}$**	**"peach"** 桃
IIa	**tau^{214}**	**$tɔ^{212}$**	**tao^{214}**	**"island"** 岛
	t^hau^{214}	**$t^hɔ^{212}$**	**t^hao^{214}**	**"to punish"** 讨
IIb	tau^{51}	$t^hɔ^{41}$	tao^{55}	"road" 道
IIIa	**tau^{51}**	**$tɔ^{44}$**	**tao^{55}**	**"to arrive"** 到
IIIb	tau^{51}	$t^hɔ^{41}$	tao^{55}	"to rob" 盗
IVa	tu^{55}	$tɔʔ^{4}$	tu^{214}	"to supervise" 督
	t^hu^{55}	$t^hɔʔ^{4}$	t^hu^{214}	"bald" 秃
	(kuo^{35}	$kɔʔ^{4}$	kuo^{214}	"country" 国)
	(pei^{214}	$pɔʔ^{4}$	po^{214}	"north" 北)
IVb	tu^{35}	$t^hɔʔ^{45}$	tu^{55}	"poison" 毒
	(ly^{51}	$lɔʔ^{45}$	lu^{214}/ ly^{214}	"green" 绿)

Of interest to the present study are the modern cross-dialectal correspondences in the MC I, II and III tone categories. As can be seen from (4.2), all three dialects have a modern category corresponding to MC tone Ia: R41 – Y31 – T55; and similarly for MC IIa: R212 – Y214 – T214. But for MC tones Ib and IIIa, there is only one modern reflex in Yantai: R45 – Y55 – T35; and R44 – Y55 – T51. We may notice that MC tones IIb and IIIb have the same modern reflex as IIIa in Yantai (Y55) and Beijing (T51) but that they have the same reflex as Ia in Rugao (R41). Researchers on historical Chinese phonology believe that register split happened in MC category II earlier than in the other categories and that after the split, *tone IIb merged into III. Later, register split in MC category III took place in some dialects (e.g., Rugao) but not others (e.g., Yantai and Beijing) (see, e.g., Lian, 1986; Chan, 1983). In Rugao, MC tones IIb and IIIb then merged with Ia. As a result, while the predominant tone category correspondences can be established as mentioned above (and highlighted in (4.2)), a small subset of modern Rugao R41 morphemes may also correspond to T51 in Beijing.

4.3 Tone Sandhi Rules in Beijing, Rugao and Yantai Chinese

4.3.1 Tone sandhis in Beijing

Tone sandhis in Beijing/Putonghua have already been described in §2.1. We repeat the T214 sandhi here: when two T214 syllables occur consecutively, the tone on the first syllable changes to [T35], or /T214.T214/ → [T35.T214]. In addition, /T35/ may be realized as [T55] when occurring as the middle of a three-syllable sequence after T55 or T35 and before a non-neutral tone.

4.3.2 Tone sandhis in Rugao

Tone sandhi processes have been observed in Rugaohua (e.g. Ting, 1966; Wú, forthcoming). For example, a prosodically weak syllable, which may or may not have an underlying tonal specification, can get its surface tone from a preceding full-toned (prosodically strong) syllable as a result of rightward tonal spreading. In another process, the checked rising tone may change into a falling tone when followed by another rising tone But these sandhi processes, especially the latter type, are probably irrelevant for the speech perception experiment to be reported in Chapters 5 and 6, as (i) tone stimuli used in the present study were recorded as monosyllables with primary stress, and (ii) the checked tones were not included in the stimulus set. Phonetic tonal realizations under the influence of a preceding tonal target in Rugao are rather similar to those found in Beijing. For instance, all preceding tones with a L target cause downstep in the following tone (Huang, 2002; Shih 1988; Xu, 1997, 2001). Thus, sandhi rules are very limited in number in Rugao. More importantly, this dialect has no neutralization rules similar to the T214 rule in Beijing. It is predicted that, all else being equal, the Rugao listeners should have a more dispersed tone space than Beijing listeners.

4.3.3 Tone sandhis in Yantai

In contrast, Yantai has a rich tone sandhi system, as illustrated by the tonal realization rules in (4.4) below.

(4.4) Yantai disyllabic tonal realization rules (Qian, 1982)

(i) /31.31/ → [35.31] ← /214.31/ (iv)
(ii) /31.31/ → [55.31] ← /55.31/ (vi)
(iii) /31.214/ → [31.214]
(v) /31.55/ → [31.55] ← /55.55/ (x)
(vii) /214.214/ → [55.214] ← /55.214/ (ix)
(viii) /214.55/ → [214.55]

As can be seen from the rules above, dissimilation happens whenever two consecutive syllables have the same tone. As a result, before /31/, underlying /31/ and /214/ neutralize into [35] (rules (i) and (iv)), and /31/ and /55/ neutralize into [31] (rules (ii) and (vi)); before /55/, underlying /31/ and /55/ neutralize into [55] (rules (v) and (x)); and before /214/, underlying /214/ and /55/ neutralize into [55] (rules (vii) and (ix)).

Rules (i) and (ii) require some explanation. Rule (i) should be seen as more dominant between the two, while Rule (ii) might have a lesser effect in comparison with rule (i), for rule (ii) only affects some sonorant-initial /Y31/ morphemes. There is a further complication with rule (ii): some of the sonorant-initial /Y31/ morphemes can also be pronounced with tone /Y55/ (Qian, 1982: 16).

From the results of the experiments reported in the previous two chapters, it is evident that sandhi rules leading to paradigmatic contrast neutralization shorten perceptual distance between two otherwise distinctive tones. It seems reasonable to hypothesize that when tested with the same set of stimuli, listeners whose dialect had numerous sandhi rules leading to neutralization of tonal contrast in different environments would have a smaller tone space, relative to that of a group of listeners whose dialect has fewer such rules. We shall test this hypothesis with Yantai (with the most neutralization rules), Beijing (having one neutralization rule) and Rugao (with no neutralization rule) by asking listeners who are native speakers of one of these three dialects to participate in all three experiments using natural speech monosyllables recorded by a male Beijing (in his early thirties), a male Rugao speaker (in his late thirties) and a male Yantai speaker (in his early twenties), respectively. The number of contrastive tones may also have an impact on perception. The Yantai listeners, with only three underlying tones and a surface [35] (derived from underlying /Y31/ or /Y214/), may not a have separate category for a rising tone. As a result, they may not discriminate the acoustically similar T35 and T214 as well as do the Beijing and Rugao listeners.

CHAPTER 5

A CROSS-DIALECTAL STUDY OF CHINESE TONE PERCEPTION (I): PROCEDURES, ERROR RATES AND ORDER EFFECTS

This chapter and the next describe a set of three cross-dialectal listening experiments using natural speech stimuli from Putonghua (Beijing), Rugao and Yantai. As mentioned in Chapter 4, the purpose of this study is to test the hypothesis that differences in tonal inventories and tone sandhi rules may lead to different perceptual performances by listeners who speak one of the three dialects natively. Also to be tested are the training effects as reported by Werker & Logan (1985) and a speeded response effect as suggested by the results in Fox (1984).

5.1 Participants in Experiments BJ, RG and YT

Twenty-four (24) listeners from Beijing, forty-eight (48) listeners from Yantai, and forty-eight (48) listeners from Rugao participated in the experiments. The Beijing listeners were high school students from Tsinghua Affiliated Middle School. The forty-eight Yantai listeners can be divided into two age groups, namely young and old. The young group (26 listeners) included high school students and people in their twenties or early thirties whose education was conducted in Putonghua. The 22 listeners in the older group were people in their forties, fifties or early sixties and were not educated in Putonghua. The forty-eight Rugao listeners form two age groups as well. The listeners in the young group were high school students whose classroom language was Putonghua, and those in the older group were people in their fifties or forties and were not educated in Putonghua. Different age groups from Yantai and Rugao were included to see whether learning Putonghua as a second language affects the listeners' perception of the tones in Putonghua as well as in their own dialects. All listeners were paid a small amount of money for their participation.

5.2 Stimuli

The Putonghua stimuli were the same natural speech /ba/ monosyllables used in Experiments 2 and 3. For Rugao and Yantai, monosyllabic tonal minimal sets were recorded ten (10) times. The segmental makeup is /ba/ for Yantai and /sa/ for Rugao. It would be ideal if we could keep the segmental makeup the same across all three dialects. Unfortunately, historical sound changes have created different systematic gaps in Putonghua (Beijing), Yantai and Rugao. The five (5) most typical productions were selected as stimuli. Since Putonghua and Rugao each has four (4) tones, there are $4 \times 4 = 16$ pairs of tones in the experiments involving stimuli from these two dialects, while in the experiment using stimuli from Yantai, which has only three citation tones, there are only $3 \times 3 = 9$ pairs.

5.3 Procedures

The procedures for this series of experiments were basically the same AX discrimination task as was used in Experiment 3, except that a roving – instead of fixed order – test was involved. That is, all tone pairs were tested in random order in each block. Block order was also randomized across listeners. Listeners were tested two at a time in a quiet room in three different cities in China, namely Beijing, Yantai and Rugao. All listeners participated in all three experiments within the same hour, with short breaks in between if the listener needed a rest. The experiments using Beijing and Rugao stimuli took twenty (20) minutes each, while the experiment using Yantai stimuli took about ten (10) minutes.

Werker & Logan (1985) found a "training" effect for acoustically non-identical pairs. That is, listeners seemed to have perfected their discriminating skill with each successive block of stimuli presentation, resulting in better performance for the repeated stimulus pairs as well as for new stimulus pairs that were presented later than others. In our present study, listeners heard the stimuli in a Latin Square fashion. That is, listeners heard stimuli from the three dialects in three different orders, namely Beijing-Rugao-Yantai, Yantai-Beijing-Rugao, and Rugao-Yantai-Beijing, which were counterbalanced for each listener group and for both genders. Because the experiments were run in a Latin Square fashion, we cannot use numbers 5, 6 and 7 to refer to them, as those numbers would indicate a misleading fixed order. Instead, I shall use two-letter abbreviations for the stimulus dialect and call them Experiment BJ, Experiment RG, and Experiment YT. When an experiment (e.g., BJ) was run as the first, second (i.e., after YT), or third (i.e., after both RG and YT) in the series, it will be called BJ-1, BJ-2 or BJ-3, respectively.

We also encouraged speeded response in these experiments by asking people to try to respond within 500 ms. Fox (1984) found that faster response led to decreased language effects. In particular, he showed that a response latency shorter than 500 ms blocks the lexical effect on perception reported by Ganong (1980). We decided to investigate whether a speeded task would reduce lexical/linguistic effects on tone perception. We also hoped to test empirically Guenther and colleague's (Guenther & Gjaja, 1996; Guenther et al., 1999; Guenther & Boland, 2002) view of auditory warping as well as Johnson's (2004) lexical distance model of speech perception.

Instructions were given primarily in writing. All listeners read instructions in Chinese. Brief oral explanations were also given to make sure that the listeners understand the task. As I was not able to find a native Yantai speaker to run the experiments in Yantai, Putonghua was used when running the experiments in the cities of Beijing and Yantai, while the Rugao dialect was used when running the experiments in Rugao. The use of Putonghua instructions with the Yantai listeners may weaken the effect of testing in the native setting.

Both reaction time and error rates were recorded as experimental data. As in the experiments reported in the previous chapters, reaction time (RT) was measured from the onset of the second stimulus in the AX discrimination tone pair. Feedback was given to the listeners throughout the tests, with reaction time and percentage correct shown on the computer screen in front of them.

5.4 Data Analyses

Only reaction time for correct "different" responses will be analyzed using the method of repeated measures ANOVA (as well as planned comparisons with Independent Samples

T test). RTs for the "same" pairs (i.e. pairs involving the same tones as in T55-T55, R41-R41, Y214-Y214, etc.) and RTs for incorrect responses were not included in the analyses.

5.5 Error Rates

As in all the experiments reported in this dissertation, listeners heard the stimuli with no background noise at a comfortable listening level. As a result, overall error rates were fairly low in Experiment-BJ (Table 5.1). For the American English listeners' overall error rate for the 12 "different" pairs was 5.34%. The older listeners from Rugao and Yantai had the highest error rates, which were both under 7%.

In the AE listeners' data, the pairs attracting the most errors are T35-T55, T55/T51, T35/T214 and T35-T51 (6.75% ~ 8.75%), while the four pairs involving the least errors are T214/T55 (3.25%) and T214/T51 (2.25%). In other words, the high versus low contrast was the most salient for the AE listeners.

In the Chinese listeners' data, pairs T55/T51 and T35/T214 attracted the most errors. The two groups of Rugao listeners made the most mistakes in pairs T55/T51, while the other groups had slightly more errors with pairs T35/T214. Most noticeably shown in the error data in Table 5.1 is that the Rugao older listeners failed to discriminate a T55-T51 or a T51-T55 pair over 20% of the time. An equally interesting thing to notice is that the Rugao older listeners also had the lowest error rate for pairs T35/T214 among the Chinese listeners. The unusually high error rate with pairs T55/T51 can be attributed to the cross-dialectal tonal category correspondence between R44 and T51 (see § 4.2). The low error rate for T35/T214 indicates that they might not have formed the T214 sandhi rule in their passive Putonghua system. It is a "passive" system because with the most homogeneous population among the three cities, the Rugao older listeners almost never use Putonghua actively in their speech. They were never taught Putonghua in the classroom, either. It seems that their contact with the standard dialect through the mass media has only allowed them to establish tonal correspondences between the two dialects. The Rugao young listeners, who also made the most mistakes with pairs T55/T51, did much better. Obviously, the young listeners have acquired the standard dialect to a higher proficiency.

On the other hand, most errors occurred with pairs T35/T214 in the Beijing listeners' data, although the margin between the error rates for pairs T55/T51 and T35/T214 is not large. The same can be said of the two groups of Yantai listeners. This pattern was expected for the Yantain young listeners. As the second most proficient group in Putonghua, they must have acquired the T214 rule. And because I had to give them the instructions in Putonghua, the test setting was probably Putonghua (rather than their native Yantai). For the Yantai older listeners, however, we may not assume that they have also learned the T214 rule in Putonghua, as some of them do not speak the standard dialect. There may be a different cause for these errors: the Yantai rule that turns Y214 into [Y35] (before Y31).

	Overall	T55/T51	T35/T214
American	5.88%	8.75%	7.5%
Beijing	5.13%	**7.61%**	**9.35%**
Rugao (y)	5.66%	10.63%	7.71%
Rugao (o)	6.6%	**21.67%**	**6.25%**
Yantai (y)	5.22%	7.5%	8.85%
Yantai (o)	6.93%	11.05%	12.63%

Table 5.1 Overall error rates recorded for the listeners in Experiment BJ. Rates for pairs with the most errors are also included here.

Error rates are higher in Experiment YT (Table 5.2). Errors are well spread out among tone pairs in the Beijing and Yantai young listeners' data. For the AE and Rugao listeners, pair Y55-Y31 induced the most errors (24/200, 55/230 and 56/240, respectively). This is obviously due to the small pitch range in which the Yantai stimulus tones were produced. As noted in Chapter 4, Y55 is not high in pitch phonetically. If the falling contour in Y31 was not well perceived, this tone would be placed close to Y55 in the speaker's pitch range. The Yantai old listeners made the most mistakes in pairs Y55-Y31 (26/200) and Y55-Y214 (30/200). Both these pairs are involved in neutralization rules (see (4.3)): the contrast between Y31 and Y55 is neutralized before Y31 or Y55, whereas the contrast between Y214 and Y55 is neutralized before Y214.

	Overall	Y55-Y31	Y55-Y214
American	6.92%	12%	9%
Beijing	8.48%	-----	-----
Rugao (y)	11.81%	23.91%	-----
Rugao (o)	7.78%	23.33%	-----
Yantai (y)	6.53%	-----	-----
Yantai (o)	8.67%	13%	15%

Table 5.2 Overall error rates recorded for the listeners in Experiment YT. Rates for pairs with the most errors are also included here. "-----" means that the listener group did not show a high error rate for the tone pair(s) listed.

Error rates are the highest in Experiment RG using the Rugao tone stimuli, except for the AE (6%) and the Rugao older (3.99%) listeners (Table 5.3). Errors are well spread out among tone pairs in these two groups of listeners' data.

For listeners with higher error rates, namely Beijing, Yantai young and old, and Rugao young listeners, rates for pairs with the most and fewest errors are reported in Table 5.3. The two tones with the highest overall pitch, i.e. R45 and R44, attracted the most errors in the Beijing listeners' data. However, pairs R41/R44 show the highest rates for Rugao young and the Yantai listeners. For the Yantai listeners whose dialect has very similar tones to Rugao, this is probably due to the sandhi rules that neutralize the contrast between Y31 and Y55 (which are similar to R41 and R44, respectively). For the Rugao young listeners, it is probably because the rising contour in R45 is perceptually more salient than the falling R41 (Gandour, 1978).

The fact that pairs R45/R212 have the lowest error rates indicates that the T214 sandhi effect in Putonghua was not transferred to the Rugao stimuli. This is probably because the Rugao stimuli, with phonetically different tone shapes from the Putonghua stimuli, induced phonetic listening in non-Rugao Chinese (especially Beijing) listeners.

To summarize, error rates were higher for YT and RG stimuli than for BJ (Putonghua) stimuli. The RG tone contrasts, especially those between the rising R45 and R44 as well as between R41 and R44, may be less distinctive compared with the similar contrasts in the BJ stimuli. The Yantai speaker had the smallest pitch range, which may have contributed to the confusability of the YT stimuli.

	Overall	R41/R44	R45/R44	R45/R212
American	6%	-----	-----	-----
Beijing	12.6%	18.95%	25.21%	4.17%
Rugao (y)	9.03%	14.79%	-----	3.54%
Rugao (o)	3.99%	-----	-----	-----
Yantai (y)	9.84%	17.69%	15.38%	4.42%
Yantai (o)	12.89%	24.41%	18.82%	6.76%

Table 5.3 Overall error rates recorded for the listeners in Experiment RG. Pairs with the most and fewest errors for the four Chinese groups higher rates are also reported here. Missing data cells mean that the listener group did not have a higher than average error rate for R41/R44 and R45/R44 or a lower rate for R45/R212.

5.6 Order Effects

As we have used only a limited number of listeners, comparisons for perceptual patterns of the Beijing, Rugao and Yatai stimuli needed to be made within listeners. However, these stimulus sets were presented in three different orders to different listeners to counterbalance the order effect (Werker & Logan, 1985). For example, Experiment-BJ was run as the first of the three experiments for 1/3 of the listeners, second for 1/3, and third for another 1/3. Before analyzing data for each stimulus set, we shall first test for order effects within each of Experiments BJ, RG and YT.

5.6.1 Experiment BJ: no order effect or "order by tone pair" interaction for any group

Main effects were only found with the within-subject factor of tone pair in the Experiment BJ reaction time data. No main effects were found with the Latin Square order. The tone pair by order interaction was not significant either for any group of listeners. These results are reported in Table 5.4. The RT plots are shown in Figure 5.1.

Listener grp	effect	*df*	error *df*	*F*	*p*	$\eta^2_{par.}$
AmEng	tone pair	10.6	180.7	9.3	< .001	.353
	order	2	17	2.6	.105	.233
	interaction	21.3	180.7	.68	.852	.074
Beijing	tone pair	9.8	195.6	14.1	< .001	.414
	order	2	20	2.1	.154	.17
	interaction	19.6	195.6	.99	.481	.09
Rugao (y)	tone pair	8.0	167.1	8.8	< .001	.295
	order	2	21	2.4	.117	.185
	interaction	15.9	167.1	1.6	.067	.134
Rugao (o)	tone pair	7.8	156.1	22.3	< .001	.527
	order	2	20	2.3	.128	.186
	interaction	15.6	156.1	1.1	.32	.103
Yantai (y)	tone pair	9.7	222.8	11.5	< .001	.333
	order	2	23	.66	.524	.055
	interaction	19.4	222.8	.76	.754	.062
Yantai (o)	tone pair	6.9	118.0	14.1	< .001	.454
	order	2	17	.85	.444	.091
	interaction	13.9	118.0	1.5	.128	.149

Table 5.4 Repeated measures ANOVA results for the RT data from Experiment BJ. No significant order effects were found for any of the listener groups.

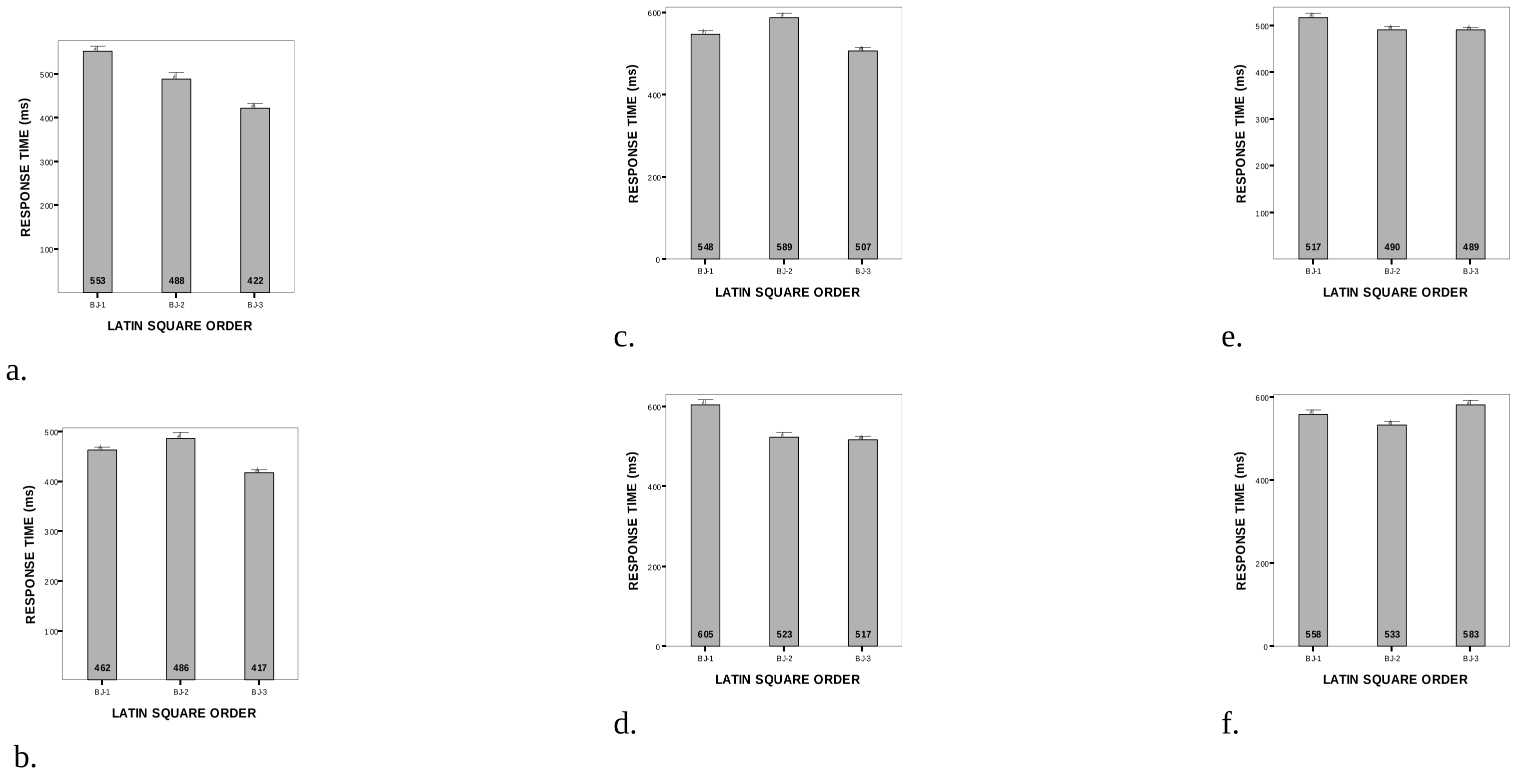


Figure 5.1 Mean reaction time plotted in the three different orders in which Experiment BJ was run. Mean RTs computed from each listener's medians for the tone pairs are shown here. The panels show American English (a), Beijing (b), Rugao young (c), Rugao older (d), Yantai young (e) and Yantai older (f) listeners' data. No significant order effects were found with any of the listener groups. Error bars show one Standard Error.

5.6.2 Experiment RG: no order effect or "order by tone pair" interaction for any group

As in Experiment BJ, main effects were only found with the within-subject factor of tone pair in the Experiment RG reaction time data. No main effects were found with order. The tone pair by order interaction was not significant for any listener group, either. These results are reported in Table 5.5. The RT plots are shown in Figure 5.2.

Listener grp	effect	*df*	error *df*	*F*	*p*	$\eta^2_{par.}$
AmEng	tone pair	10.0	170.7	2.4	.009	.126
	order	2	17	1.0	.386	.106
	interaction	20.1	170.7	1.5	.091	.149
Beijing	tone pair	8.4	176.8	11.5	< .001	.354
	order	2	21	1.4	.264	.119
	interaction	16.8	176.8	1.1	.395	.092
Rugao (y)	tone pair	10.2	213.6	9.4	< .001	.308
	order	2	21	1.4	.279	.114
	interaction	20.3	213.6	1.3	.152	.114
Rugao (o)	tone pair	10.7	224.2	18.2	< .001	.464
	order	2	21	.72	.499	.064
	interaction	21.3	224.2	.856	.65	.075
Yantai (y)	tone pair	11	253	9.4	< .001	.29
	order	2	23	1.2	.316	.095
	interaction	22	253	1.2	.249	.094
Yantai (o)	tone pair	11	209.0	7.2	< .001	.274
	order	2	19	.55	.588	.054
	interaction	22	209.0	1.2	.269	.11

Table 5.5 Repeated measures ANOVA results for the RT data from Experiment RG. No significant order effects were found for any of the listener group.

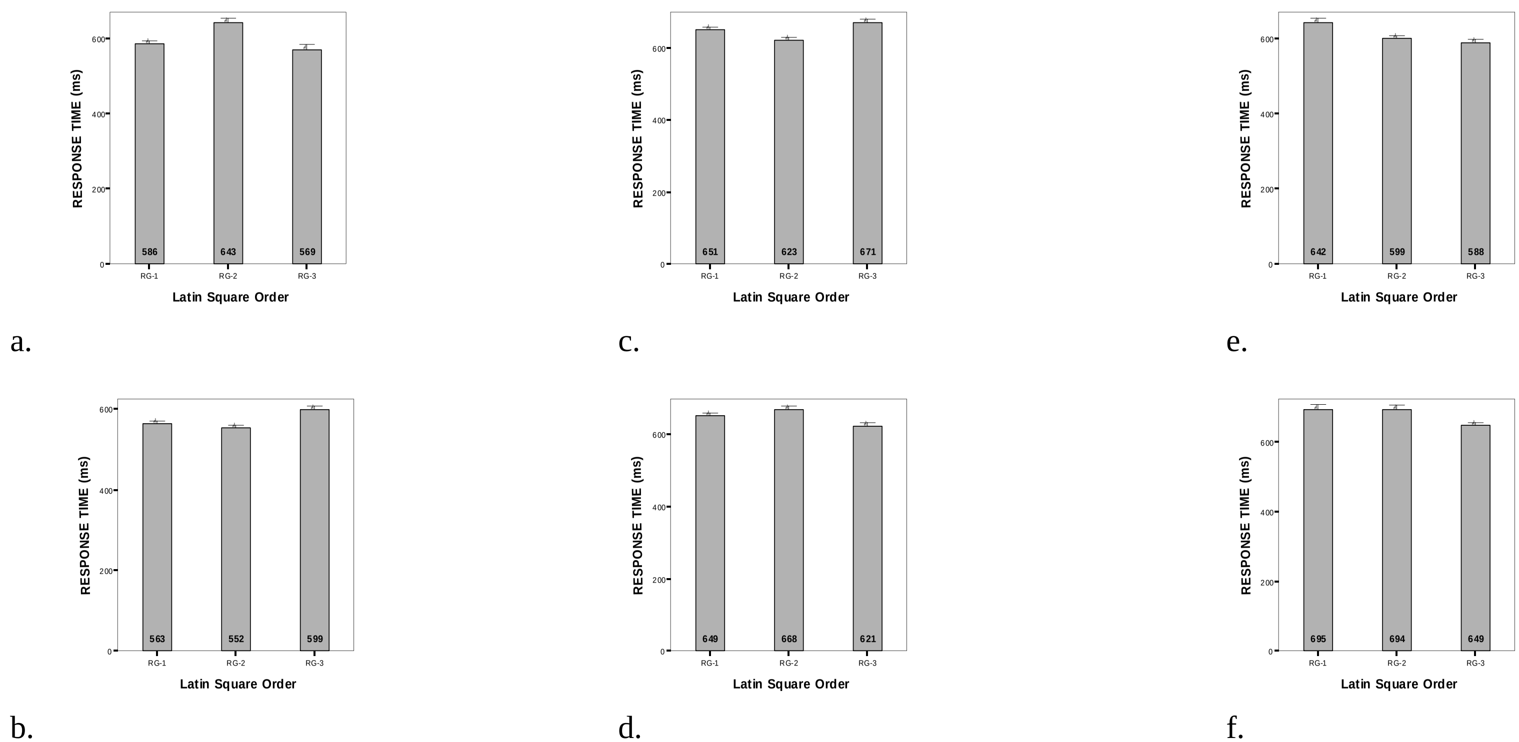


Figure 5.2 Mean reaction time plotted in the three different orders in which Experiment RG was run. Mean RTs computed from each listener's medians for the tone pairs are shown here. The panels show American English (a), Beijing (b), Rugao young (c), Rugao older (d), Yantai young (e) and Yantai older (f) listeners' data. No significant order effects were found with any of the listener groups. Error bars show one Standard Error.

5.6.3 Experiment YT: main effect with order for Beijing and Rugao young listeners

Repeated measures ANOVA only found some marginally significant between-subject effects of the Latin Square order (3 levels) in Experiment YT with the Beijing listeners' data [$F(2, 21) = 5.764$, $p = .01$, partial $\eta^2 = .354$] and the Rugao young listeners' data [$F(2, 20) = 3.556$, $p = .048$, partial $\eta^2 = .262$]. While the within-subject effect of tone pairs were also significant for both the Beijing and the Rugao young listeners ($p < .05$), the order by tone pair interaction was not significant for either group [$F(9.589, 100.68) = .805$, $p = .62$, partial $\eta^2 = .071$] for the Beijing listeners and [$F(10, 100) = .979$, $p = .467$, partial $\eta^2 = .089$] for the Rugao listeners. Post-hoc tests further revealed that the significant difference for the Beijing listeners came from between BJ-1 and BJ-3 ($p = .009$) and that BJ-2 is not significantly different from either BJ-1 or BJ-3. For the Rugao young listeners, the effect was too weak to show up in post-hoc tests, but there was a marginal effect between BJ-1 and BJ-3 ($p = .06$).

Latin Square order did not have a significant effect on the Yantai young listeners' RT data from Experiment YT, [$F(2, 23) = .63$, $p = .542$, partial $\eta^2 = .052$]. But the within-subject effect of "tone pair" was significant, [$F(4.234, 97.374) = 5.602$, $p < .001$, partial $\eta^2 = .196$]. So was the tone pair by order interaction, $F(8.467, 97.374) = 2.351$, $p = .021$, partial $\eta^2 = .17$]. Similar patterns were found with the AE listeners' RT data. No main effect was found with the between-subject factor of order, [$F(2, 17) = .902$, $p =.424$, partial $\eta^2 = .096$]. But a significant main effect was found with the within-subject factor of tone pair, [$F(5, 85) = 4.208$, $p=.002$, partial $\eta^2 = .198$]. The interaction of tone pair by order was also marginally significant, [$F(10, 85) = 2.197$, $p = .025$, partial $\eta^2 = .205$].

However, planned comparisons using independent samples T tests revealed no significant difference in RTs with any tone pair between any two orders for either group of listeners.

No main effect of Latin Square order was found in the Yantai older listeners' Experiment YT reaction time data, [$F(2, 18) = 1.044$, $p = .372$, partial $\eta^2 = .104$]. The tone pair effect was significant, $F(5, 90) = 6.549$, $p < .001$, partial $\eta^2 = 267$]. No significant effect was found with the tone pair by order interaction, [$F(10, 90) = 1.04$, $p = .417$, partial $\eta^2 = .104$]. The results are very similar in the Rugao older listeners' Experiment YT data. There was a main effect with tone pair, [$F(4.055, 85.15) = 7.118$, $p < .001$, partial $\eta^2 = .253$]. But the order effect was not significant, [$F(2, 21) = .906$, $p = .419$, partial $\eta^2 = .079$]. Nor was there a significant effect with the tone pair by order interaction, [$F(8.11, 85.15) = .322$, $p = .957$, partial $\eta^2 = .03$].

The repeated measures ANOVA results are summarized in Table 5.6. The RT data for all groups of listeners from Experiment YT are plotted in Figure 5.3. The YT reaction time data for listener groups showing significant effects in the repeated measures ANOVA, i.e. the Beijing and the Rugao young listeners, are marked with asterisks (panels *b* and *c*).

The main effects of order and tone pair as well as interaction of order by tone pair in Experiments BJ, RG and YT are summarized in Tables 5.6 and 5.7. Contrary to Werker & Logan's (1985) results, no large scale order effects were found in this series of experiments. This does not really falsify their findings, because the three stimulus sets in our experiments were produced by three speakers with very different pitch ranges. So, the listeners might have made some adjustment each time a new set of stimuli was presented.

Listener grp	effect	*df*	error *df*	*F*	*p*	$\eta^2_{par.}$
AmEng	tone pair	5	85	4.21	.002	.198
	order	2	17	.902	.424	.096
	interaction	10	85	2.20	.025	.205
Beijing	tone pair	4.79	100.7	7.71	< .001	.268
	order	**2**	**21**	**5.76**	**.01**	**.354**
	interaction	9.59	100.7	.81	.62	.071
Rugao (y)	tone pair	5	100	2.89	.018	.126
	order	**2**	**20**	**3.56**	**.048**	**.262**
	interaction	10	100	.98	.467	.089
Rugao (o)	tone pair	4.055	85.2	7.12	< .001	.253
	order	2	21	.91	.419	.079
	interaction	8.11	85.2	.322	.957	.03
Yantai (y)	tone pair	4.23	97.4	5.6	< .001	.196
	order	2	23	.63	.542	.052
	interaction	8.47	97.4	2.35	.021	.17
Yantai (o)	tone pair	5	90	6.55	< .001	.267
	order	2	18	1.04	.417	.104
	interaction	10	90	1.5	.128	.149

Table 5.6 Repeated measures ANOVA results for the RT data from Experiment YT. Significant order effects were found for the Beijing and Rugao (young) listener groups (and highlighted in bold).

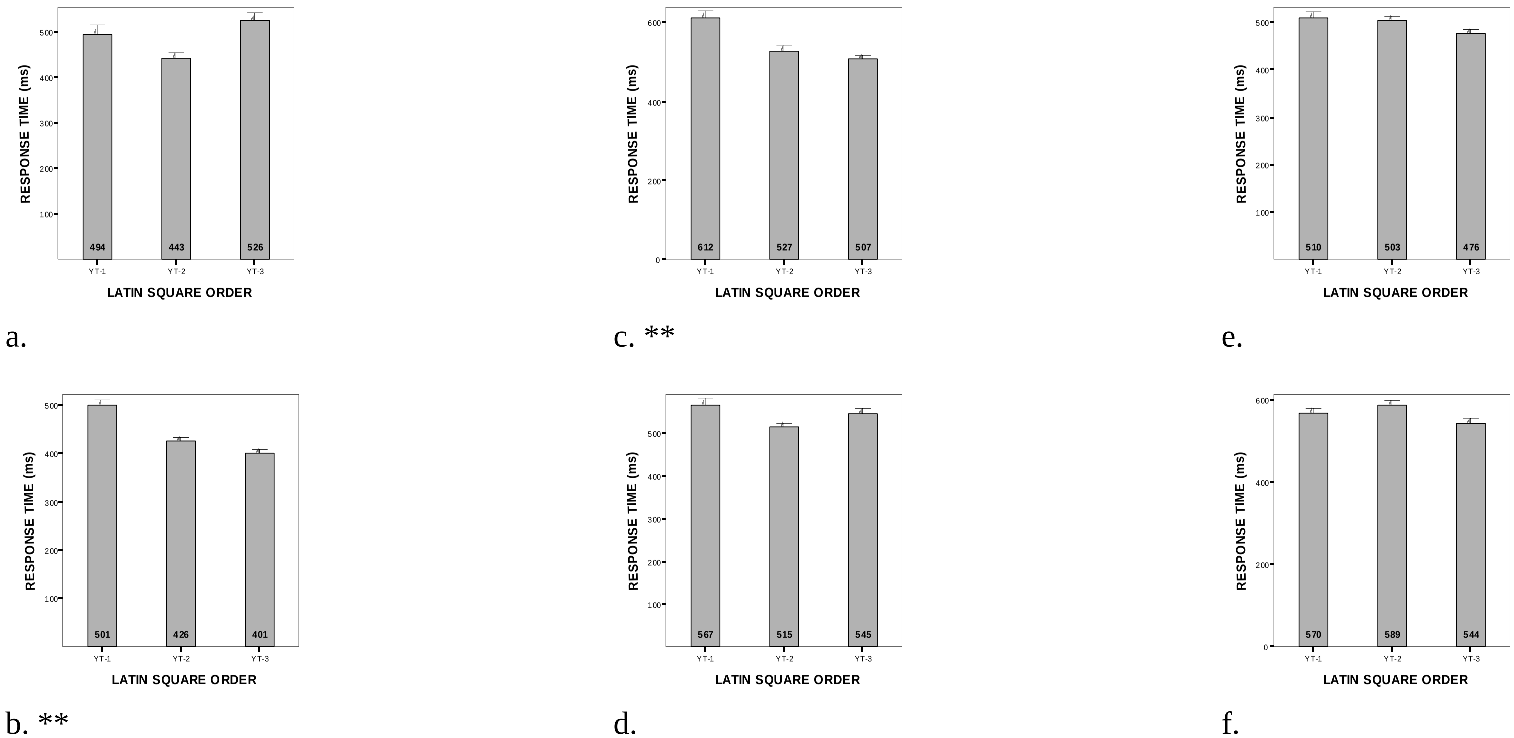


Figure 5.3 Mean reaction time plotted in the three different orders in which Experiment YT was run. Significant order effects were found between YT-1 and YT-3 in the Beijing (panel b, $p = .01$) and Rugao young (panel c, $p = .048$) listeners' RT data. No significant order effects were found with the American English (a), Rugao older (d), Yantai young (e) or Yantai older (f) listeners' data. Mean RTs computed from each listener's median RTs for the tone pairs are shown here. Error bars show one Standard Error.

Stimuli	listener grp	order effect	tone pair effect	interaction of order * tone pair
	AmEng		***	
	Beijing		***	
BJ	Rugao (y)		***	
	Rugao (o)		***	
	Yantai (y)		***	
	Yantai (o)		***	
	AmEng		**	
	Beijing		***	
RG	Rugao (y)		***	
	Rugao (o)		***	
	Yantai (y)		***	
	Yantai (o)		***	
	AmEng		**	*
	Beijing	**	***	
YT	Rugao (y)	*	***	
	Rugao (o)		***	
	Yantai (y)		***	
	Yantai (o)		***	

Table 5.7 Main effects of order and tone pair as well as interaction of order by tone pair in Experiments BJ, RG and YT. (A single asterisk * indicates $p < .05$; double asterisks ** indicate $p < .01$; triple asterisks *** indicate $p < .001$.)

CHAPTER 6

A CROSS-DIALECTAL STUDY OF CHINESE TONE PERCEPTION (II): RESULTS AND ANALYSES

Since no strong order effect was found, the reaction time (RT) data from different Latin Square orders will be pooled together. And since the listener's native language as well as listener age (or rather their educational background and L2 competence in Putonghua) may play a role in perception, data will be analyzed in four combinations using the method of repeated measures ANOVA for each of Experiments BJ, YT and RG. That is, data from the two age groups of Rugao and Yantai listeners will be compared with those from Beijing and American English listeners separately; and data from the Rugao and Yantai listeners will be compared across age groups within each dialect group. Planned comparisons with an independent samples T test will also be made for each pair of listener groups.

Recall that this series of experiments were designed to test for language-specific effects in speech perception using a psychophysical methodology that does not require listeners to categorize the stimuli in terms of the native-language tone categories. This methodology makes it possible to contrast the claims of Guenther et al. (1999) and

Johnson (2004) regarding the language-specificity of perceptual maps. Thus, the data will be analyzed in multidimensional scaling (INSCAL, Carroll & Chang, 1970) as well as the more customary hypothesis-testing statistics.

It is predicted that if an AX discrimination task with a short ISI and speeded response tap on auditory processing of speech stimuli, as claimed by various researchers (e.g. Pisoni, 1973; Carney et al., 1977; Fox, 1984; Johnson, 2004), no language-specificity would surface in these experiments.

6.1 Results and Analysis: Experiment BJ

6.1.1 Repeated measures ANOVA and planned comparisons: young listeners

The data of the young listeners from Rugao and Yantai were compared with those of the Beijing and American English listeners[1] in a repeated measures analysis of variance, with "listener language" (4 levels) as the between-subjects factor and "tone pair type" (12 levels) as the within-subjects factor. Significant differences were found for "listener language", sig.[$F(3, 89) = 5.666$, $p = .001$, partial $\eta^2 = .16$], and "tone pair type", sig.[$F(9.099, 809.771) = 36.616$, $p < .001$, partial $\eta^2 = .291$]. In addition to the main effects, there was also a significant effect with the "listener language" by "tone pair type" interaction, sig.[$F(27.296, 809.771) = 1.867$, $p= .005$, partial $\eta^2 = .059$]. Post-hoc tests and pairwise comparison showed that the between-subject effects mainly came from the significant difference ($p = .001$) between the Beijing listeners, who scored the shortest RTs, and the Rugao young listeners, who had the longest RTs among the four groups (Figure 6.1 and Table 6.1). No other two groups of listeners were significantly different from each other in overall RTs.

[1] We had only one group each of Beijing and AE listeners, all of whom are young.

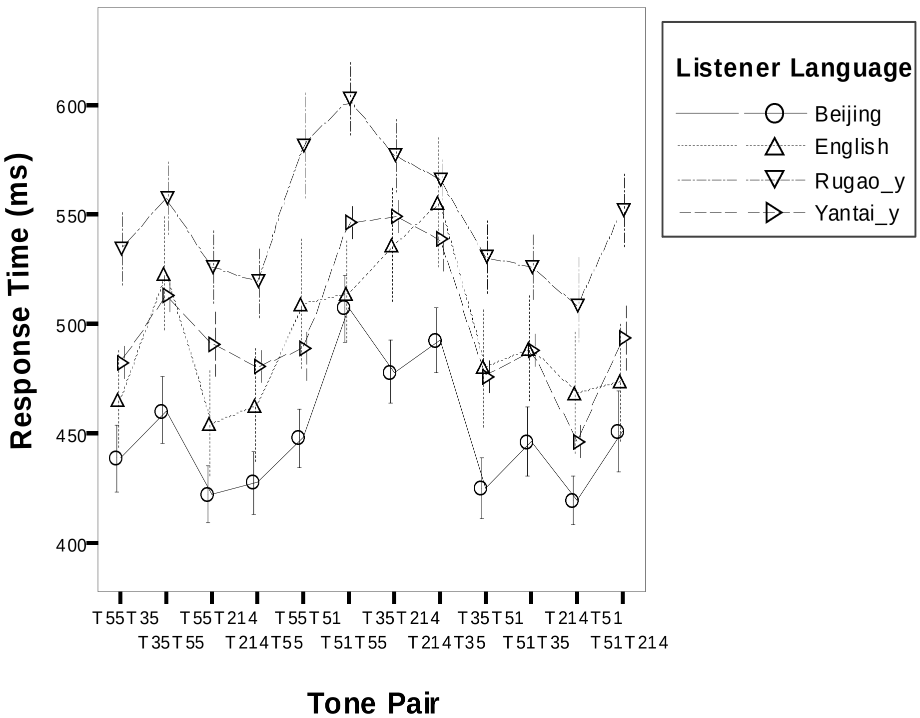


Figure 6.1 Mean response time for the Beijing, American English, Rugao (young), and Yantai (young) listeners in Experiment BJ. Significant difference was only found between the Beijing (solid line) and the Rugao (young) listeners (dashed line) in the repeated measures ANOVA. Error bars show one standard error.

Tone Pair	English	Beijing	Rugao (y)	Rugao (o)	Yantai (y)	Yantai (o)
T55/T35	495	450	546	532	497	582
T55-T35	466	438	534	513	482	566
T35-T55	523	461	557	550	512	597
T55/T214	459	423	523	517	485	526
T55-T214	455	420	526	521	490	536
T214-T55	462	426	520	512	480	516
T55/T51	512	477	591	630	517	582
T55-T51	509	448	581	611	488	557
T51-T55	514	506	600	648	546	606
T35/T214	546	483	571	586	543	599
T35-T214	535	474	576	594	548	620
T214-T35	556	491	566	578	538	577
T35/T51	485	436	528	518	482	540
T35-T51	480	424	530	518	475	545
T51-T35	489	448	526	518	488	535
T214/T51	471	435	527	515	470	515
T214-T51	468	420	503	502	446	491
T51-T214	473	450	551	527	493	538

Table 6.1 Mean RTs (in milliseconds) for correct "different" responses in Experiment BJ. These mean RT values were computed from each listener's median RT for each tone pair.

Planned comparison using the Independent samples T tests were also performed on the RT data to compare the means between Beijing and each of the other three groups. Two things can be said about the comparison between the Beijing and the American listeners: (i) equal variance cannot be assumed for any of the 12 pairs, and (ii) although no tone pairs show significant difference, effects with four pairs (T35-T55, T35/T214, and T55-T51) are marginal, with the confidence for considering RTs for these pairs the same under 10%. The standard deviation measures in the AE listeners' RT data can be twice as large as those in the Beijing listeners' data.

Planned comparisons of the Beijing and the Yantai (young) listeners' mean RTs with the T test showed significant difference or borderline effects for almost all pairs (Table 6.2). As can be seen in Figure 6.1, the peaks and valleys align very well in the Beijing and Yantai data, except for T35/T214, where the Yantai (young) listeners have a longer RT for T35-T214 (instead of T214-T35 as in the Beijing listeners' data). As a result, pair T35-T214 showed up as the most different between the two groups, [$t(47) = -3.937$, $p < .001$, $\eta^2 = .245$]. This suggests that learning the T214 rule as an explicitly taught L2 rule may have made the Yantai listeners more conscious of the rule than the native Beijing listeners.

Tone pair	*t(47)*	*p*	η^2
T55-T35	-2.179	.034	.092
T35-T55	-2.551	.014	.123
T55-T214	-3.386	.001	.196
T214-T55	-2.804	.007	.143
T35-T214	-3.937	< .001	.245
T214-T35	-2.154	.036	.090
T35-T51	-2.549	.014	.121

Table 6.2 Results of Independent Samples T Test on the RT data from Beijing and Yantai (young) listeners in Experiment BJ.

Since the Rugao (young) listeners have the slowest responses among the four groups analyzed so far, their RTs are predictably very different from the Beijing listeners. One possible explanation for why these listeners were slower is that they might have had less keyboard experience compared with the other groups of young listeners. Indeed, RTs for all pairs were found to be significantly different by planned comparisons (Table 6.3). As can be noted from Figure 6.1, RT curves for these two groups align fairly well, except that unlike any other group, T55/T51 and T35/T214 form one big peak in the Rugao data, with a slight skew toward pairs T55/T51. As in the Yantai (young) listeners' data, T35-T214 also has a slightly longer RT than T214-T35, albeit insignificant.

Tone pair	*t(45)*	*p*	η^2
T55-T35	-4.153	< .001	.277
T35-T55	-3.527	.001	.217
T55-T214	-4.687	< .001	.328
T214-T55	-4.085	< .001	.271
*T55-T51	-4.366	< .001	.298
T51-T55	-3.636	.001	.227
T35-T214	-4.352	< .001	.296
T214-T35	-3.567	.001	.220
*T35-T51	-4.330	< .001	.294
T51-T35	-3.423	.001	.207
*T214-T51	-3.362	.002	.201
T51-T214	-3.897	< .001	.252

Table 6.3 Results of Independent Samples T Test on the RT data from Beijing and Rugao (young) listeners in Experiment BJ. For asterisked (*) pairs T55-T51, T35-T51 and T214-T51, equal variance was not assumed and the degree of freedom (df) values in these cases are 33, 40 and 34, respectively.

Rugao and Yantai young listeners have very similar RT curves, except for pair T55-T51, where the peak is missing in the Yantai young listeners' data. Planned means comparisons with an independent samples T test showed only borderline effects for almost all tone pairs, except for pair T55-T51, whose RT means are significantly different for the two groups, [t (48) = 3.019, p = .004, $\eta^2 = .16$], suggesting that this pair is more confusable for the Rugao group. This is very likely due to the inter-dialectal tone category correspondence of R44 to T51.

The repeated measures analysis was also performed on the RT data for each language group separately to further explore the within-subjects "tone pair type" effect. Significant effects were found for each group ($p < .001$). To the American English listeners, pairs T214-T35, T35-T214 and T35-T55 are the most confusable (i.e. having the longest RTs) and significantly different ($p < .05$) from seven (7), six (6), and four (4) other pairs, respectively. Pairs T55/T51 fall in the middle, with T55-T51 not being significantly different from any other pair. The other pairs form the less confusable group.

For the Beijing listeners, pair T51-T55 is the most confusable and significantly different ($p < .05$) in RT from ten (10) other pairs. Pairs T214-T35, T35-T55 and T35-T214 are also among the more confusable, different from six (6), four (4) and three (3) pairs, respectively. On the other hand, pairs T55/T214, T35-T51, and T214-T51 are the least confusable, each of which differs from the four most confusable pairs.

For the Rugao (young) listeners, pairs T51-T55, T35-T214 and T214-T35 are the more confusable pairs, while pairs T214-T51, T55-T214, T214-T55 and T51-T35 are the least confusable. These eight tone pairs form two significantly different groups ($p < .05$). The other four tone pairs fall in between. Noticeably among this last group T55-T51 has a very large mean RT value. Yet due to large variance, this pair is not significantly different from any other pair.

For the Yantai (young) listeners, T35-T214, T214-T35 and T51-T55 are the most confusable and are significantly different ($p < .05$) from T214-T51, T214-T55, T35-T51, T51-T35 and T55-T35, which are the least confusable. Pairs T35-T55, T55-T214, T55-T51 and T51-214 fall in between.

	most confusable	"middle"	least confusable
English	T214-T35, T35-T214, T35-T55	T51-T55, T55-T51	T51-T35, T55-T35, T214/T51, T214-T55, T35-T51, T55-T214
Beijing	T51-T55, T214-T35, T35-T55, T35-T214	T55-T51, T214-T51, T55-T35, T51-T35	T55-T214, T214-T55, T35-T51, T214-T51
Rugao (young)	T51-T55, T35-T214, T214-T35	T55-T51, T35-T55, T51-T214, T55-T35, T35-T51	T55-T214, T51-T35, T214-T55, T214-T51
Yantai (young)	T35-T214, T214-T35, T51-T55	T35-T55, T51-T214, T55-T214, T55-T51	T55-T35, T214-T55, T51/T35, T214-T51
Rugao (old)	T51-T55, T55-T51, T35-T214, T214-T35	T35-T55	T51-T214, T55-T214, T55-T35, T35/T51, T214-T51, T214-T55
Yantai (old)	T35-T55, T35-T214, T51-T55, T214-T35	T55-T51, T55-T35, T35/T51, T51-T214	T214-T55, T55-T214, T214-T51

Table 6.4 Tone pair groupings based on their confusability as perceived by different groups of listeners in Experiment BJ. (Data from the older listeners of Rugao and Yantai – to be discussed later – are also included here.)

Table 6.4 summarizes the sub-groupings of tone pairs mentioned above. As can be seen from Table 6.4, pairs T51-T55, T35-T214 and T214-T35 are the most confusable for the Chinese listeners, while pairs T214-T35, T35-T214 and T35-T55 are the most confusable for the American English listeners. In addition to acoustic similarities, the common factor affecting all groups of listeners, the T214 tone sandhi and other tonal realization rules (e.g. the T35 rule in Putonghua and the Y31 and Y214 rules in Yantai, as well as downstep in Beijing and Rugao) may have played a role in the Chinese listeners' tone perception. For the Rugao and Yantai listeners, inter-dialectal tone category correspondences of R44 to T51 and Y55 to T51 between their native dialects and Putonghua may also have interfered with tone perception.

The differences between the Beijing and the other three groups of listeners again suggest a linguistic effect. Comparison of the Beijing data against those of the American English indicates that the longer RT with T51-T55 on the Beijing curve cannot be attributed completely to psychophysical factors. Differences between Beijing and Yantai (young) as well as those between Beijing and Rugao (young) further indicate dialectal influences on perception. The acquisition of the T214 sandhi rule by the Rugao and Yantai listeners through explicit classroom instruction appears to have made the listeners more aware of the process, leading to further shortening of perceptual distance between T35 and T214. Inter-dialectal tone category correspondences also interfere in perception, although a higher competence in and more frequent use of Putonghua may reduce the magnitude of this effect, as is evident in the difference between the Rugao young (infrequent use) and the Yantai young (frequent use) listeners' data. Higher proficiency in Putonghua has also enabled the Yantai young listeners to discriminate T55 and T51 better than the Rugao young listeners. This suggests that they were able to separate their native Yantai tone system from the L2 Putonghua system better than the Rugao listeners, as the correspondence between Y55 and T51 did not seem to interfere in their L2 perception as much as the correspondence between R55 and T51 did in the Rugao listeners' L2 perception.

6.1.2 Repeated measures ANOVA and planned comparisons: older listeners

The Rugao and Yantai older listeners' data were also compared with the Beijing and American English listeners' data. Again, the between-subject factor of "listener language" had a significant effect, sig.[$F(3, 82) = 7.06$, $p < .001$, partial $\eta^2 = .205$]. The within-subject effect of "tone pair type" was also significant, sig.[$F(8.946, 733.583) = 48.791$, $p < .001$, partial $\eta^2 = .373$]. There was also a significant effect with the interaction of "listener language" and "tone pair type", sig.[$F(33, 733.583) = 4..13$, $p < .001$, partial $\eta^2 = .131$]. Post-hoc tests show that the significant between-subject main effect came from the differences between the Beijing and the Rugao older listeners ($p = .001$) and between the Beijing and the Yantai older listeners ($p = .002$).

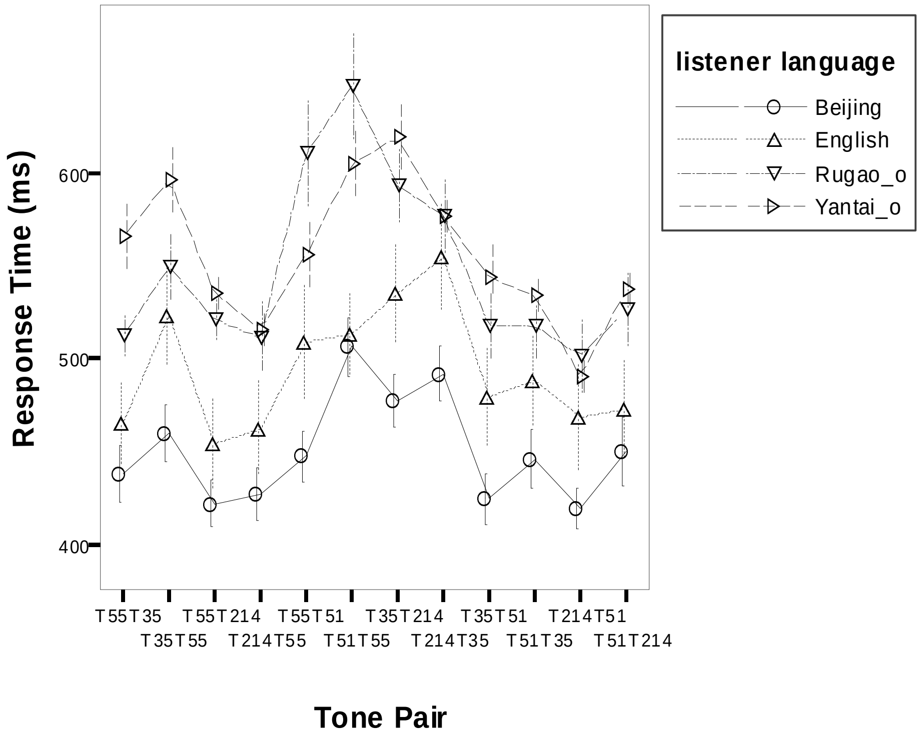


Figure 6.2 Mean response time for Beijing, American English, Rugao (old), and Yantai (old) listeners in Experiment BJ. Significant difference was found between Beijing listeners (solid line) and Rugao (old) listeners (dashed line) as well as between Beijing listeners and Yantai older listeners (dash-dotted line). Error bars show one standard error.

RT means for each tone pair in the Rugao (older) and Yantai (older) listeners' data were compared between the two groups and against the Beijing and the AE listeners' data with independent samples T tests. As shown in Tables 6.5 and 6.6, all tone pairs are significantly different between Beijing and Rugao (older) as well as between Beijing and Yantai (older). Significant differences were also found with five pairs between Yantai (older) and AE listeners (Table 6.7). Differences between Rugao and AE were limited to three pairs: T51-T55 [t (41) = 3.422, p = .001, η^2 = .222], T51-T55 [t (41) = 2.335, p= .025, η^2 = .117], and T55-T214 [t (41) = 2.229, p = .031, η^2 = .108]. Only T55-T51 has a marginally significant RT difference between Rugao and Yantai, sig.[t (41)= -2.056, p= .046, η^2 = .093].

As can be seen in Table 6.5, the largest RT differences between Beijing and Rugao (older) listeners are in pairs T55/T51 and T55-T214. As has been mentioned above, the peak with T55/T51 in the Rugao (older) listeners' data was probably caused by the R44-T51 correspondence between the two dialects. The difference with T55-T214 could be because the T214 counterpart in Rugao (i.e., R212), has a rather shallow dip if at all and is very likely distinguished from the other tones by its low feature in Rugao. As a result, the Rugao (older) speakers may have carried this strategy over to their perception of Putonghua tones without comparing difference in the T55 and T214 contours – hence, a slightly larger than average difference between the two groups for this pair. Note that pairs T35/T214 have shorter RTs than T55/T51 in the Rugao (older) listeners' data, suggesting a weak T214 sandhi effect if at all.

The most robust RT differences between the Yantai and Beijing listeners can be found with pairs T55/T35, T55-T214, T35-T214, and T35-T51 (Table 6.6). Significant differences were also found with T55/T35 between AE and Yantai (Table 6.7), suggesting that these two pairs are special for the Yantai (older) listeners.

Tone pair	*t(44)*	*p*	η^2
T55-T35	-3.252	.002	.194
T35-T55	-3.667	.001	.234
T55-T214	-4.667	< .001	.331
*T214-T55	-3.410	.001	.209
T55-T51	-4.953	< .001	.447
*T51-T55	-4.215	< .001	.347
T35-T214	-4.255	< .001	.292
*T214-T35	-3.219	.003	.210
T35-T51	-4.016	< .001	.268
T51-T35	-2.928	.005	.163
*T214-T51	-3.254	.003	.249
T51-T214	-2.619	.012	.135

Table 6.5 Results of Independent Samples T Test on the RT data from Beijing and Rugao (older) listeners in Experiment BJ. Equal variance cannot be assumed for the two groups for the asterisked (*) pairs T214-T55, T51-T55, T214-T35 and T214-T51 and degree of difference values are 30, 33, 39 and 32 respectively.

Tone pair	*t(41)*	*p*	η^2
T55-T35	-5.262	< .001	.403
T35-T55	-5.456	< .001	.421
T55-T214	-5.611	< .001	.434
T214-T55	-4.09	< .001	.290
T55-T51	-4.754	< .001	.355
T51-T55	-3.376	.002	.218
T35-T214	-5.473	< .001	.422
T214-T35	-3.721	.001	.252
T35-T51	-5.454	< .001	.420
T51-T35	-3.756	.001	.256
T214-T51	-4.091	< .001	.290
T51-T214	-3.46	.001	.226

Table 6.6 Results of Independent Samples T Test on the RT data from Beijing and Yantai older listeners in Experiment BJ.

Tone pair	*df*	*t*	*p*	η^2
T55-T35	38	3.229	.003	.215
T35-T55	35	2.16	.038	.118
T55-T214	32	2.702	.011	.186
T51-T55	38	2.578	.014	.149
T35-T214	38	2.358	.024	.128
T35-T51	32	1.934	.062	.105
T51-T214	31	1.976	.057	.111

Table 6.7 Results of Independent Samples T Test on the RT data from AE and Yantai older listeners in Experiment BJ.

Repeated measures analyses of variance were then performed on each of Rugao (older) and Yantai (older) listeners' data. The within-subject effect of "tone pair type" was found to be significant for both groups, sig.[$F(7.057, 155.244) = 22.048$, $p < .001$, partial $\eta^2 = .501$] for Rugao and sig.[$F(7.024, 133.463) = 14.156$, $p < .001$, partial $\eta^2 = .427$] for Yantai. Pairwise comparison of RT means showed that for Rugao (older) listeners, pairs T55/T51 and T35/T214 are the most confusable, T35-T55 falls in the middle, and the other pairs are less confusable. For the Yantai (older) listeners, T214-T51 stands out among all pairs and is the most distinctive. Pairs T35-T55, T35-T214, T51-T55 and T214-T35 are more confusable, and the rest fall in the middle. (See summary in Table 6.4.)

T35-T55 turned out to be the most confusable pair for the Yantai (older) listeners. This is different from all other Chinese groups. This seemingly surprising pattern can be explained by the diachronic process that merged a previous tone category – corresponding to T35 – with Y55. Perhaps their passive knowledge of the standard language through the mass media is not sufficient for them to form two tone categories of /35/ and /55/. In fact, I have noticed the mispronunciation of T35 morphemes as T55 or T51 (T51 being the Y55 counterpart) in some of my listeners' speech. Pairs T35/T214 are also among the most confusable. The cause might be the Yantai tone sandhi that changes Y214 to a surface [35]. This rising tone only exists in connected speech. It is also related to Y55 in the synchronic tonology, as both [35] and Y55 are sandhi forms of Y214, depending on contexts (see (4.2)).

6.1.3 Repeated measures ANOVA and planned comparisons: young vs. old listeners

No significant effect was found between the two age groups of Rugao listeners in the repeated measures ANOVA, [$F(1, 45) = .006$, $p = .941$, partial $\eta^2 < .001$]. or the T test. The main effect of tone pair was significant, [$F(7.105, 319.717) = 28.009$, $p < .001$, partial $\eta^2 = .384$]. So was the interaction of tone pair by age group, [$F(7.105, 319.717) = 2.215$, $p = .032$, partial $\eta^2 = .047$]. But planned comparisons of RT means between the two groups failed to yield significantly difference in RTs for any tone pair, although RTs for pair T51-T55 was more different (albeit still insignificant, $p = .192$) than the other pairs (Figure 6.3). A higher competence in Putonghua seems to have helped the young listeners to block the inter-dialectal R44-T51 correspondence interference to a certain extent.

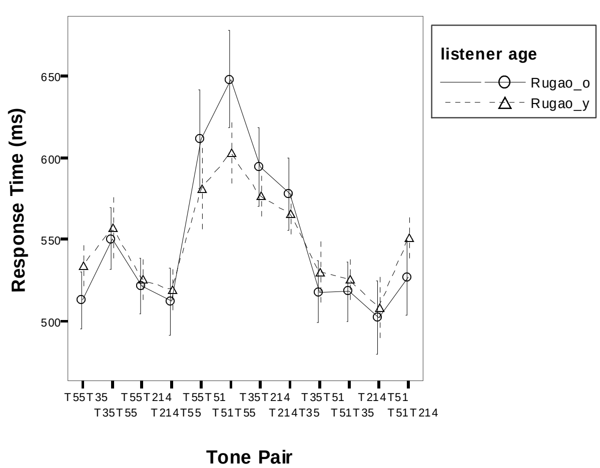


Figure 6.3 No significant effect was found with listener age for the Rugao listeners in Experiment BJ. Error bars show one standard error.

Since the Yantai young listeners have much shorter RTs than the older group, a repeated measures ANOVA found the between-subjects effect of listener age to be significant, sig.[$F(1, 44) = 9.11$, $p = .004$, partial $\eta^2 = .172$]. The main effect with tone pair was also significant, [$F(8.929, 392.878) = 24.656$, $p < .001$, partial $\eta^2 = .359$]. But the interaction of tone pair by age group was not significant, [$F(8.929, 392.878) = 1.629$, $p = .106$, partial $\eta^2 = .036$]. An independent samples T test shows that almost all tone pairs are significantly different for the two age groups but that the largest differences are in pairs T55/T35, T35-T51 and T55-T51 (Table 6.8). For the older listeners, with the historical merger of the Yantai counterpart of T35 with Y55 (i.e., the Yantai counterpart of T51), T35 is easily confusable with T51 and T55 (the latter of which being the phonetically most similar tone to Y55). On the other hand, learning Putonghua has apparently helped the Yantai young listeners to form a separate rising /35/ tone category. And because of their high competence in Putonghua, when the BJ stimulus tones were compared, relevant linguistic information might have been retrieved from their Putonghua lexicon, without much interference from their Yantai lexicon.

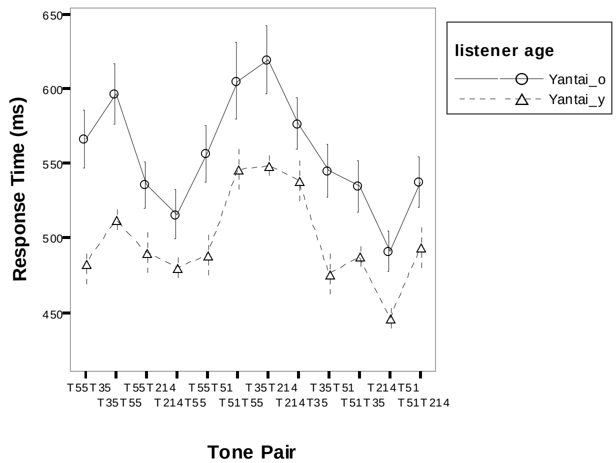


Figure 6.4 Experiment BJ Significant effects were found with listener age for the Yantai listeners in both the repeated measures ANOVA and T test. Error bars show one standard error.

Tone pair	*t(44)*	*p*	η^2
T55-T35	-3.61	.001	.229
T35-T55	-3.624	.001	.230
T55-T51	-2.833	.007	.154
T51-T55	-2.107	.041	.092
T35-T214	-2.908	.006	.161
T35-T51	-3.024	.004	.172
T51-T35	-2.258	.029	.104
T214-T51	-2.449	.018	.120
T55-T214	-2.003	.051	
T214-T55	-1.075	.085	
T214-T35	-1.624	.112	
T51-T214	-1.869	.068	

Table 6.8 Results of Independent Samples T Test on the RT data from Yantai young and old listeners in Experiment BJ.

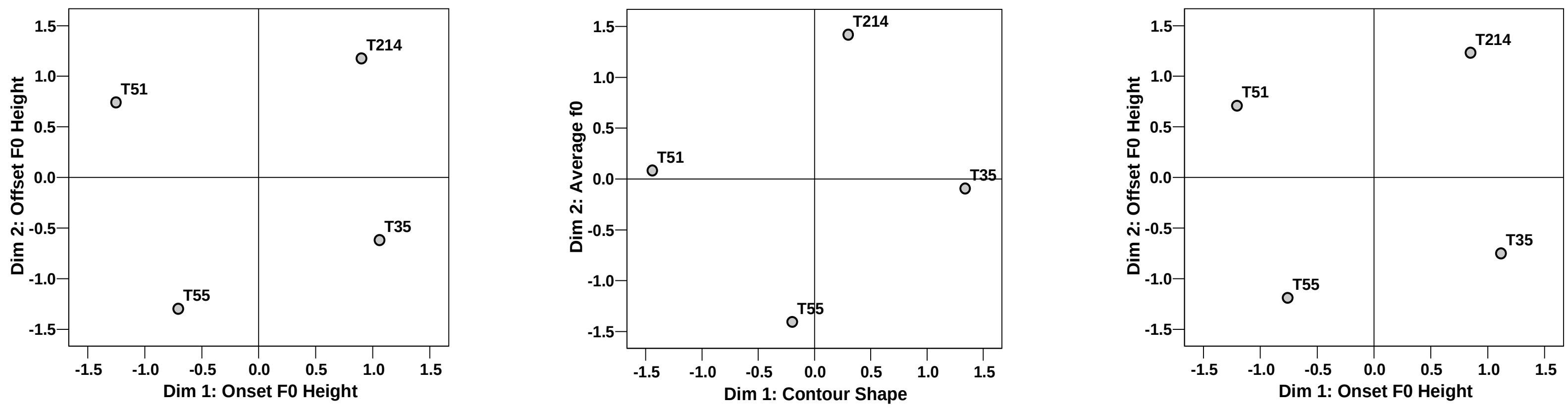
6.1.4 MDS: all listeners in Experiment BJ

As in the previous experiments reported in earlier chapters, we have seen that reaction time (RT) in this speeded AX discrimination task differs for different groups of listeners and that the salience of certain tonal distinctions depends on the listener' native language/dialect. In this section, I report the results of multidimensional scaling analyses of the tone spaces for each group of listeners. Treating RT as an indication of perceptual distance between tones (as in Chapter 2; see references cited therein), these analyses permit us to visualize the perceptual spaces of the Putonghua (Beijing) tones for each group of listeners.

The RT data were converted to perceptual distances by the reciprocal function and entered into six matrix files, with one file for each group of listeners and one matrix for each listener. These were analyzed using the INSCAL model (Carroll & Chang, 1970). Two listeners from Beijing and two from Yantai (older) were excluded, either because they had relatively high weirdness numbers in the inclusive analyses or because they guessed too much in the test and made too many mistakes. One Rugao (older) listener's data were lost. The MDS stimulus configurations are shown in Figure 6.5. The stress and RSQ values for these configurations are reported in Table 6.9.

Listener group	No. of Subj.	Stress	RSQ	Dim. weights	
				Dim 1	Dim 2
American	20	.164	.918	.5601	.3575
Beijing	22	.157	.918	.4744	.4432
Rugao (young)	24	.165	.910	.5606	.3497
Rugao (older)	23	.163	.913	.5015	.4118
Yantai (young)	26	.176	.895	.5671	.3279
Yantai (older)	19	.162	.913	.5289	.3840

Table 6.9 Stress and RSQ values for MDS group stimulus configurations in Experiment BJ.

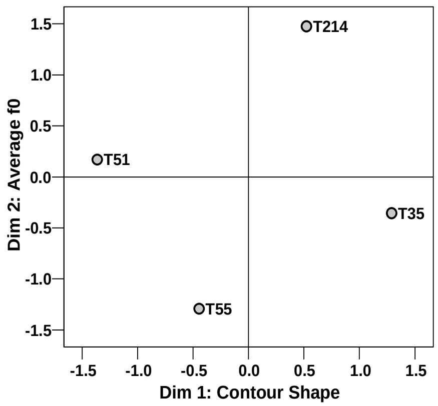


a. b. c.

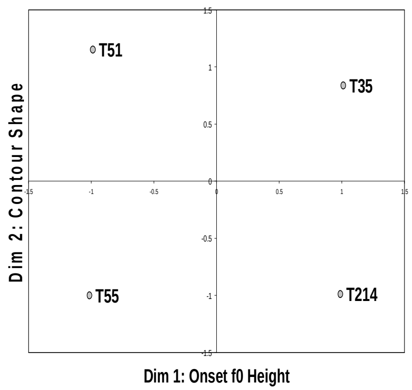
Figure 6.5 Experiment BJ group MDS stimulus spaces for American English (panel a), Beijing (panel b), Rugao young (panel c), Rugao old (panel d), Yantai young (panel e) and Yantai old listeners. Note that the positions of T35 and T214 are switched in the Yantai young listeners' space (panel e) as compared to the other five spaces.

(To be continued on next page)

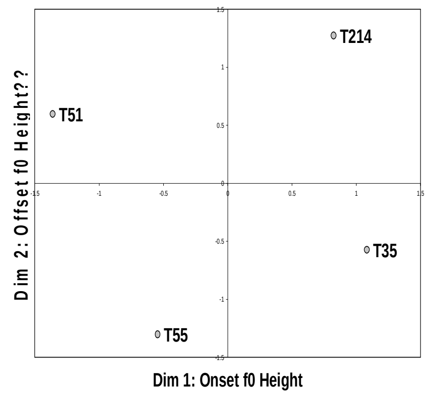
Figure 6.5: Continued.



d.



e.



f.

As shown in Figure 6.5, there are similarities and differences in the MDS spaces. The dimensions in the American English listeners' space correspond to onset and offset pitch heights rather nicely. The same labels can be applied to the Rugao (young) and Yantai (older) listeners' spaces, although the tones are a bit off along dimension 2 in the latter space. For the Beijing listeners, dimension 1 can be interpreted as contour shape: T51 has a falling contour, T55 a high level, T214 a low tone with a shallow dip, and T35 a rising contour. Dimension 2 in this space may be labeled average pitch height, with the low tone and the high tone placed at the two ends of the dimension. The same interpretations can be made of the Rugao (older) listeners' space. The Yantai (young) listeners' space is special in that T35 and T51 are placed very close to each other along dimension 2, instead of diagonally from each other as in all the other spaces. Thus, their contour shape dimension shows a dynamic vs. static contour contrast. Recall that this is also the pattern seen in the Chinese listeners' tone space in Experiment 1 (Figure 2.3a, Chapter 2). It is worth mentioning that such a placement of tones is not as unusual as it appears: some Beijing listeners have this pattern in their individual spaces. Except for the Yantai (young) listeners' tone space, the rising T35 and the falling T51 are placed diagonally, as are the high T55 and the low T214, in all other spaces.

The Euclidean distances in the spaces reflect the patterns shown in the RT plots (Figures 6.1 and 6.2). Distance between T35 and T214 is the shortest in the English listeners' space. This distance is also the shortest for the Beijing listeners, although T55 and T51 are also close to each other. The same can be said of the Yantai (young) listeners. For the Rugao listeners (young and older), T55 and T51 are the closest, but T35 and T214 are also very close. For the Yantai older listeners, T35 is close to both T55 and T214. In addition, T55 is also close to T51. Thus, again the MDS also reveal an age group difference for the Yantai and Rugao listeners.

6.1.5 Summary of discussion

As is obvious from the analyses on the RT data from Experiment BJ, no strong effect of the T214 rule showed up in the Beijing listeners' data. This could be due to the blocking effect of a speeded response (Fox, 1984). Nevertheless, there were still differences found between Beijing and AE listeners (e.g., the status of pairs T55/T51 along the confusion scale), and between Beijing and Yantai (young) listeners (e.g., a longer RT with T35-T214 in the Yantai listeners' data, suggesting a more conscious knowledge of the rule), although all three groups had mean RT under 500ms (Fox, 1984). Thus, it would be safe to conclude that procedures conducive to auditory listening (i.e., a short ISI of 100ms, an AX discrimination task, and a speeded response with a 500ms deadline) may weaken language-specific effects substantially but that it may not take away the effects completely.

Cross dialectal differences as well as age group differences were also observed in the RT data of the Chinese listeners. In general, perception of Putonghua (Beijing) tones by the older listeners, who have only passive knowledge of the standard language, was affected by their own native phonology as well as by the cross-dialectal tone category correspondences between their native dialect and the standard dialect. The younger listeners, especially the Yantai young listeners, who have higher competency in the standard language, perceived the tones more like the Beijing listeners, with their L2 Putonghua knowledge overriding the effects from their native phonology and cross-dialectal tone category correspondences. This is consistent with Elman, Diehl & Buchwald (1977), where different degrees of bilingualism were also reported to be reflected in the perception of a /ba-pa/ continuum by English-Spanish bilinguals.

6.2 Results and Analysis: Experiment RG

This section presents the results of Experiment RG – perception of Rugaohua tones by American English (AE), Beijing, Rugao (young and older), and Yantai (young and older) listeners. The outline and format of the results presentation is the same as for Experiment BJ in §6.1.

6.2.1 Repeated measures ANOVA and planned comparisons (T tests): young listeners

When Rugao (young) and Yantai (young) listeners' data were analyzed along with Beijing and American English listeners' RT data, the between-subjects effect from "listener language" (4 levels) was found to be significant, sig.[$F(3, 90) = 4.178$, $p = .008$, partial $\eta^2 = .122$]. The within-subject effect of tone pair was also significant, sig.[$F(9.947, 895.192) = 23.08$, $p < .001$, partial $\eta^2 = .204$]. So was the interaction of tone pair by listener language, sig.[$F(29.84, 895.192) = 2.098$, $p = .001$, partial $\eta^2 = .065$]. Post-hoc test further reveals that the between-subject effect came from the difference between Beijing (fastest, Figure 6.6, solid line, grand mean RT = 571ms) and Rugao young (slowest, Figure 6.6, dashed line, grand mean RT = 649ms) listeners ($p = .004$). No overall RT difference between any other pairs of listener groups was found. Relatively Low p values were found with Beijing and Yantai (.36), Rugao and Yantai (.574), as well as Rugao and English (.328). We should expect significant differences with some tone pairs between these listener groups showing up in planned comparisons of means.

Planned comparisons using Independent samples T tests found significant effects for all tone pairs except R45-R44 between Beijing and the Rugao young listeners. There were also marginal effects for certain pairs between Beijing and the Yantai young listeners, between Rugao and English listeners, as well as between Rugao young and Yantai young listeners. These results are reported in Tables 6.10, 6.11, 6.12 and 6.13. There was only a borderline effect for pair R45-R41 between Beijing and American English listeners (p = .06), where similarity between the offset f0 of R45 and the onset of R41 seems to have made the tones more confusable for the English listeners.

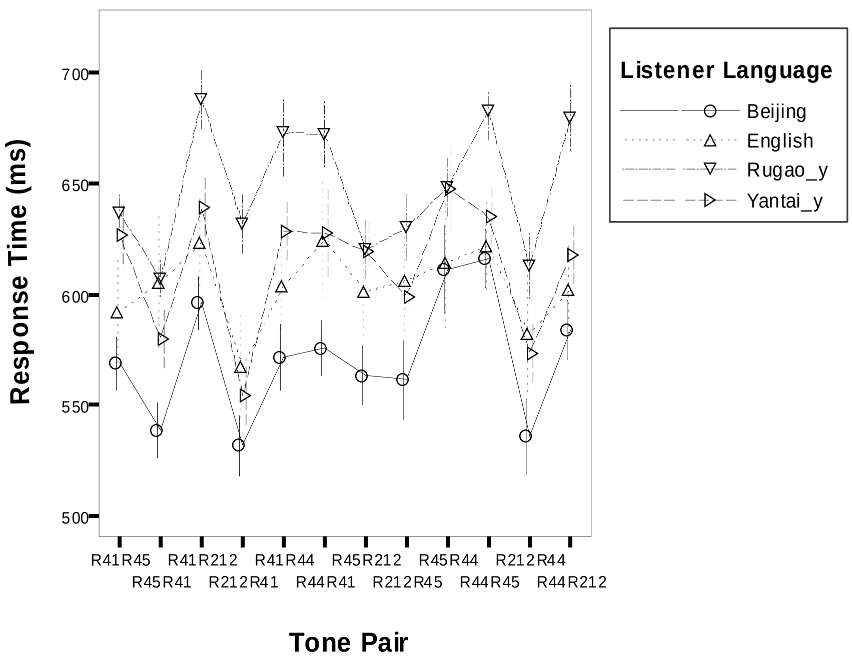


Figure 6.6 Experiment RG: Significant effects were found with listener language between the Beijing (solid line) and the Rugao young (dashed line) listeners in both the repeated measures ANOVA and the T test. No significant effects otherwise. Error bars show one standard error.

Tone pair	*t(45)*	*p*	η^2
R41-R45	-4.128	< .001	.275
R45-R41	-3.965	< .001	.259
R41-R212	-5.033	< .001	.360
R212-R41	-4.895	< .001	.347
R41-R44	-4.098	< .001	.272
R44-R41	-4.413	< .001	.302
R45-R212	-2.910	.006	.158
R212-R45	-2.931	.005	.160
R44-R45	-3.710	.001	.234
R212-R44	-3.342	.002	.199
R44-R212	-4.691	< .001	.328
R45-R44	-1.741	.089	

Table 6.10 Results of Independent Samples T Test comparing the RT data from Rugao (young) and Beijing listeners in Experiment RG.

Tone pair	*t(47)*	*p*	η^2
R41-R45	-2.576	.013	.124
R41-R212	-2.067	.044	.088
R41-R44	-2.230	.031	.096
R45-R212	-2.953	.005	.156
R45-R41	-1.87	.068	
R44-R41	-1.954	.057	
R212-R45	-1.702	.095	
R44-R212	-1.807	.077	

Table 6.11 Results of Independent Samples T Test comparing the RT data from Beijing and Yantai (young) listeners in Experiment RG.

Tone pair	*t(48)*	*p*	η^2
R41-R212	2.203	.032	.092
R212-R41	3.545	.001	.207
R44-R45	2.314	.025	.100
R44-R212	2.69	.01	.131
R44-R41	1.682	.099	
R212-R44	1.716	.093	

Table 6.12 Results of Independent Samples T Test comapring the RT data from Rugao (young) and Yantai (young) listeners in Experiment RG.

Tone pair	*t(42)*	*p*	η^2
R41-R212	-2.639	.012	.142
*R212-R41	-2.18	.037	.105
R41-R44	-2.297	.027	.112
R44-R45	-2.526	.015	.132
R44-R212	-2.791	.008	.156
R41-R45	-1.663	.107	

Table 6.13 Results of Independent Samples T Test comparing the RT data from Rugao and English listeners in Experiment RG. Equal variance cannot be assumed for pair R212-R41 and the degree of freedom value was 31.

It can be noticed in Figure 6.6 that the three groups of Chinese listeners have rather similar maxima and minima along the RT curves. Therefore, if there is a significant effect between two groups, we should expect it to be reflected in all tone pairs.

Such is the case between Beijing and Rugao (young) listeners. The fact that there is only a borderline effect with R45-R44 shows that this pair is more distinctive for the Rugao (young) listeners. On the other hand, since the RT differences between Beijing and Yantai (young) listeners are smaller, marginal effects are expected across the board. However, for pair R45-R212, there is a significant difference ($p = .005$), suggesting that this pair is more confusable for the Yantai (young) listeners. For the Beijing listeners, the acoustic difference between R45 and T35, as well as that between R212 and T214, may be large enough to override the T214 sandhi effect. Or maybe the difference is simply too large to evoke the sandhi effect at all. Note also that Yantai listeners were faster in discriminating R212-R41 than the Rugao (young) listeners ($p = .001$). This is somewhat surprising, given the fact that /Y31/ and /Y214/ both surface as [35] before /Y31/ in Yantai. Maybe the larger pitch range of the Rugao speaker made the R41 stimuli less similar to Y31: after all, their falling contour cover about 20 more Hz than the Y31 stimuli. Imaginably, despite of the neutralization rule, the falling versus low (dipping) contrast may be also more robust in the three-tone Yantai system than in the four-tone Rugao system, which explains at least in part why the Yantai (young) listeners were much faster than the Rugao (young) listeners in discriminating this pair of tones.

Repeated measures ANOVAs were also performed on each listener group's RT data to further investigate the effect of the within-subject factor "tone pair type" (see summary in Table 6.14 below). The English listeners had a rather flat RT curve. The only outstanding and most dissimilar pair R212-R41 was only marginally different from three other pairs R41-R212 ($p = .042$), R212-R45 ($p = .018$) and R45-R44 ($p = .029$). Pairs R41-R212 and R45-R44 have a rather small pitch transition from the first stimulus tone to the second. Although the f0 difference between the offset of R212 and the onset of R45 is quite large, the rising contour in R45 seemed to have been interpreted by the English listeners as being composed of a low start and a high end. On the other hand, the f0 difference between the offset of R212 and the onset of R41 seemed to be large enough for these non-tone language speaking listeners to fairly easily discriminate pair R212-R41. Perhaps the falling contour was also parsed into two pitch targets, a high start and a low end in this case. As we shall see later in this chapter, these interesting patterns also showed up in the MDS analysis.

The most confusable pair for the Beijing listeners is R44-R45. As noted in Chapter 4, acoustically the f0 onsets are very similar for these tones. As a result, the rising contour in R45 may not be as prominent as in T35 of Putonghua. Pairs R45-R44 and R41-R212 are also very confusable for this group of listeners. The most dissimilar pairs for the Beijing listeners are R45-R41, R212-R41 and R212-R44. The first of the last three pairs involves a pitch range difference and contour shape difference. The perception of the other two pairs probably involved "undoing" the downstepping effect of the low R212, making the f0 onsets in R41 and R44 more different from the f0 onset of R212.

For the Rugao (young) listeners, pairs R41-R212, R44-R45, and R44-R212 are the most confusable, while pairs R45/R212, R212-R44 and R45-R41 are the least confusable. As mentioned earlier, despite the relatively high f0 onset found in instrumental studies (Huang 2002 and the present study), R41 has been described as /21/ in previous impressionistic studies by native linguists, which probably reflect the native perception of it as a fairly low overall pitch. It is thus no wonder that R41-R212 is the most confusable pair. Since R45 and R212 have less obvious pitch movements compared with T35 and T214 in Putonghua and since R44 has a relatively low pitch level in comparison to T55, it is not surprising that R44-R45 and R44-212 also have short perceptual distances.

The Yantai (young) listeners responded as if pairs R212-R41, R212-R44 and R45-R41 were the most dissimilar. Recall that these are also the most dissimilar pairs for the Beijing listeners. All other pairs have about the same confusability for the Yantai (young) listeners, with no significant difference between any two pairs.

Listener group	Confusable	Dissimilar
AmEng	R212-R45, R45-R44 R41-R212	R212-R41
Beijing	R44-R45	R45-R41, R212-R41, R212-R44
Rugao (young)	R41-R212, R44-R45, R44-R212	R45/R212, R212-R44, R45-R41
Rugao (old)	R41-R212, R44-R45, R44-R212	R212-R44, R45-R41, R45-R44
Yantai (young)	-----	R212-R41, R212-R44, R45-R41
Yantai (old)	R41-R212, R44/R45, R44-R212	R45-R41, R212-R44, R212-R41, R212-R45

Table 6.14 Experiment RG: Confusability rankings of Rugao tones.

6.2.2 Repeated measures ANOVA and planned comparisons: old listeners

Repeated measures analysis on Rugao (older), Yantai (older), Beijing and American English listeners' RT data in Experiment RG revealed significant language effect, sig.[$F(3, 86) = 6.977$, $p < .001$, partial $\eta^2 = .196$]. Post-hoc test and pairwise comparisons showed that the significant effect came from three sources: Beijing was significantly different from Rugao ($p = .018$) and from Yantai ($p < .001$), and AE was marginally different from Yantai ($p = .035$).

The within-subjects factor of "tone pair" also had a significant effect, sig.[$F(9.903, 851.619) = 25.666$, $p < .001$, partial $\eta^2 = .23$]. The interaction of "listener language*tone pair" was significant as well, sig.[$F(29.708, 851.619) = 2.696$, $p < .001$, partial $\eta^2 = .086$].

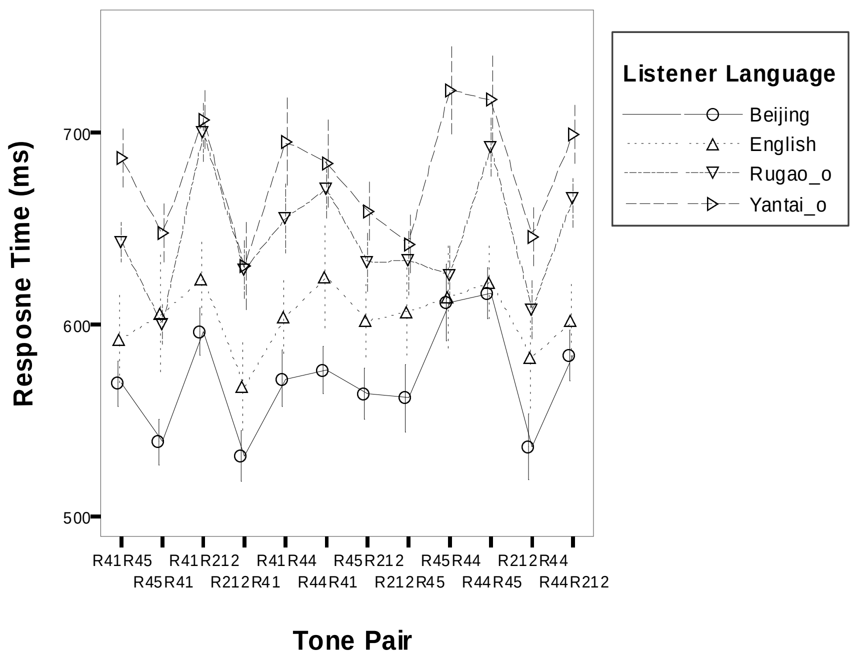


Figure 6.7 Significant effects were found with listener language (i) between Beijing and Rugao (old) listeners ($p = .018$), (ii) between Beijing and Yantai (old) listeners ($p < .001$), and (iii) between English and Yantai (old) listeners ($p = .035$) in the repeated measures ANOVAs. Error bars show one standard error.

The independent samples T tests further revealed the sources of the significant differences found in the repeated measures analysis of variance. Beijing and Rugao (older) listeners had significantly different RTs for all pairs except for R45-R44, where Rugao listeners showed better than average discrimination. Beijing and Yantai (older) listeners' RTs differed significantly for all tone pairs. The English and and Yantai (older) listeners were at least marginally different for most of the tone pairs. These results are reported in Tables 6.15, 6.16 and 6.17.

In addition, the Rugao (older) listeners differed from the English listeners in several pairs, especially pairs R41-R212 and R44-R45, where the AE listeners seem to have found larger onset pitch differences. On the other hand, we note two pairs, i.e. R45-R44 and R45-R41, where the Rugao (older) and the English listeners had very similar RTs. In other words, these pairs were more confusable for the English listeners but relatively distinctive for the Rugao (older) listeners. As in the Putonghua pairs found more confusable by the English listeners, these pairs have small f0 transitions from the offset of the first stimulus tone to the onset of the second stimulus tone, which did not affect the perception by the Rugao (older) listeners, who were probably also able to use contour information to a certain extent in their discrimination. There was also one pair, namely R45-R44, significantly different for the Rugao (older) and the Yantai (older) listeners. Obviously, with the rising tone only as a surface tone in Yantai and the similarity in contour shapes between R45 and R44, the Yantai (older) listeners found the distinction between these tones hard to detect.

Tone pair	*t(45)*	*p*	η^2
R41-R45	4.131	< .001	.275
R45-R41	3.101	.003	.176
R41-R212	4.858	< .001	.344
R212-R41	4.351	< .001	.296
R41-R44	3.415	.001	.206
R44-R41	4.239	< .001	.285
R45-R212	3.110	.003	.177
R212-R45	2.913	.006	.159
R44-R45	3.817	< .001	.245
R212-R44	2.99	.005	.166
R44-R212	4.182	< .001	.280
R45-R44	.759	.452	

Table 6.15 Results of Independent Samples T Test comparing the RT data from Rugao (older) and Beijing listeners in Experiment RG.

The results of T test comparing the RT data from Rugao (older) and Beijing listeners reported in Table 6.15 are comparable to Table 6.10 (T test results from comparisons of Rugao (young) and Beijing listeners' RT data) in every aspect, suggesting that the two age groups of Rugao listeners reacted to the Rugaohua stimuli in very similar ways.

Tone pair	*t(43)*	*p*	η^2
R41-R45	-5.196	< .001	.386
R45-R41	-4.238	< .001	.295
R41-R212	-4.869	< .001	.355
R212-R41	-3.651	.001	.237
R41-R44	-3.864	< .001	.258
*R44-R41	-3.491	.002	.290
R45-R212	-3.988	< .001	.270
R212-R45	-3.139	.003	.186
R45-R44	-3.704	.001	.242
R44-R45	-3.635	.001	.235
R212-R44	-3.897	< .001	.261
R44-R212	-4.807	< .001	.350

Table 6.16 Results of Independent Samples T Test comparing the RT data from Beijing and Yantai (old) listeners in Experiment RG. (* Equal variance was not assumed for pair R44-R41 and degree of freedom was 30.)

Tone pair	*t(40)*	*p*	η^2
R41-R45	-3.021	.004	.186
R41-R212	-2.875	.006	.171
R41-R44	-2.437	.019	.129
R45-R44	-2.823	.007	.166
R44-R45	-2.793	.008	.163
R44-R212	-3.091	.004	.192

Table 6.17 Results of Independent Samples T Test comparing the RT data from American English and Yantai (old) listeners in Experiment RG.

Separate repeated measures ANOVAs were also done on the RT data from the Rugao (older) and Yantai (older) to further test the within-subject effect of "tone pair type". For the Rugao (older) listeners, pairs R41-R212, R44-R45 and R44-R212 are the most confusable, while pairs R212-R44, R45-R41 and R45-R44 are the least confusable. This is almost the exact pattern found in the Rugao (young) listeners' data. Yantai (older) listeners showed a pattern similar to the Rugao listeners. They also reacted as if R41-R212, R44/R45 and R44-R212 were the most confusable, while R45-R41, R212-R44, R212-R41 and R212-R45 were the most dissimilar (see summary in Table 6.14).

6.2.3 Repeated measures ANOVA and planned comparisons: Rugao young vs. old listeners

Repeated measures ANOVA did not find a significant main effect between the two age groups of Rugao listeners (Figure 6.8), [$F(1, 46) = .017$, $p = .897$, partial $\eta^2 < .001$]. The within-subject effect of tone pair was significant, [$F(9.192, 422.834) = 24$, $p < .001$, partial $\eta^2 = .343$]. The interaction of tone pair by age group was not significant, [$F(9.192, 422.834) = .91$, $p = .518$, partial $\eta^2 = .019$].

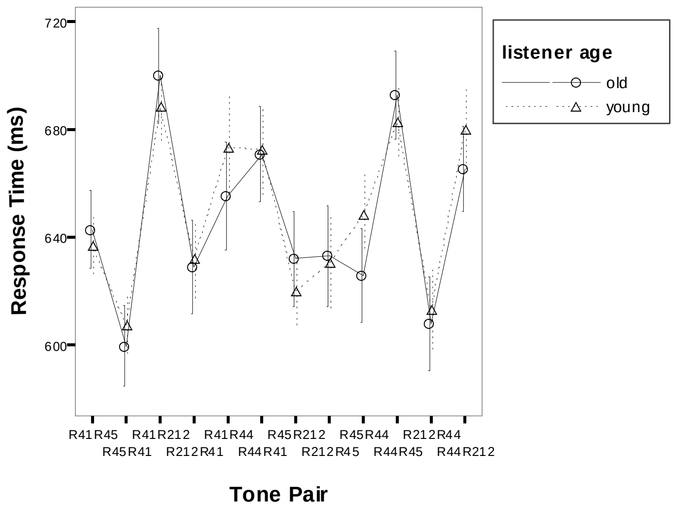


Figure 6.8 No significant effects were found with listener age for the Rugao listeners in Experiment RG. Error bars show one standard error.

6.2.4 Repeated measures ANOVA and planned comparisons: Yantai young vs. old listeners

Repeated measures ANOVA found a marginally significant main effect between the two age groups of Yantai listeners in Experiment RG, sig.[$F(1, 46) = 6.731$, $p = .013$, partial $\eta^2 = .128$). The within-subject effect of tone pair was significant, sig.[$F(10.287, 473.199) = 15.907$, $p < .001$, partial $\eta^2 = .257$). Since the two groups have very similar RT curves, the tone pair by order interaction was not significant, [$F(10.287, 473.199) = .833$, $p = .6$, partial $\eta^2 = .018$]. Independent samples T test (Table 6.18) also showed that eight (8) of the 12 pairs were at least marginally different and that the confidence levels for claiming no difference with the other four pairs were quite low (highest $p = .103$). This indicates that the pattern of tone discrimination was the same for the young and older Yantai listeners, despite an overall RT difference.

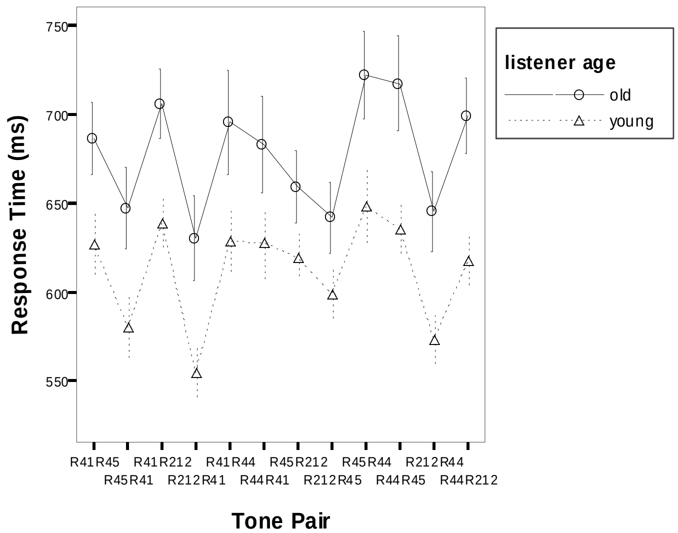


Figure 6.9 Significant effects were found with listener age for the Yantai listeners in both the repeated measures ANOVA and the T test (Table 6.18). But the patterns of tone discrimination were remarkably similar, as can be seen from the resemblance of the RT curves here. Error bars show one standard error.

Tone pair	*t(46)*	*p*	η^2
R41-R45	-2.087	.042	.086
R45-R41	-2.376	.022	.109
R41-R212	-2.586	.013	.127
R212-R41	-2.747	.009	.141
R45-R44	-2.293	.026	.103
R44-R45	-2.768	.008	.143
R212-R44	-2.604	.012	.128
R44-R212	-3.102	.003	.209

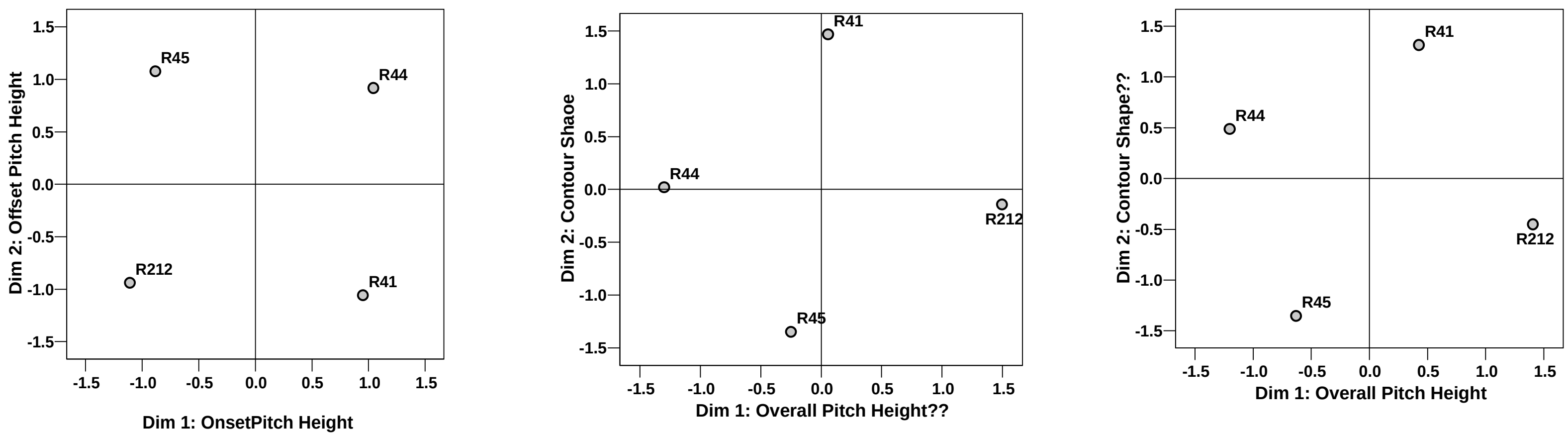
Table 6.18 Results of Independent Samples T Test on the RT data from Yantai young and old listeners in Experiment RG.

6.2.5 MDS: all listener groups in Experiment RG

As in the previous experiments, reciprocals of RTs for the test pairs were entered as perceptual distances into six (6) separate matrices for INSCAL analyses, which yielded the group tone spaces in Figure 6.10. The stress and RSQ values for these configurations are reported in Table 6.19. The two dimensions in all spaces have about equal weights. All listener groups reached their respective MDS solution rather unanimously, without very high weirdness numbers.

Listener group	No. of matrices	Stress	RSQ	Dim. weights	
				Dim 1	Dim 2
American	20	.177	.898	.4715	.4268
Beijing	24	.175	.896	.4573	.4390
Rugao (young)	24	.175	.895	.4655	.4298
Rugao (older)	24	.169	.904	.4731	.4307
Yantai (young)	26	.179	.891	.4535	.4378
Yantai (older)	20	.182	.888	.4725	.4158

Table 6.19 Stress and RSQ values for MDS group stimulus configurations in Experiment RG.

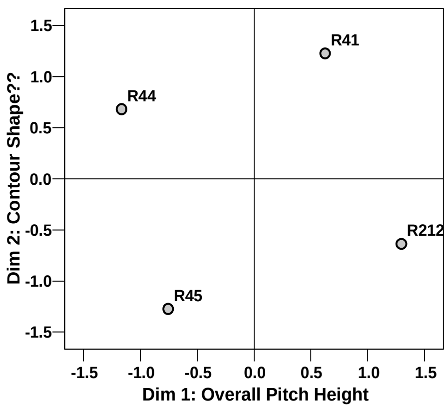


a. b. c.

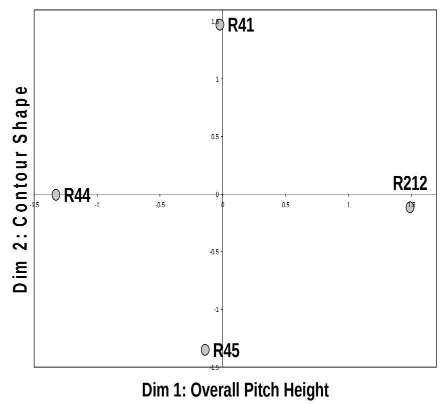
Figure 6.10 The perceptual spaces of Rugao tones as revealed in the American English (panel a), Beijing (panel b), Rugao young (panel c), Rugao older (panel d), and the Yantai young (panel e) and older (panel f) listeners' RT data.

(To be continued on the next page.)

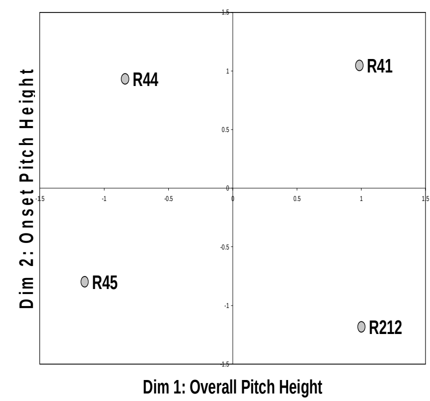
Figure 6.10: Continued.



d.



e.



f.

As can be seen in Figure 6.10, all groups placed R44 diagonally from R212 and R45 diagonally from R41, giving the basic configuration of the tones. R44 and R212 may represent a high versus low register contrast, while R45 and R41 a rising versus falling contrast. Everyone agrees that R45 and R44 are among the most confusable, although the Rugao (especially the older) listeners discriminated these tones better than others, making fewer mistakes (see §5.5, Chapter 5).

As before, the American English listeners placed the tones in extremely neat alignments along both dimensions, which can be interpreted as "Onset Pitch Height" and "Offset Pitch Height". As mentioned above, the rising contour of R45 seems to have been analyzed into a relatively low f0 onset and a higher f0 offset, while the falling contour of R41 into a relatively high onset and a lower offset. As a result, R45 is placed closer to R212 and R41 closer to R44 on the "onset" dimension. Dimension 2 may also be called "Overall Pitch Height", but such a label would miss the fact that pitch offsets were important for these listeners (see § 6.2.1).

In the Beijing listeners' tone space, dimension 2 is clearly a "contour shape" dimension. Along dimension 1, the low tone R212 is placed farther away from the other three tones, suggesting that this is an "overall pitch height" dimension. We might expect R44 and R45 to be even closer together along this dimension, but then the "contour shape" dimension would have to be distorted. Since the two dimensions have about equal weights in these spaces (Table 6.19), dimension 1 was not strong enough to alter the arrangements of the tones along dimension 2. The same can be said of Yantai young listeners' tone space, which has the same configuration as Beijing listeners'.

The two age groups of Rugao listeners have very similar tone spaces. (In fact, a MDS combining the data from both listener groups yielded a similar space, too.) While dimension 1 in their spaces can be labeled "overall pitch height", dimension 2 is a bit hard to interpret. Maybe it is similar to dimension 2 in the Beijing listeners' space and can be called a "contour shape" dimension. It is just distorted because the ratio of dimensional weights is bit larger than that in the Beijing listeners' space. It does not make much sense to call it an "onset pitch" dimension, because even though R45 might be analyzed as having a phonological low onset, it certainly would not be perceived psychoacoustically lower than the onset of R212.

The Yantai (older) listeners' space is different from all other groups of Chinese listeners. It is somewhat similar to the English listeners in that both spaces are relatively small and that the tones occupy the four corners of the spaces. The dimensions can be interpreted as "overall pitch" and "onset pitch", respectively.

6.2.6 Summary of discussion

The AE listeners had the same psychoacoustic listening for the Rugao tones as in the previously reported experiments using Putonghua/Beijing stimuli in our study. The Chinese listeners, with "onset pitch", "contour shape" and "overall pitch" showing up as dimensions in the MDS analyses, seemed again to have ignored offset pitch information to a certain extent. The two age groups of Rugao listeners showed the same pattern in their perception of RG tones. They had some native advantage in distinguishing R45 and R44, which were more confusable for the non-Rugao Chinese listeners.

6.3 Results and Analysis: Experiment YT

Although some between listener order effects were found with the YT data (as reported in §5.6, Chapter 5), the overall patterns in the three Latin Square orders are rather similar for all listeners groups. As a result, the YT data will also be pooled together in the analyses below as was done in the analyses for Experiments BJ and RG.

6.3.1 Repeated measures ANOVA and planned comparisons (T tests): young listeners

Repeated measures ANOVA found a significant between-subject listener language effect, sig.[$F(3, 89) = 6.301$, $p = .001$, partial $\eta^2 =.175$]. Post-hoc tests showed that the source of this effect was the large RT differences between the Beijing and the Rugao (young) listeners. The within-subject factor "tone pair type" also had a significant effect, sig.[$F(4.974, 442.653) = 14.082$, $p < .001$, partial $\eta^2 =.137$]. But no significant effect was found with the interaction of "listener language" and "tone pair type". This suggests that (within the variability in these data) the RT curves can basically be considered to be parallel (Figure 6.11). This was also confirmed by planned comparisons using the Independent samples T test, which found no significant RT difference with any tone pair between the Yantai (young) listeners and the American English listeners, between the Rugao (young) listeners and the American English listeners, or between the Beijing listeners and the American English listeners.

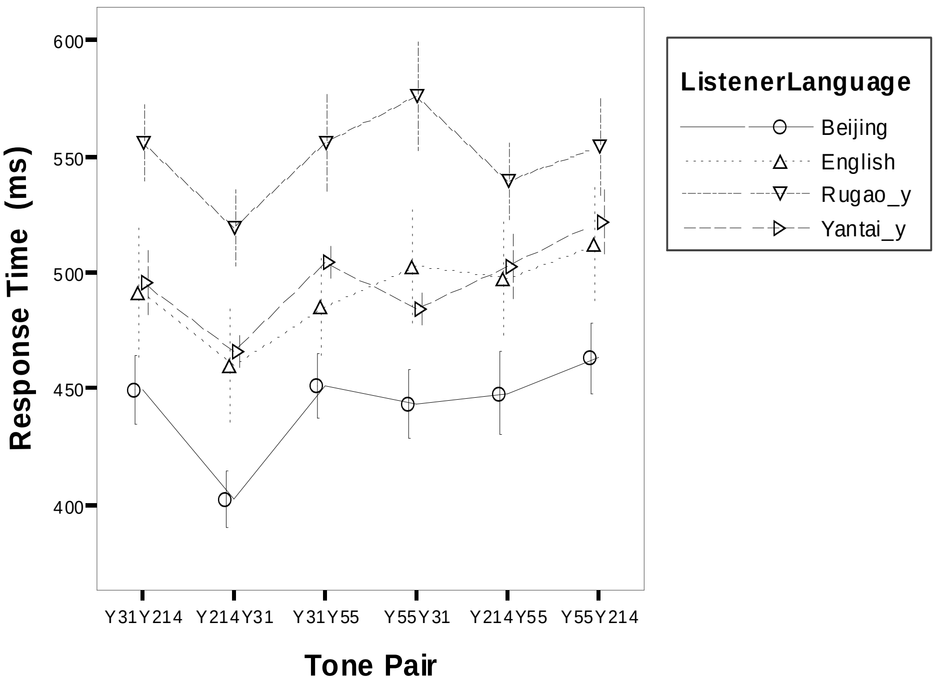


Figure 6.11 Experiment YT: Repeated measures ANOVA only found significant language effect between Beijing and Rugao (young) listeners' RT data. Error bars show one standard error.

The Yantai (young) listeners differed significantly from the Rugao listeners in just one pair, Y55-Y31, sig.[t (47)= 3.123, p = .004, η^2 = .227]. There were marginal differences with two other pairs. We should also note that Y31-Y55 was not significantly different for the two listener groups (p = .058), i.e., the Yantai listeners were relatively slow in discriminating this pair than Y55-31. This directional difference may be linked to the tonal realization rules in Yantai. A surface [Y31-Y55] may come from one of two sources: /Y31.Y55/ or /Y55.Y55/. Thus, this Y55 rule has the same effect as the T214 rule in Putonghua; that is, before Y55, the contrast between Y31 and Y55 is neutralized. Interestingly, such an effect was not found with Y55-Y31, although a surface [Y55-Y31] may also be derived from two underlying sequences: /Y31.Y31/ (when the first Y31 syllable has a sonorant onset) and /Y55-Y31/. As noted earlier, this realization rule for /Y31.Y31/ is a bit unusual in that it only applies to certain Y31 morphemes with sonorant onsets (which were not included in the YT stimuli). It is rather likely that Yantai young listeners may not have acquired this odd rule that only applies to a small subset of Y31 morphemes. It would not be a complete surprise if they simply replaced this rule with the more prevalent rule of /Y31.Y31/ → [Y35.Y31]. If this were the case, [Y55-Y31] would only have one corresponding underlying sequence, /Y55-Y31/. That is, there would be no neutralization rules involved with this pair of tones in this environment. Another possibility is that these so-called underlying /Y31.Y31/ sequences were simply acquired as /Y55.Y31/ words, instead of two separate morphemes and a derivation rule. This is likely the case because historically these morphemes should have had tone contours similar to the tone category that later merged with Y55. More research into the phonology of Yantai is needed to verify this speculation. The true Yantai counterpart of the Putonghua T214 sandhi is the Y214 sandhi, /Y214.Y214/ → [Y55.Y214]. At first sight, this rule should translate to Putonghua /T214.T214/ → [T51.T214], because Yantai tone category Y55 corresponds to T51. However, Y55 is a merged category that includes historical T51 cognates as well as T35 cognates. If, instead of saying the historical Yantai counterpart of T35 merged into the Yantai counterpart of T51, we theorize the merger the other way around, then the Yantai rule should translate into Putonghua as the T214 sandhi (see also Jansche 1999, ms.). As a result, we should expect the Yantai listeners to have longer reaction time for pairs Y214/Y55. This seems true with the data, as these are the only two pairs where the Yantai young listeners' RTs were most similar to those of the Rugao listeners ($p >= .158$).

Significant RT differences were also found with two pairs (Y214-Y31 and Y31-Y55) between the Yantai (young) and the Beijing listeners (Table 6.20). There were also marginal differences with the other pairs.

Although the Beijing listeners differed significantly from the Rugao (young) listeners in all pairs (Table 6.21), it is rather unremarkable: it is just an overall slow versus fast RTs difference.

Tone pair	*t(48)*	*p*	η^2
Y31-Y214	-2.106	.04	.085
Y214-Y31	-3.848	< .001	.236
Y31-Y55	-2.754	.008	.136
Y55-Y31	-2.049	.046	.084
Y214-Y55	-2.221	.031	.093
Y55-Y214	-2.613	.012	.125

Table 6.20 Results of Independent Samples T test on Yantai (young) and Beijing listeners' RT data in Experiment YT.

Tone pair	*df*	*t*	*p*	η^2
Y31-Y214	45	4.536	< .001	.314
Y214-Y31	36	4.945	< .001	.405
Y31-Y55	37	3.922	< .001	.293
Y55-Y31	35	4.419	< .001	.356
Y214-Y55	45	3.453	.001	.209
Y55-Y214	38	3.218	.003	.215

Table 6.21 Results of Independent Samples T test on Rugao (young) and Beijing listeners' RT data in Experiment YT.

6.3.2 Repeated measures ANOVA and planned comparisons (T tests): older listeners

There was a significant between-subjects effect from listener language, sig.[F(3, 86) = 8.754, p < .001, partial η^2 = .234]. Post-hoc test revealed that the effect came from the differences between Beijing and Yantai (older) listeners (p < .001) and between Beijing and Rugao (older) listeners (p < .001). The effect was marginal between the American English and the Yantai listeners (p = .083). The within-subject factor of tone pair had a significant effect, [F(4.756, 409.042) = 19.155, p < .001, partial η^2 = .182]. The interaction of tone pair and listener language was also significant, [F(14.269, 409.042) = 1.922, p = .017, partial η^2 = .065].

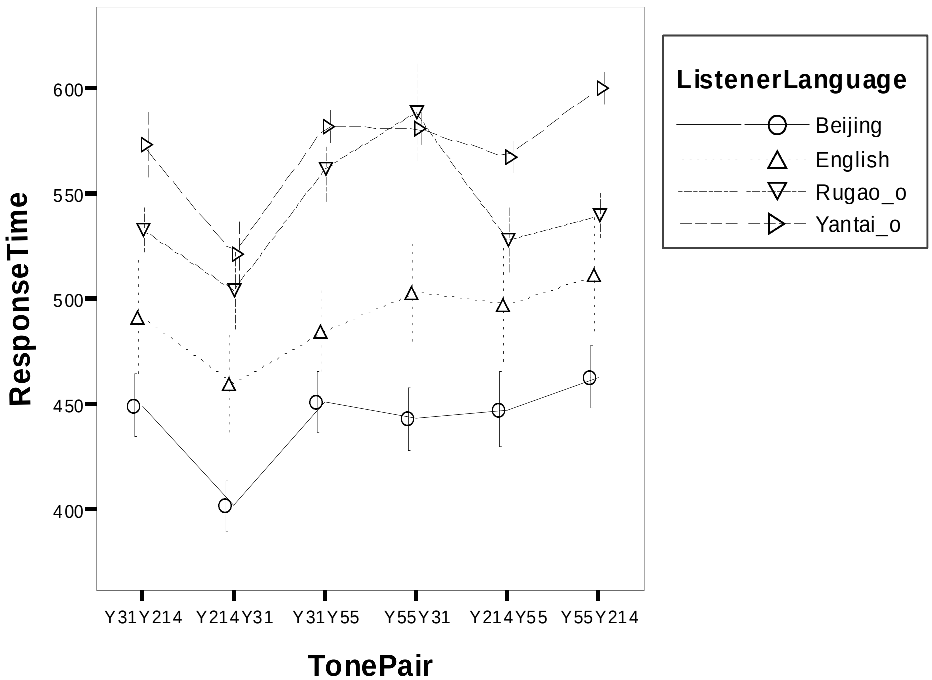


Figure 6.12 Experiment YT: Repeated measures ANOVA found a significant language effect (between Beijing and Rugao (older) listeners' RT data). Error bars show one standard error.

An independent samples T test comparing Yantai (older) and English RT data found at least marginal effects for almost all tone pairs (Table 6.22). In addition, the American English listeners' RT data had larger variances, with standard deviation and standard error values being about twice as large as those in the Yantai (older) listeners' data in most cases. As a result, equal variance was assumed for only one pair, namely

Y214-Y31, which showed only a marginal effect between the two groups ($p = .061$). Comparison between Yantai and Beijing uncovered significant RT difference in each pair ($p < .001$) (Table 6.23).

Significant RT difference was only found for pair Y55-Y214 between Yantai and Rugao listeners; pairs Y214-Y55 and Y31-Y214 had marginal effects ($p = .091$ and $p = .076$, respectively). As in the Yantai young listeners' data, Y55 and Y214, the two tones involved in the Yantai equivalent of the T214 sandhi, were found to be the most confusable by the Yantai (older) listeners.

In addition, the Rugao (older) listeners' data were significantly different from the American English listeners' in pairs Y31-Y55 [$t(42) = -2.872$, $p = .006$, $\eta^2 = .164$] and Y55-Y31 [$t(42) = -2.270$, $p = .028$, $\eta^2 = .109$]. Significant RT differences were found across the board between the Rugao and Beijing listeners' data (Table 6.24).

Tone pair	*df*	*t*	*p*	η^2
Y31-Y214	30	2.46	.020	.168
Y214-Y31	39	1.926	.061	.087
Y31-Y55	31	3.786	.001	.314
Y55-Y31	30	2.551	.016	.180
Y214-Y55	29	2.234	.033	.149
Y55-Y214	30	2.798	.009	.209

Table 6.22 Results of Independent Samples T test on Yantai (older) and American English listeners' RT data in Experiment YT.

Tone pair	*t(43)*	*p*	η^2
Y31-Y214	5.621	< .001	.424
Y214-Y31	5.786	< .001	.438
Y31-Y55	6.701	< .001	.511
Y55-Y31	6.522	< .001	.497
Y214-Y55	5.068	< .001	.374
Y55-Y214	6.319	< .001	.481

Table 6.23 Results of Independent Samples T test on Yantai (older) and Beijing listeners' RT data in Experiment YT.

Tone pair	*t(46)*	*p*	η^2
Y31-Y214	-3.961	< .001	.254
*Y214-Y31	-4.522	< .001	.345
Y31-Y55	-5.169	< .001	.367
Y55-Y31	-4.828	< .001	.336
Y214-Y55	-3.199	< .001	.182
Y55-Y214	-3.648	< .001	.224

Table 6.24 Results of Independent Samples T test on Rugao (older) and Beijing listeners' RT data in Experiment YT. Equal variance is not assumed for Y214-Y31 (degree of freedom = 39).

6.3.3 Repeated measures ANOVA and pairwise comparison: two age groups of Rugao

The between-subject effect of listener age was not significant for the Rugao listeners [$F(1, 45) = .091$, $p = .764$, partial $\eta^2 = .002$]. Tone pair had a significant effect, [F(4.457, 200.546) = 9.411, p < .001, partial $\eta^2 = .173$]. No significant effect was found with the tone pair by age group interaction, [F(4.457, 200.546) = .748, p = .574, partial $\eta^2 = .016$].

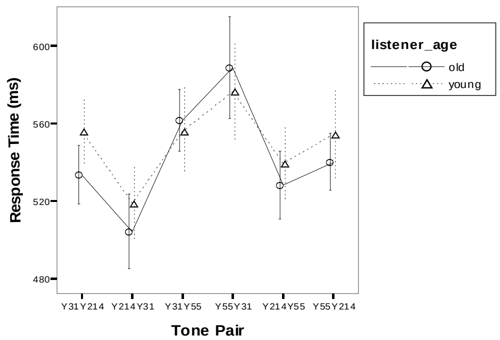


Figure 6.13 Experiment YT: No significant between-subjects effect was found between the two age groups of Rugao listeners. Error bars show one standard error.

6.3.4 Repeated measures ANOVA and pairwise comparison: two age groups of Yantai

The between-subjects factor of "listener age" was found to be significant for the two age groups of Yantai listeners in the repeated measures ANOVA, sig.[F(1, 46) = 8.412, p = .006, partial $\eta^2 = .155$]. The within-subjects factor "tone pair" also had a significant effect, sig.[F(4.638, 213.34) = 11.402, p< .001, partial $\eta^2 = .199$]. But the interaction of listener age by tone pair was not significant, [F(4.638, 213.34) = 1.033, p = .397, partial $\eta^2 = .022$]. So, the patterns may be similar for the Yantai young and older listeners, despite the overall RT difference (but see below).

An independent samples T test found significant difference for each tone pair between the two age groups (Table 6.25). The young listeners were much faster in general. The two RT curves are essentially parallel to each other, (as indicated by the non-significant tone pair by listener age interaction), except for pairs Y55-Y31 and Y31-Y55, where the differences in RTs are the greatest between the two groups. As mentioned earlier, the Yantai young listeners may only have one sandhi rule for the /Y31.Y31/ sequences, namely /31.31/ → [35.31], while the older listeners have an additional /31.31/ → [55.31], which neutralizes the contrast between Y31 and Y55 before Y31. As a result, tones Y31 and Y55 may be more contrastive for the younger listeners.

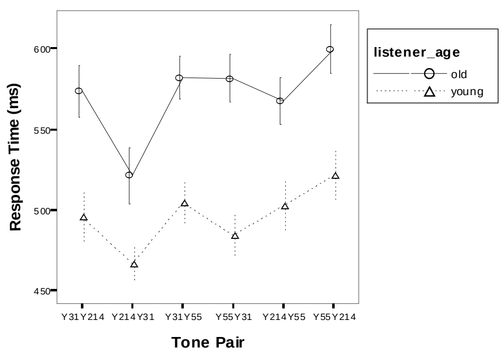


Figure 6.14 Experiment YT: significant between-subjects effect was found between the two age groups of Yantai listeners in both the repeated measures ANOVA and Independent Samples T test. Error bars show one standard error.

Tone pair	*t(46)*	*p*	η^2
Y31-Y214	-2.625	.012	.100
Y214-Y31	-2.06	.045	.084
Y31-Y55	-2.705	.01	.137
Y55-Y31	-3.387	.001	.200
Y214-Y55	-2.009	.05	.081
Y55-Y214	-2.539	.015	.123

Table 6.25 Results of Independent Samples T test on Yantai young and older listeners' RT data in Experiment YT.

6.3.5 Repeated measures ANOVA and pairwise comparison: each listener group

Separate repeated measures ANOVA on the AE listeners' RT data in Experiment YT revealed only marginal differences between the most dissimilar Y214-Y31 and the more similar Y55-Y31 ($p = .066$) and Y55-Y214 ($p = .051$). But in the Yantai (young) listeners' data, the most dissimilar pair Y214-Y31 was significantly different from Y31-Y55 and Y214/Y55 ($p < .05$), while in the Yantai (older) listeners' data it was significantly different from all other pairs except Y214-Y31.

The Beijing listeners also found Y214-Y31 to be the most dissimilar pair, which is significantly different from all other pairs except Y55-Y31. Pair Y214-Y31 was also the most dissimilar to the Rugao young listeners, although there were only marginal effects between this pair and Y31-Y214 ($p = .064$) and Y55-Y31 ($p = .067$). In the Rugao (older) listeners' data, Y55-Y31 was the most similar and was significantly different from pairs Y31/Y214 ($p < .05$).

Listener grp	most confusable	"middle"	most dissimilar
AmEng	Y55-Y214, Y55-Y31	Y31-Y55, Y31-Y55 Y214-Y55	Y214-Y31
Beijing	Y31-Y214, Y31-Y55, Y214-Y55, Y55-Y214	Y55-Y31	Y214-Y31
Rugao (y)	Y31-Y214, Y55-Y31	Y31-Y214, Y214-Y55, Y55-Y214	Y214-Y31
Rugao (o)	Y31-Y55, Y55-Y31	Y31-Y214, Y214-Y55, Y55-Y214	Y214-Y31
Yantai (y)	Y31-Y55, Y214-Y55, Y55-Y214	Y55-Y31, Y31-Y214	Y214-Y31
Yantai (o)	Y31-Y214, Y214-Y31, Y31-Y55, Y55-Y31	Y31-Y214	Y214-Y31

Table 6.26 Confusability rankings of YT tone pairs. Results were obtained from pairwise comparisons of RTs within each listener group.

6.3.6 Summary of discussion on Experiment YT

Analyses on the RT data in Experiment YT revealed some cross-linguistic differences. Perception of YT tones by the Chinese listeners, especially the Yantai older listeners, was again influenced by the sandhi rules in their dialect (e.g. the /Y31/ and /214) rules).

6.4 Summary of Discussion on Experiments BJ, RG and YT

This series of experiments investigated effects of tone sandhi on perception with three different Mandarin dialects, namely Beijing (or Standard Mandarin Putonghua, with four tones and one major neutralization rule, namely the T214 rule as well as the secondary T35 rule), Yantai (a Northern Mandarin dialect with only three citation tone categories but several tonal neutralization rules that changes the first of two consecutive identical tones to a different tone category or a surface tone: namely /Y55.Y55/ → [Y31.Y55], /Y214.Y214/ → [Y55.Y214], /Y31.Y31/ → [Y35.Y31], /Y31.Y31/→[Y55.Y31]), and Rugao (a Southern Mandarin dialect with four tones and no neutralization sandhi). A fourth control group of American English listeners also participated in the experiments. It was found that acoustic similarities in tonal contours and onset f0 may influence the perception of all groups of listeners. Thus, T35 and T214 in Putonghua (Beijing), as well as R44 and R45 in Rugao, were confusable for everyone. But such psychophysical effects may be strengthened or weakened by one's knowledge of a particular tonal system (or systems, as in the case of Yantai young listeners who are highly competent in two dialects). For example, T51/T55 were found confusable by all groups of Chinese listeners (especially the Rugao listeners), but less so by the English listeners. Explicitly taught sandhi rules may have an even stronger effect, as is evidenced by the T214 sandhi effect on the Yantai young listeners' perception of Putonghua (Beijing) tones.

Cross-dialectal tone category correspondence such as R44 to T51 and Y55 to T51 also seem to have an impact on Rugao and Yantai listeners' perception of tones in Putonghua. This effect is most noticeable with the older listeners, most of whom might only have a passive knowledge of the standard dialect and do not use it in their daily speech. This is especially true for the Rugao older listeners, who have a very homogeneous linguistic community. With Yantai being an immigrant city, the Yantai older listeners might have a better knowledge of the standard dialect. But this did not block the Yantai sandhi effects in their perception.

Age difference was found between the two Yantai listener groups in their perception of Yantai tones as well. It seems that the younger Yantai listeners have formed a simpler tonology than the older generation by getting rid of the less transparent rule of /31.31/ → [55.31] that only affects a small subset of /31/ morphemes. As a result, the Y31 and Y55 contrast seems stronger for the younger listeners. This indicates a language change between two generations of Yantai speakers. It seems that learning Putonghua and using it actively in their daily life has also helped the Yantai young listeners form a separate /35/ category. These changes in their linguistic system have enabled the young Yantai listeners to discriminate tone pairs involving the relevant tones better than the older Yantai listeners. It is also likely that information is retrievable from both their native Yantai and the L2 Putonghua lexica in the perceptual processes, although one system may be more active than the other depending on the stimuli (i.e., Putonghua/Beijing stimuli will bring forth their Putonghua lexicon, while Yantai stimuli may activate their Yantai lexicon).

As in the previous experiments reported in earlier chapters, f0 transitions from the offset of the first stimulus tone to the onset of the second stimulus tone are important for the English listeners but not the Chinese listeners.

CHAPTER 7

CONCLUSION

In these concluding remarks, I shall first summarize the main findings from the seven experiments reported in this dissertation. Then, I shall evaluate two speech perception models, namely the neural model of an auditory cortical map (Guenther & Gjaja, 1996; Guenther et al. 1999) and the lexical distance model (Johnson, 2004) against the results from these experiments.

7.1 Summary of Discussion

The data from the tone perception experiments reported in this dissertation point to strong language-specific effects, supporting our hypothesis that tonology influences tone perception. The cross-linguistic differences in tone perception found in these experiments may be attributed to differences in the tone inventories among the different groups of listeners as well as to different tone sandhi rules operating in the listeners' respective native languages/dialects.

In general, we have seen that the American English listeners, with no lexical tones in their native language system, tend to rely more on acoustic cues, especially the pitch height of the onset and offset of a tonal contour, to discriminate tones in a stimulus pair. When these pitch heights differ, as in T214/T55, the tones are perceived as being very different; when these pitch heights are similar, as in T35/T214 and R44/R41 (where the onset pitch heights match) or T35-T51 and T35-T55 (where the offset pitch of T35 and the onset pitch of T51 or T55 match), the tones are perceived as being similar. On the other hand, two tones seem more confusable for the Chinese listeners if the contrast between them has been weakened by a neutralization rule, e.g. the T214 rule as seen in Experiment 1 (Chapter 2), the T214 and T35 rules in Experiments 3, 4 and BJ (Chapters 3 and 6), and the Y214, Y55 and Y31 rules (Experiment YT, Chapter 6).

The fact that T35 and T214 were perceived as being the most similar/confusable by the Chinese listeners in Experiment 1, with pairs T35/T214 showing RTs significantly different from all other tone pairs in their data (as opposed to more maxima and minima on the AE listeners' RT curve, Figure 2.2) points to language-specific differences in the Chinese and the AE listeners' tone perception. Patterns revealed in the RT data and MDS analyses suggest that the two groups used different strategies: the AE listeners focused attention on acoustic details of the start and end pitch points of the contours, while the Chinese listeners might have paid more attention to the contours as a whole. But if the Chinese listeners only compared the contours of the T35 and T214 stimuli in these experiments, they should have found T35 and T214 less confusable in Experiment 1 where T214 had a low (falling) contour (as opposed to the rising contour in T35). Thus, when making the comparison the Chinese listeners must have consulted the representations of T35 and T214 in the lexicon, where all phonetic shapes of T214 may be available and where some morphemes may be cross-specified for both T35 and T214 due to the sandhi rule (Peng, 1996; Wang, 1995). This means that the Chinese listeners may have processed the tone stimuli at a higher cognitive level.

In Experiments 2 and 3 reported in Chapter 3 (as well as Experiment BJ, Chapter 6), where the ISI was shortened to just 100ms and where a low memory demand limited stimulus set AX discrimination task (or a speeded response AX discrimination in Experiment BJ) was used, the RT differences between pairs T35/T214 (as well as T55/T35) and the other tone pairs were smaller in the Chinese listeners' data. These low memory-load tasks probably tapped on lower level auditory processing (e.g., Pisoni, 1973). So, pitch contours of the first and second stimulus tones were compared using memory traces of these tones, which did not decay completely due to the short ISI (further aided by a speeded response in Experiment BJ). That is, the Chinese listeners might have switched from mainly higher level cognitive processing to mainly auditory processing, yielding perceptual data patterns more similar to those found in the AE listeners' data. Yet even under these experimental conditions, there were some differences found between the Chinese (Putonghua) and the AE listener groups, indicating that there were still some language-specific effects, which cannot be accounted for by higher level cognitive processing (Pisoni, 1973; Fox, 1984; Johnson, 1988; Guenther et al. 1999; Johnson, 2004). This suggests that language-specific pattern may also surface in "low-level" auditory responses.

Comparing the performances of the different groups of Chinese listeners in Experiment BJ, we found some interesting cross-dialectal differences, as well as some age group differences between the Yantai young and older listeners. Rugao listeners showed very little age difference. The largest RT differences between Beijing and Rugao listeners were found with pairs T55/T51, which can be explained by the inter-dialectal tone category correspondence of R44 to T51 between Rugaohua and Putonghua/Beijing.

Recall from Chapter 5 that these are also the tone pairs that attracted the most errors in Rugao listeners' data, especially the older listeners. The largest RT differences between Beijing and Yantai older listeners were found with pairs T55/T35, T55-T214, T35-T214 and T35-T51, which can be attributed to inter-dialectal tone category correspondence between the two dialects (Y55 to T51) as well as tone sandhi rules (/Y214.Y214/ → [Y55.Y214]; /Y214.Y31/ → [Y35.Y31]) and historical tone merger (the Yantai counterpart of T35 with Y55) in Yantai. The largest RT difference between Beijing and Yantai young listeners lies in pair T35-T214. It thus seems that learning Putonghua as L2 and acquiring the rule through explicit classroom instruction made Yantai young listeners more conscious about the T214 sandhi rule. This pattern of different degrees of L2 effect was consistent with Elman et al.'s findings (1977). Thus, language-specificity was again found in the supposedly low-level auditory processing tapped in these experiments.

In Experiment RG, we observed the same psychoacoustic listening by the AE listeners. Among the Chinese listeners, Rugao listeners showed some native advantage and made fewer mistakes when discriminating the acoustically similar R44 and R45. There were no obvious age group differences in Rugao and Yantai listeners' RT data, except that the young listeners were faster. The effect of T214 sandhi was not observed in either the Beijing or the Yantai young listeners' data. This is probably because the pitch contours of R45 and R212 do not bear enough resemblance to their Beijing counterparts T35 and T214 to evoke the T214 rule. With the addition of R45 and a different speaker pitch range, the Rugao tones also failed to bring out any Yantai sandhi effects in the Yantai listeners' data. Nevertheless, as is evident the confusability rankings in Table 6.14, language-specificity surfaced again in this experiment.

The Yantai tones discriminated in Experiment YT were both easier (because there were only three tones to compare) and harder (because the speaker had the smallest pitch range) for the listeners. The Yantai listeners' perception was influenced by the tone sandhi rules in their dialect, especially that of /Y214.Y214/ → [Y55.Y214], which is the Yantai counterpart of the T214 rule in Putonghua. The older Yantai listeners' performance was further affected by two other rules, namely /Y55.Y55/ → [Y31.Y55] as well as /Y31.Y31/ → [Y55.Y31], the latter of which may not exist in the younger Yantai listeners' system. Thus, as in the other experiments reported in this dissertation, language-specificity was present in Experiment YT, despite experimental procedures conducive to auditory perception predicted to reveal only universal patterns by various researchers (e.g., Pisoni, 1973; Carney et al., 1977; Fox, 1984; Johnson, 2004).

7.2 Evaluation of Speech Perception Models

To refresh our memory, Johnson's (2004) lexical distance model accounts for language-specific perceptual warping by adjusting the amount of lexical influence on auditory perceptual distance. Specifically, stimuli belonging to non-occurring or infrequent sound categories cause none or little activation in the lexicon and the perceptual distance between them would be determined mainly by their inherent auditory distance. It is predicted by the model that low memory demand tasks such as limited stimulus set AX discrimination or speeded response AX discrimination would tap on the auditory trace mode of stimulus processing. That is, the lexicon will not be consulted in these tasks. As a result, only auditory distance between the stimuli matters. Consequently, no language specific effects would be observed.

The neural model proposed by Guenther and colleagues (Guenther & Gjaja, 1996; Guenther et al., 1999; see also Bauer et al. 1996) suggests that linguistic experience may lead to warping in the auditory cortex such that between-category perception is enhanced and within-category discriminability reduced. Different cross-linguistic perceptual patterns in warped perceptual spaces are then reflections of different landmarks of the auditory maps. The model accounts well for empirical data such as *categorical perception* (Liberman, et al., 1957) and the *perceptual magnet effect* (Kuhl, 1991). It is also supported by fMRI data from neurophysiological studies using synthetic vowel /i/ stimuli (Guenther & Bohland, 2002; Guenther et al. 2004).

A recent perceptual study by McGuire (2004, ms.) also replicated part of Guenther et al.'s (1999) findings with natural speech stimuli. Three groups of American English listeners were tested with naturally produced /fa, θa, xa, ha/ monosyllables, among which the voiceless fricatives /f/ and /θ/ form a native contrast. The non-native /x/ was introduced to form a contrast with the native /h/. Both contrasting pairs involve acoustically similar sounds. One group of AE listeners received categorization training, a second group discrimination training, and a third control group no training. Comparison between the pre-training and post-training AX discrimination (with a short interstimulus interval of 100ms and a speeded response deadline of 500ms) test results revealed larger significant difference for the categorization training group than for the discrimination training group and the control group. Further analysis revealed that only the fricative pair /h/-/x/ showed significant differences between training conditions. When data were analyzed separately for each fricative, only /x/ showed significant training condition effects, with the categorization training group showing a significant decrease in discrimination accuracy between pre- and post-training tests, the discrimination training group a significant improvement in discrimination accuracy, and the control group no difference. Assuming the existence of an auditory map, since the native categories have been formed and reinforced throughout the listener's life, it is hard to imagine that a short training session would alter those categories – hence the auditory map – dramatically. The intense input of the novel /x/ stimuli, on the other hand, should provide some stimulation to the cortical cells, resulting in different effects for the categorization and discrimination training groups. Furthermore, if a short 100ms ISI, a 500ms response deadline and an AX discrimination task tapped auditory listening, such results seem to support to the auditory cortical map proposal (Guenther & Gjaja, 1996; Guenther et al., 1999).

Except for the difference rating task in Experiment 4, the experiments reported in this dissertation all involved tasks of low uncertainty AX discrimination. The ISI in Experiment 1 was a bit longer than in the rest of the experiments, 300ms. But it can still be considered short, given Pisoni's (1973) results. Experiments 2 and 3 involved a short 100ms ISI and a limited stimulus set AX discrimination task using non-speech synthetic tones and natural speech tones, respectively. Experiments BJ, RG and YT used a short 100ms ISI and a speeded AX discrimination with a 500ms response deadline. Although we were only able to bring three groups of listeners' (namely, Beijing, AE and the Yantai young listeners) mean RTs within 500ms in these last three experiments, in a sense the deadline worked with the other groups as well, for RTs were much shorter than in Experiment 3, where no response deadline was set. In any case, these experiments should have tapped the auditory trace processing mode more than some higher level processing mode. According to the predictions by the lexical distance model (Johnson, 2004), no language-specific effects would be found. The fact that the T214 sandhi effect was weakened in Experiments 2, 3 and BJ is certainly consistent with these predictions. But the results reported in this dissertation also suggests that the model may have placed too much restriction on how much lexical influence is allowed in the so-called low-level auditory perception tasks. The language-specific patterns observed in our data seem to support the hypothesis of an auditory cortical map, which has a neurophysiological basis and whose landmarks should be reflected in the perception data, regardless of the task. But the neural model, as it stands now, does not provide an explicit treatment for different degrees of the language-specific effects. Although Guenther et al. (1999) reported that a short ISI (250ms) and no interfering noise between two stimuli "basically eradicates" the effect (1999: 2909), it is not clear how the suppression of auditory warping effect is handled by the model. We may infer from the fMRI data reported in Guenther & Bohland (2002) and Guenther et al. (2004) that attention shifts from one cortical area to another may partially explain such differences. If this is the case, we may actually have an auditory cortical map that is warped to efficiently serve our specific linguistic needs in normal speech communication situations. We may further hypothesize that this auditory map (with a neurophysiological basis) and the lexical distances (calculated through consultation to the lexicon at a higher cognitive level) complement each other in accounting for language-specificity at different levels of auditory stimuli processing.

BIBLIOGRAPHY

Abramson, A., 1962. The vowels and tones of Standard Thai: Acoustic measurements and experiments. In *International Journal of American Linguistics*, 28 (2).

Abramson, A. & L. Lisker, 1970. Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the Sixth International Congress of phonetic Sciences* (Academia, Prague), pp. 569-563.

Baxter, William H., 1992. *A Handbook of Old Chinese Phonology*. Berlin, New York: Mouton de Gruyter.

Beckman, Mary Esther, 1984. Toward Phonetic Criteria for a Typology of Lexical Accent. Ph.D. dissertation, Cornell University.

Beddor, P. S. & W. Strange, 1982. Cross-language study of perception of the oral-nasal distinction. In *Journal of the Acoustical Society of America*, 71 (6), 1551-1561.

Best, C. T., 1995. A Direct Realist View of Cross-Language Speech Perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience Issues in Cross-Language Research* (pp. 171-204). Baltimore: York Press.

Best, C. T., G. W. Mc Roberts, and E. Goodell, 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. In *Journal of the Acoustical Society of America,* 109 (2), 775-794.

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

Brown, Roger W. and Donald C. Hildum, 1956. Expectancy and the perception of syllables. In *Language*, 32, 411-419.

Carney, A. E., G. P. Widin & N. F. Viemeister, 1977. Noncategorical perception of stop consonants differing in VOT. In *Journal of the Acoustical Society of America*, 62: 961-970.

Carroll, J. D. and J.-J. Chang, 1970. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. In *Psychometrika*, Vol. 35, No. 3, 283 – 319.

Chan, M. K.M., 1983. Lexical diffusion and two Chinese case studies re-analyzed. In *Acta Orientalia*, 44:118-152.

Chan, M. K.M., 1985 Fuzhou Phonology: A Non-Linear Analysis of Tone and Stress. Ph.D. dissertation, University of Washington, Seattle, U.S.A.

Chan, Marjorie K. M., 1989. On the status of "basic" tones. In *Acta Linguistica Hafbiensia*, 21.2: 5-34. The Linguistic Circle of Copenhagen.

Chang, Kun, 1975. Tonal Development among Chinese Dialects. In *BIHP*, Vol. 46: 636-709.

Chao, Y-R (赵元任), 1928. (in Chinese) Xiandai Wuyu de Yanjiu. [Research on Modern Wu Dialects]. Beiping: Tsinghua Xuexiao Yanjiuyuan.

-----, 1930. A System of Tone Letters. *La Maître phonétique* 45:24-27.

-----, 1935. *Xin Guoyu Liushengpian Keben*. Shanghai: Commercial Press.

-----, 1948. *Mandarin Primer.* Cambridge, MA: Harvard University Press.

-----, 1965. *A Grammar of Spoken Chinese*. Berkeley and Los Angeles: University of California Press.

-----. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chen, Matthew Y., 1976. From Middle Chinese to Modern Peking. In JCL, 4.2/3: 113-277.

Chen, Matthew Y. 2000. *Tone Sandhi patterns across Chinese dialects*. Cambridge, UK: Cambridge University Press.

Cheng, Chin-Chuan, 1968. English stresses and Chinese tones in Chinese sentences. Phonetica, 18: 77-88.

-----. 1973. *A Synchronic Phonology of Mandarin Chinese.* The Hague: Mouton.

Chomsky, N. & M. Halle, 1968. *The sound pattern of English*. New York, Harper & Row.

Chuang, C.K., S. Hiki, T. Sone, & T. Nimura, 1971. The acoustic features and perceptual cues of the four tones of standard colloquial Chinese. 7th International Congress on Acoustics, Budapest 1971.

Curtis, D. W., Paulos, M. A., & S. J. Rule, 1973. Relation between disjunctive reaction time and stimulus difference. *Journal of Experimental Psychology*.

Downer, G. B., 1963. Chinese, Thai, and Miao-Yao. In H. L. Shorto (ed.), *Linguistics Comparison in Southeast Asia and the Pacific*. London.

Duanmu, S., 1997. "Recursive constraint evaluation in Optimality Theory: evidence from cyclic compounds in Shanghai," *Natural Language and Linguistic Theory* 15: 465-507.

Duanmu, S., 2000. *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Elman, J. L., R. L. Diehl & S. E. Buchwald, 1977. Perceptual switching in bilingualism. In Journal of the Acoustical Society of America, Vol. 62 (4): 971-974.

Flege, James E. and Chipin Wang, 1989. Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/ - /d/ contrast. In Journal of Phonetics (1989) 17, 299-315.

Fon, J., Chiang W.-Y., & H. Cheung, 1999. Production and Perception of T2 and T3 in Taiwan Mandarin. Ms.

Fox, Robert Allen, 1984. Effect of Lexical Status on Phonetic Categorization. In *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 10, No. 4, 526-540.

Fox, R. A. and Y.-Y. Qi, 1982. Context effects in the perception of lexical tone. In Journal of Chinese Linguistics, Vol. 18, 261-284.

Fox, Robert Allen, 1992. Perception of Vowel Quality in a Phonologically Neutralized Context, in Y. Tokura, E. Vatikiotis-Bateson, & Y. Sagisaka (eds.), *Speech Perception*, *Production and Linguistic Structure*, Tokyo: Ohmsha and IOS Press (jointly), pp. 21-42.

Fox, R. A. and J. Unkefer, 1985. The Effect of Lexical Status on the Perception of Tone. In *Journal of Chinese Linguistics*, Vol.13, No.1, 69-90.

Gandour, Jack, 1981. Perceptual dimensions of tone: Evidence from Cantonese. In *Journal of Chinese Linguistics,* 9: 20-36.

Gandour, Jack, 1983. Tone perception in Far Eastern languages. In *Journal of Phonetics,* 11: 149-176.

Gandour, Jack, 1984. Tone Dissimilarity Judgments by Chinese Listeners. In *Journal of Chinese Linguistics,* 12.2: 235-61.

Gandour, J. T. & R. A. Harshman, 1978a. Cross-Language Differences in Tone Perception: A Multidimensional Scaling Investigation. In *Language and Speech* 21: 1-33.

Gandour, J. T. & R. A. Harshman, 1978b. Cross-Language Study of Tone Perception. In D. Sankoff (ed.), *Linguistic Variation*: *Models and Methods*. Pp.139-147. New York: Academic Press.

Ganong, William F., 1980. Phonetic Categorization in Auditory Word Recognition. In *Journal of Experimental Psychology: Human Perception and performance*, Vol. 6, No. 1, 110-125.

Goldstone, R. L., 1998. Perceptual learning. In *Annual Review of Psychology*, 49: 585-612.

Gottfried, T. L. & T. L. Suiter, 1997. Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics* 25: 207-231.

Guenther, Frank H. & Marin Gjaja, 1996. The Perceptual Magnet Effect as an Emergent Property of Neural Map Formation. In *Journal of the Acoustical Society of America* (1996), Vol. 100, pp. 1111-1121.

Guenther, Frank H., Ratima T. Husain, Michael A. Cohen & Barbara G. Shinn-Cunningham, 1999. Effects of Categorical and Discrimination Training on Auditory Perceptual Space. In *Journal of the Acoustical Society of America*, 23(4), pp. 213-221.

Guenther, Frank H. & Jason W. Bohland, 2002. Learning Sound Categories: A Neural Model and Supporting Experiments. In *Acoustical Science and Technology* (1996), Vol. 100, pp. 1111-1121.

Guenther, F.H., Nieto-Castanon, A., Ghosh, S.S., and Tourville, J.A. (2004). Representation of sound categories in auditory cortical maps. J*ournal of Speech, Language, and Hearing Research,* 47(1), pp. 46-57.

Guion, S. G., 1998. The Role of Perception in the Sound Change of Velar Palatalization. In *Phonetica*, 55: 18-52.

Hall, K. C., E. Hume, K. Johnson & A. Reiter, 2004. The Influence of Language Experience on Perception: The Case of Spanish and English . Paper presented to the OSU Hispanic Linguistics Colloquium. May 28, 2004.

Harnsberger, J. D., 2001. The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis. In *Journal of Phonetics*, 29: 303-327.

Haudricourt, Andre-Georges, 1954a. De L'Origine des Tons en Vietnamien. In *Journal Asiatique* 242, 69-82.

Haudricourt, Andre-Georges, 1954b. Comment Reconstruire le Chinois Achaique. In *Word* 10, 351-64.

Ho, Da-an 何大安, 1994. 声调的完全回头演变是否可能? [Can tone change process be completely reversed?] In *BIHP*, Vol. 65: 1-18.

Hombert, J.-M., 1978. Consonant Types, Vowel Quality, and Tone. In Victoria Fromkin (ed.), *Tone: A Linguistic Survey*. Pp. 77-111. New York: Academic Press.

Hombert, J.-M., J. J. Ohala, and W. G. Ewan, 1979. Phonetic Explanations for the Development of Tones. In *Language*, 55(1): 37-58.

Huang, Tsan, 2001. The interplay of perception and phonology in Tone 3 sandhi in Chinese Putonghua. In Hume, E. and K. Johnson (eds.), *Studies on the interplay of speech perception and phonology* (OSU working papers in linguistics, No. 55), pp. 23-42.

Huang, Tsan, 1999. A First Look at Rugaohua. Paper presented at the Colloquium Fest, The Ohio State Unviersity, Department of Linguistics.

Huang, Tsan, 2002. Downstep and other tonal interactions in Rugaohua. Paper presented at the Linguistics and Phonetics Conference, September 2 – 6, Tokyo, Japan. Also published in the LP proceedings.

Hume, E. & K. Johnson, 1999. The Role of Speech Perception Phenomena in Phonology. Paper presented at the ICPhs satellite meeting "The Role of Speech Perception in Phonology", July, 1999, San Francisco, California.

Hume, E. & K. Johnson, 2001. A Model of the Interplay of Speech Perception and Phonology. In E. Hume & K. Johnson (eds.) *The Role of Perception in Phonology*. New York: Academic Press.

Hume, E. & K. Johnson, 2003. The Impact of Partial Phonological Contrast on Speech Perception. In *Proceedings of the XVth International Congress of Phonetic Sciences*.

Hume, E., K. Johnson, M. Seo, G. Tserdanelis & S. Winters, 1999. A Cross-linguistic Study of Stop Place Perception. In *Proceedings of the XIVth International Congress of Phonetic Sciences*. Pp. 2069-2072.

Hume, E. & M. Seo. 2004. From Speech Perception to Optimality Theory: Metathesis in Faroese and Lithuanian. In *Nordic Journal of Linguistics* 27:1. 35-60.

Hura, S. L., B. Lindblom, & R. L. Diehl, 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech*, 35(1,2): 59-72.

International Phonetic Association, 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

Iverson, P. & P. Kuhl, 1996. Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. In *Journal of the Acoustical Society of America*, 99 (2), 1130-1140.

Jakobson, R., Fant, G., and M. Halle, 1952. *Preliminaries to speech analysis*: the distinctive features and their correlates. Cambridge: Acoustics Laboratory, Massachusetts Institute of Technology.

Jakobson, R. & M. Halle, 1956. *Fundamentals of language*. Gravenhage: Mouton & Co.

Janson, Tore, 1983. Sound change in perception and in production. In *Language,* 59 (1): 18 – 34.

Jiangsu sheng he Shanghai shi fangyan diaocha zhidao zu (ed.), 1960. *Jiangsu Sheng he Shanghai Shi Fangyan Gaikuang* 《江蘇省和上海市方言概況》[Language situations in Jiangsu Province and Shanghai City]. Nanjing: Jiangsu Renmin Chubanshe. (In text as JSFG, 1960.)

Johnson, Keith, 1988. Processes of Speaker Normalization in Vowel Perception. Unpublished PhD dissertation, Department of Linguistics, The Ohio State University.

-----, 2004. Cross-linguistic perceptual differences emerge from the lexicon. In Augustine Agwuele, Willis Warren, and Sang-Hoon Park (eds.) *Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception.* Sommerville, MA: Cascadilla Press. pp. 26-41.

Kiriloff, C. 1969. On the Auditory perception of Tones in Mandarin. *Phonetica* 20: 63-67.

Kohler K. 1990. Segmental reduction in connected speech: Phonological facts and phonetic explanations. In W.J. Hardcastle & A. Marchal (eds.) *Speech Production and Speech Modeling* (pp. 69-72). Dordrecht: Kluwer Academic Publishers.

Kratochvil, Paul, 1968. *The Chinese Language Today: features of an emerging standard*. London: Hutchinson.

Kuhl, Patricia K., 1991. Human adults and human infants show a "perceptual magnet effect" for the protypes of speech categories, monkeys do not. In *Perception & Psychophysics*, 1991, 50 (2), 93-107.

Kuhl, P. K., K. A. Williams, F. Lacerda, K. N. Stevens, & B. Lindblom, 1992. Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. In *Science*, Vol. 255: 606-607.

Lahiri, A. & W. D. Marslen-Wilson, 1991. The mental representation of lexical form: a phomological approach to the recognition lexicon. In *Cognition*, 38: 245-294.

Lea, W. A., 1973. Segmental and Suprasegmental Influences on Fundamental Frequency Contours. In L. Hyman (ed.), *Southern California Occasional Papers in Linguistics, No.1: Consonant types and tones*. Pp. 15-70.

Lee, Y-S, Vakoch, D.A., & L. H. Wurm, 1996. Tone Perception in Cantonese and Mandarin: A Cross-Linguistic Comparison. *Journal of Psycholinguistic Research*, 25: 527-542.

Lehiste, Ilse and G. E. Peterson, 1961. Some basic considerations in the analysis of intonation. In *Journal of the Acoustical Society of America*, 33: 419-425.

Li, Fang Kuei, 1980. Laryngeal Features and Tone Development. In *BIHP*, Vol. 51: 1-13.

Li, Rong 李荣. 1989a. "Zhongguo-de yuyan he fangyan" [Languages and dialects in China] Fangyan 《方言》1989.3: 161-167.

Li, Rong 李荣, 1989b. Hànyǔ fāngyán de fēnqū. Fāngyán 汉语方言的分区, [Classification of the Chinese dialects] *Fangyan* 《方言》1989.4: 241-261. Běijīng.

Liberman, A. M., K. S. Harris, H. S. Hoffman & B. C. Griffith, 1957. The discrimination of speech sounds within and across phoneme boundaries. In *Journal of Experimental Psychology*, 54: 358 – 368.

Lien, Chinfa, 1986. Tone merger in the dialects of Northern Chinese. In *Journal of Chinese Linguistics*, 14.3: 243-291.

Liljencrants, J. & B. Lindblom, 1972. Numerical simulation of vowel quality systems: the role of perceptual contrast. In Language, Vol. 48, No. 2, pp. 839-862.

Lin, H.-B., and B. H. Repp, 1989. Cues to the perception of Taiwanese tones. In *Language and Speech*, 32(1): 25-44.

Lin & Wang, 1985. In *Journal of Chinese Linguistics*.

Lindblom, B., 2000. Developmental Origins of Adult Phonology; The Interplay between Phonetic Emergents and the Evolutionary Adaptations of Sound Patterns. In *Phonetica*, 57: 297-314.

Macmillan, N. A., 1987. In Harnad, S. (ed.) *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press.

Makashay, M.J. & Johnson, K. (1998) Surveying auditory space using vowel formant data. *The Sound of the Future: A Global View of Acoustics in the 21st Century, Proceedings 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America.* pp. 2037-8.

Maran, L., 1973. On becoming a tone languages: a Tibeto-Burman modal of tonogenesis. In Larry Hyman (ed.), *Consonant Types and Tone*, 97-114. Southern California Occasional Papers in Linguistics, No.1.

Maspero, H., 1912. Etudes sur la phonetique historique de la langue annamite: les initiales. In *Bulletin de l'Ecole Francaise d'Extreme Orient* 12.

Massaro, D. W. and M. M. Cohen, 1983. Phonological context in speech perception. In Perception & Psychophysics, 1983, 34 (4), 338-348.

Massaro, D. W., M. M. Cohen & Tseng, 1985.

Matisoff, James A., 1973. Tonogenesis in Southeast Asia. In Larry Hyman (ed.), *Consonant Types and Tone*, 71-95. Southern California Occasional Papers in Linguistics, No.1.

Mei, T-L, 1977. Tones and Tone Sandhi in 16th Century Mandarin. In *Journal of Chinese Linguistics* 5: 237－260.

Mielke, J., 2003. The Interplay of Speech Perception and Phonology: Experimental Evidence from Turkish. In *Phonetica* 60 (3): 208-229.

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

Norman, Jerry, 1973. Tonal Development in Min. In *Journal of Chinese Linguistics*, Vol. 1, No.2, 222-238.

Nosofsky, R. M. 1992. Similarity scaling and cognitive process models. In *Annual Review of Psychology*, 43: 25-53.

Ohala, J., 1981. The listener as a source of sound change. In C. S. Masek, R. A. Hendrik, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior: Chicago Linguistics Society* (pp. 178-203). Chicago: Chicago Linguistics Society.

Peng, S-H, 1996. Phonetic Implementation and Perception of Place Coarticulation and Tone Sandhi. PhD dissertation, The Ohio State University.

Pisoni, D., 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. In *Perception & Psychophysics* 13: 253-260.

Pitt, Mark A., 1998. Phonological process and the perception of phonotactically illegal consonant clusters. In *Perception & Psychology*, 60 (6), 941-951.

Polka, L., 1991. Cross-language speech perception in adults: phonemic, phonetic, and acoustic contributions. In *Journal of the Acoustical Society of America*, 89 (6), 2961-2977.

Qian, Zengyi, 1982. (in Chinese) *Yantai Fangyan Baogao*. Jinan : Qi Lu Shu She.

Qian, Z., S. Zhang & F. Luo, 2001. (in Chinese) *Shangdong Fangyan Yanjiu*. Jinan : Qi Lu shu she.

Rúgāo Xiànzhi editing committee, 1994. Fāngyán 方言 [Dialect], in Lǐ Shíqiū, Cài Chūnmíng, and Shào Yuánlǐ (eds.), *Rúgāo xiànzhi*《如皋县志》, Vol. 20. Hongkong: New Asia Press. (In text as RGXZ, 1994.)

Seo, M., 2001. A Perception-based Study of Sonorant Assimilation in Korean. In Hume, E. and K. Johnson (eds.), *Studies on the interplay of speech perception and phonology* (OSU working papers in linguistics, No. 55), pp. 43-69.

Seo, M. & E. Hume, 2001. A comparartive account of metathesis in Faroese and Lithuanian. In E. Hume, N. Smith, & J. van de Weijer (Eds.), *Surface syllable structure and segment sequencing*. Leiden: Holland Institute of Linguistics

Shanghai Library, 1980. 钜宋广韵 *Ju Song Guang Yun* [The Broad Rhymes]. 上海古籍出版社 Shanghai Classical Publisher.

Shen, Jiong, 1994. (in Chinese**)** 北京话上声连读的调型组合和节奏形式 [Tone 214 Sandhi in Connected Beijing Speech], in *Zhongguo Yuwen*. Vol. 241. pp. 274-281.

Shen, X. S. & Lin, M. 1991. A perceptual study of Mandarin Tone 2 and 3. *Language and Speech*, 34 (2), 145-156.

Shepard, R. N.1962. The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function (Part II). In *Psychometrika*, 27: 219- 246.

-----, The circumplex and related topological manifolds in the study of perception. In Shye, S. (Ed.), *Theory construction and data analysis in the social sciences*. San Francisco: Jossey-Bass, 1978.

Shepard, R. N., Kilpatric, D. W., & J. P. Cunningham, 1975. The internal representation of numbers. *Cognitive Psychology*, 1975, 7, 82-138.

Shih, Chi-Lin, 1988. Tone and Intonation in Mandarin. In *Working Papers of the Cornell Phonetics Laboratory* 3: 83-109.

-----, 1997. Mandarin third tone sandhi and prosodic structure. In Wang Jialing and Norval Smith (eds.) *Studies in Chinese Phonology*. Berlin & New York: Mouton de Gruyter, 1997.

Stagray, J. R., & D. Downs, 1993. Differential sensitivity for frequency among speakers of a tone and a nontone language. In Journal of Chinese Linguistics, 21 (1): 143-163.

Steriade, D., 2001. A perceptual account of directional asymmetries in assimilation and cluster reduction. In E. Hume & K. Johnson (eds.) *The Role of Perception in Phonology*. New York: Academic Press.

Strange, W., 1995. Cross-Language Studies of Speech Perception: A Historical Review. In Strange, W. (Ed.), *Speech Perception and Linguistic Experience Issues in Cross-Language Research*. Baltimore: York Press.

Svantesson, Jan-Olof, 2001. Tonogenesis in Southeast Asia – Mon-Khmer and beyond. In Shigeki Kaji (ed.), *Proceedings of the Symposium: Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Japanese Accentology, and Other Topics*. ILCAA, Tokyo University of Foreign Studies. Pages 45-58.

Tai, James H.-Y., and Marjorie K.-M. Chan, 1999. Some Reflections on the Periodization of the Chinese Language. In Peyraube, A. and Sun C.-F. (eds.), *Studies in Chinese*

Historical Syntax and Morphology: Linguistic Essays in Honour of Mei tsu-lin [=Collection des Cahiers de Linguistique d'Asie Orientale]. Pages 223-239.

Takane Y. & J. Sergent, 1983. Multidimensional Scaling Models for Reaction Times and Same-different Judgments. *Psychometrika*, 1983, Vol. 48, No. 3, 393-423.

Takane, Y., F. W. Young, & J. DeLeeuw, 1977. Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika* 42: 7-67. (ALSCAL)

Tan, Fu, 1987. Tone Sandhi in the Tianjin Dialect. In *Journal of Chinese Linguistics*, 15.2: 228-246.

Ting, Pang-hsin 丁邦新, 1966. Rugao Fangyan de yinyun 如皋方言的音韵 [Phonology of the Rugao Dialect]. Zhongyang Yanjiuyuan lishi yuyan yanjiusuo jikan, Vol. 36: 573-633. Taipei.

Ting, Pang-hsin 丁邦新, 1989. 汉语声调的演变 [Tonal Development in Chinese], in 中央研究院第二界国际汉学会议论文集, 台北, 395-408. [Reprinted in 丁邦新语言学论文集， 106-126. Beijing: Commercial Press, 1998.]

Trubetzkoy, N. S. 1969. *Principles of phonology*. C. Baltaxe (translator). Berkeley and Los Angeles: University of California Press, 1969. (Translated from N. S. Trubetzkoy, 1939. *Grundzuger der Phonologie*. Travaux du Cercle Linguistique de Prague 7, Prague.)

Tserdanelis, G., 2001. A Perceptual Account of Manner Dissimilation in Greek. In Hume, E. and K. Johnson (eds.), *Studies on the interplay of speech perception and phonology* (OSU working papers in linguistics, No. 55), pp. 172-199.

Wang, Li 王力, 1983. 汉语语音史上的条件音变 [Conditioned Changes in the History of Chinese Phonology], in 语言研究 *Yuyan Yanjiu* (Half-yearly), 5: 1-5.

Wang, S.-H. & J.-L. Huang, 1996. Yangzhou Fangyan Cidian [Dictionary of the Yangzhou Dialect]. Nanjing: Jiangsu Jiaoyu Chubanshe.

Wang, S-Y. William, 1967. Phonological features of tone. In *International Journal of American Linguistics*, 2: 93-105.

-----, 1976. Language change. In *Annals of New York Academy of Sciences*. Pp.61-72.

Wang, W. S-Y. & Li, K-P., 1967. Tone 3 in Pekinese. *Journal of Speech and Hearing Research*, 10: 629-636.

Wang, H. Samuel, 1993. (In Chinese.) The Psychological Characteristics of Taiwanese Tone Sandhi. In *Journal of Tsinghua University* (Taiwan), [new] Vol. 23 (2): 175-192.

Werker, J. F. and R. C. Tees, 1984. Phonemic and phonetic factors in adult cross-language speech perception. In *Journal of the Acoustical Society of America*, 75(6), 1866-1878.

Werker, J. F. and J. S. Logan, 1985. Cross-language evidence for three factors in speech perception. In *Perception & Psychophysics*, 37 (1), 35-44.

Wú, Fèngshān 吴凤山, forthcoming. (In Chinese.) *Rúgāo fāngyān yánjiū* 《如皋方言研究》[Research on the Rugao Dialect].

Xu, Yi, 2001. Fundamental Frequency Peak Delay in Mandarin. *Phonetica* 58:26-52.

Xu, Yi, 1997. Contextual tonal variations in Mandarin**.** *Journal of Phonetics*, 25:1; 61-83.

Yip, M. 1980. The Tonal Phonology of Chinese. Ph.D. dissertation. MIT. Distributed by the Indiana University Linguistics Club.

Zhang, Zheng-sheng, 1988. Tone and Tone Sandhi in Chinese. OSU Ph.D. dissertation.

Zhou, Deqing, 周德清 1324 ??. 《中原音韵》*Zhongyuan Yinyun* [Sounds and Rhymes of the Central Plain]. Reprinted in Chen, Xinxiong 陈新雄, 1976. 中原音韵概要 *Zhongyuan Yinyun Gaiyao* [An Introduction to Zhongyuan Yinyun]. Taipei: Xuehai Press.

Zong, Fu-Bang, 1964. (In Chinese.) On the split of the Yin-ping tone in Cantonese. In *Zhongguo Yuwen*, 132: 376-389.