

The Handbook of East Asian Psycholinguistics

Volume 1: Chinese

Edited by

Ping Li

University of Richmond

Li Hai Tan

University of Hong Kong

Elizabeth Bates

University of California, San Diego

Ovid I. L. Tzeng

Academia Sinica, Taiwan

General Editor:

Ping Li

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo
Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521833332

© Cambridge University Press 2006

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2006

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this book is available from the British Library

ISBN 10 0 521 83333 7 hardback

ISBN 13 978 0 521 83333 2



Cambridge University Press has no responsibility for the persistence or accuracy of
URLs for external or third-party internet websites referred to in this book, and does not
guarantee that any content on such websites is, or will remain, accurate or appropriate.

The Handbook of East Asian Psycholinguistics

Volume 1: Chinese

Edited by

Ping Li

University of Richmond

Li Hai Tan

University of Hong Kong

Elizabeth Bates

University of California, San Diego

Ovid I. L. Tzeng

Academia Sinica, Taiwan

General Editor:

Ping Li

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521833332

© Cambridge University Press 2006

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2006

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this book is available from the British Library

ISBN 10 0 521 83333 7 hardback

ISBN 13 978 0 521 83333 2



Cambridge University Press has no responsibility for the persistence or accuracy of
URLs for external or third-party internet websites referred to in this book, and does not
guarantee that any content on such websites is, or will remain, accurate or appropriate.

18 Perception and production of Mandarin Chinese tones

*Allard Jongman, Yue Wang, Corinne B. Moore,
and Joan A. Sereno*

Introduction

Lexical tones are pitch patterns that serve to provide contrasts in word meaning. They perform this function analogously to segments such as voiced and voiceless stops, except that tone may extend beyond one segment. Mandarin Chinese phonemically distinguishes four tones, with tone 1 having high-level pitch, tone 2 high-rising pitch, tone 3 low-dipping pitch, and tone 4 high-falling pitch (Chao, 1948). The same segmental context carries different meanings depending on the tone. For example, the meaning of Mandarin Chinese *ma* with tone 1 is "mother," the tone 2 version means "hemp," and the tones 3 and 4 meanings are "horse" and "scold," respectively.

Pitch or tone is a function of the rate of vocal fold vibration (see Ohala, 1978, for a review of tone production). As stipulated in the myoelastic-aerodynamic theory of vocal fold vibration (e.g. Van den Berg, 1958), it is a cyclic process that is achieved by a complex combination of aerostatic and aerodynamic forces, operating together with air flow and laryngeal muscle forces. Changes in fundamental frequency (or in rate of vocal fold vibration) are made by manipulating tension in the vocal folds. This tension is increased or decreased by the laryngeal muscles, particularly the cricothyroid muscle and the thyroarytenoids. The cricothyroid has been linked to both rises and falls in pitch, while sternohyoid activity is associated with pitch lowering (Sagart et al., 1986).

The rate of vocal fold vibration is quantified as the fundamental frequency (F0), expressed in Hertz (Hz). For example, when a speaker has a F0 of 300 Hz, her vocal folds are vibrating at a rate of 300 times per second. Rate of vocal fold vibration gives rise to the perceived pitch of speech sounds. Although the terms *pitch* and *fundamental frequency* are often used interchangeably, there is not a direct, one-to-one correspondence between perceived pitch and fundamental frequency. Psychophysical studies with pure tones suggest that while the perception of pitch as a function of fundamental frequency is approximately linear below 500 Hz, changes in pitch above that level are perceived more closely on a logarithmic scale relative to the physical changes in fundamental frequency.

Table 18.1 Average F0 values (in Hz) of the four Mandarin tones produced by one male and one female speaker (Wu, 1986)

	Tone 1	Tone 2	Tone 3	Tone 4
Male	190-177	124-178	124-68-141	223-80
Female	307-305	222-318	221-165-242	352-166

Note: The first and last values represent the onset and offset values, respectively, while the middle value for tone 3 represents the fundamental frequency at the Turning Point.

Acoustic and perceptual characteristics of Mandarin tones

The four lexical tones in Mandarin Chinese are typical of many other tonal inventories in Chinese languages because F0 contour as well as F0 height is used to make phonemic distinctions. Phonetic analysis of these tones typically centers on measurement of fundamental frequency, although intensity and duration parameters have also been examined.

Fundamental frequency

Phonetic studies have examined the fundamental frequency contours of Mandarin Chinese tones (e.g. Liu, 1924; Dreher & Lee, 1966; Wang, 1967; Dreher, Young & Lee, 1969; Howie, 1970; Chuang et al., 1972; Rumanancev, 1972; Moore & Jongman, 1997). These studies indicate that F0 height and F0 contour are the primary acoustic parameters to characterize Mandarin tones. Wu (1986) measured F0 values of the four tones from narrow-band spectrograms. His measurements are shown in table 18.1.

These acoustic characteristics are in good agreement with Chao's (1948) phonological description. Figure 18.1 shows an illustration of the F0 contours of the four Mandarin tones produced by a female speaker. In general, tone 1 is high and relatively level over most of the tone's duration. Tone 2 exhibits a rise for much of its duration, and the onset of the rise occurs in the middle region of the F0 range and ends at a point approaching the F0 height of tone 1. The tone 3 contour occupies the lowest region of the F0 range overall, although extending at least to the midpoint of the range by the offset. Tone 3 onset is variable and can be close in frequency to that of tone 2. Tone 4 begins high and falls to the bottom of the range.

The importance of F0 as a primary cue to the perception of Mandarin tones is supported by a variety of studies, including Howie (1976) who performed perception experiments using synthetic speech. Three conditions were contrasted—synthetic speech modeled after natural F0 patterns, synthetic stimuli in which the F0 contours were made to sound monotone, and stimuli synthesized to sound

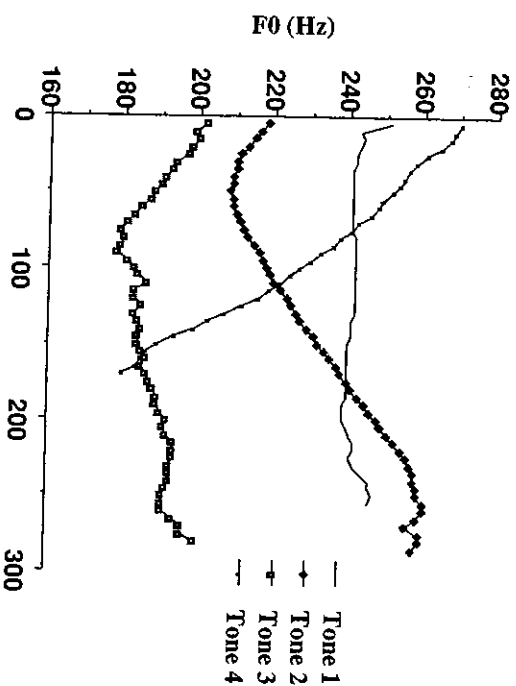


Figure 18.1 F0 contours for each of the four Mandarin Chinese tones for the segmental context *ma* spoken in isolation by a female speaker (Moore & Jongman, 1997)

like a whisper. Results showed that subjects were much better at identifying stimuli in which the pitch pattern was maintained; the monotone and whisper stimuli were identified just slightly above chance level.

Gandour (1984) explored whether F0 height or F0 contour serves as a more important perceptual cue. His results suggest that while both dimensions are important, Mandarin listeners seem to attach slightly more importance to F0 contour than F0 height. However, in their investigation of only tones 1 and 2, Massaro, Tseng, and Cohen (1985) reported that neither F0 height nor F0 contour alone was sufficient for correct identification, but rather that both cues were used by subjects.

Some tones exhibit similar F0 contours. Acoustic properties for these particular tones have been studied in an effort to discover the extent to which F0 may distinguish the tones. Specifically, tones 2 and 3 can be characterized in terms of two measures: Turning Point and $\Delta F0$ (Moore & Jongman, 1997). Turning Point is the point in time at which the contour changes from falling to rising, and $\Delta F0$ is the F0 change from the onset of the tone to the Turning Point. Tone 2 typically has an earlier Turning Point and smaller $\Delta F0$ than tone 3. Perception experiments have shown that these two measures are important cues to the distinction between tones 2 and 3 (e.g. Blicher, Diehl, & Cohen, 1990; Shen & Lin, 1991; Shen, Lin & Yan, 1993; Moore & Jongman, 1997).

Similarly, acoustic analyses of Mandarin tones 3 and 4, both of which have a falling contour over some portion of the tone, suggest several invariant properties of each tone: tone 4 starts at a peak, or reaches one very quickly, then gradually falls to its low point, which may not be achieved until the next syllable; tone 3, on the other hand, may either have a lower onset F0 than tone 4, or it may reach its valley quickly, to maintain a low F0 level particularly over the second half of the vowel (Gårding et al., 1986). Perception experiments in Garding et al. (1986) also showed that stimuli containing an early peak and a steep fall after the Turning Point were perceived as tone 4, while stimuli with long durations at a low F0 level were perceived as tone 3. These results concur with those of earlier studies indicating that F0 is the primary cue in tone perception.

Temporal properties of tones

Production data have shown that, in addition to F0 differences, tones may maintain consistent temporal differences. Mandarin tones differ in terms of overall duration (Dreher & Lee, 1966; Kratochvil, 1971; Chuang et al., 1972; Howie, 1976; Nordenhake & Svantesson, 1983). Tones 2 and 3 tend to be the longest, tone 4 the shortest. The relative duration of any tone may change as a function of position in the sentence, however (Nordenhake & Svantesson, 1983). Blicher, Diehl, and Cohen (1990) showed that overall duration differences between tones 2 and 3 in Mandarin are perceptually relevant in that stimuli ambiguous between tones 2 and 3 were identified more often as tone 3 when they were lengthened. In addition to overall duration, the temporal location of the Turning Point has been evaluated as a primary cue to the distinction between tones 2 and 3. While Turning Point and $\Delta F0$ together typically specify tonal categories, it has been shown that, within a certain range of $\Delta F0$, Turning Point alone can change listeners' percepts from tone 2 to tone 3 (e.g. Shen & Lin, 1991; Shen, Lin & Yan, 1993; Moore & Jongman, 1997).

Amplitude

In addition to F0 and duration, studies have also investigated amplitude and its perceptual relevance in tone perception. Chuang et al. (1972) showed that tone 4 has the highest overall amplitude while tone 3 has the lowest. Lin (1988) attempted to categorize tone amplitude into five patterns: level, higher at onset, higher at offset, higher in the middle, and a double-peak amplitude contour. However, these patterns did not occur systematically with any one tone type, except for the double-peak pattern, which always occurred for tone 3 tokens produced by male speakers. Lin's subsequent perceptual results suggest that F0

is the predominant cue, while duration and amplitude manipulations have little effect on tone perception.

There is some evidence, however, that amplitude contours alone can be utilized as perceptual cues. Using signal-correlated noise stimuli, which removes F0 and formant structure of natural speech but retains amplitude information, Whalen and Xu (1992) showed that subjects were able to identify all but tone 1 tokens from the amplitude contours alone.

In sum, phonetic studies on acoustic characteristics of tones have found that fundamental frequency, $\Delta F0$, Turning Point, duration, and amplitude all constitute phonetic correlates and perceptual cues for tone, with fundamental frequency typically being the most important. The following sections discuss phonetic studies that examine production and perception of tones as they interact with surrounding context and with the broader prosodic system of intonation and stress.

Tonal coarticulation

In the first comprehensive study of tonal coarticulation in Mandarin, Shen (1990) showed that coarticulation is bidirectional, and that coarticulation affects F0 height of the entire tone, rather than only onset and offset values. Most affected were tones following tone 2 and then tone 1, whose high offset F0 values had the effect of raising the onset F0 of the following tones; also affected was tone 4, whose high onset raised the offset of the preceding tone. Not affected were lexical tones preceding tones 2 and 3 whose onset falls in the middle of the frequency range. Shen's other findings were that tone contour was not affected by coarticulation, and that tonal coarticulation did not extend beyond one syllable.

Xu (1993, 1994) also finds evidence for both anticipatory and carryover coarticulation effects in tone production. Examining both production and perception of tones, Xu showed that coarticulation may alter the canonical F0 contour of the tone to the point that tones are not identifiable. However, when semantic information was removed from the tones through waveform editing, listeners compensated for the coarticulation as if the coarticulation effects were actually present in the signal, although these altered tones are unidentifiable when presented in isolation. While Xu's results confirm those from Shen (1990) regarding bidirectional coarticulation, he also finds that carryover effects are greater than anticipatory effects. Furthermore, his findings contradict Shen's in that he observes carryover effects not just in adjacent syllables, but extending beyond neighboring syllables into the second or third syllable as well. Another interesting result captured by Xu is that anticipatory coarticulation in Mandarin tones is dissimilatory for all but the low tone. This research reveals that variation

in tone production and perception is systematic, and that this variation may be utilized by listeners in tone identification.

Interactions of tones, stress, and intonation

Tones interact with other prosodic domains including stress and intonation. The interaction between tones and intonation has generated studies examining tone patterns in sentential environments. In general, the properties of tonal contour, relative duration, and peak amplitude seem to be maintained across different intonation types (Ho, 1976, 1977; Shen, 1990b). Indeed, Connell, Hogan, and Rozsypal (1983) showed that the perception of tones is stable across a variety of intonation types, a finding that made the authors conclude that this stability allows intonation to function in Mandarin.

As in many other languages, F₀, duration, and amplitude are acoustic correlates of stress in Mandarin (Krachochvil, 1969). Using discriminant analysis on a corpus of spontaneous speech produced by a single speaker, Coster and Krachochvil (1984) showed that the F₀ dimension most successfully corresponded with the actual tone and stress level identifications for the syllables. The amplitude parameter corresponded with the actual categories at better than chance level, while duration was not found to be successful at determining stress and tone categorization.

Moore (1993) compared F₀ and duration properties of stressed and unstressed tones with underlyingly atonic syllables. Atonic syllables consist of a small group of lexical items that include grammatical markers or lexicalized syllables (Chao, 1968; Duannu, 1990). Atonic syllables carry the neutral tone (tone 0) whose F₀ pattern is predictable depending on the preceding tone. The neutral tone is also known to be short in duration. Results showed that unstressed syllables remain distinct from toneless syllables in that their durations are longer, their canonical F₀ contours are not lost, and unlike toneless syllables, onset F₀ values were distinct from preceding tonal offsets. Stress was observed to modify the F₀ peak of tones, and relative durations between the tones were preserved in the unstressed tones as compared to the stressed tones. Thus, the phonological influence of stress alters acoustic characteristics to make some properties more prominent with respect to others in the phrase, but stress alone may not be responsible for unrecoverable changes to the intrinsic characteristics of the tone.

The importance of duration as a perceptual cue to stress was also examined by Shen (1993), who found that listeners identified tones with longer durations more often as stressed and concludes that duration, rather than F₀, is the primary acoustic correlate of stress in Mandarin tones.

The consensus of these studies is that stress affects both F₀ and duration properties of tones. F₀ highs become higher, and to a lesser extent, lows become

lower (Shen, 1990; Shih, 1988). Underlying tone contours are resilient to the influence of stress, however. Durations are longer for stressed tones; tones short enough in duration to make identification difficult may be perceived as completely unstressed and toneless (Lin, 1985). These general findings have contributed to an understanding of how acoustic parameters vary as a function of stress level. The intonational level only appears to impose constraints on F₀ while stress employs both F₀ and duration.

The phonological pattern of tone sandhi has also inspired several phonetic studies. Tone sandhi refers to the phonological phenomenon whereby a tone takes on features of a neighboring tone in a particular environment. An example that has gained widespread attention is Mandarin tone 3 sandhi, in which a tone 3 preceding another tone 3 becomes a tone 2. Wang and Li (1967; see also Peng, 2000) showed that listeners were not able to distinguish the tone 3 that had undergone tone sandhi from a tone 2.

The influence of speaker F₀ range

As suprasegmentals, tones are perceived relative to other tones, although they are also distinguished by intrinsic acoustic properties. For tones that contrast in both of these dimensions, intrinsic F₀ information may be sufficient for correct identification. To identify tones differing only in F₀ height, however, listeners must refer to their knowledge of the speaker's F₀ range, and where tones occur within that range. For example, a low tone produced by a high-pitched speaker and a high tone produced by a low-pitched speaker may be acoustically very similar. The process by which listeners adjust perception according to speaker-specific acoustic information is referred to as speaker normalization. While a few previous studies have investigated the role of extrinsic F₀ (e.g. Leather, 1983; Fox & Qi, 1990), Moore and Jongman (1997) have provided the most convincing evidence for speaker normalization in tone perception.

Moore and Jongman (1997) investigated speaker normalization in the perception of tones 2 and 3 by examining listeners' use of F₀ range as a cue to speaker identity. Two speakers were selected such that tone 2 of the low-pitched speaker and tone 3 of the high-pitched speaker occurred at equivalent F₀ heights. Three tone continua varying in either Turning Point, ΔF_0 , or both acoustic dimensions, were then appended to a natural precursor phrase from each of the two speakers. Results showed identification shifts such that identical stimuli were identified as low tones for the high precursor condition, but as high tones for the low precursor condition. Stimuli varying in Turning Point showed no significant shift, suggesting that listeners normalize only when the precursor varies in the same dimension as the stimuli. Overall, these results suggest that tone identification is influenced by changes in F₀ range, and that this information

is used as a frame of reference according to which ambiguous tones may be interpreted.

Hemispheric specialization for tone

Previous research indicates that the left hemisphere is more adept at phonemic processing, including phonemes, syllables, and words (Kimura, 1961; Shankweiler & Studdert-Kennedy, 1967; Studdert-Kennedy & Shankweiler, 1970) while the right hemisphere is better at melodic and prosodic processing, including music, pitch contours, and affective prosody (Kimura, 1964; Curry, 1967; Bryden, 1982). The processing of lexical tone therefore presents an interesting case. The fact that lexical tones are used to make phonemic contrasts would suggest that they are primarily processed by the left hemisphere. However, since lexical tones are instantiated by modulations of F0, one could instead expect the right hemisphere to be dominant in their processing.

Using a task in which listeners were asked to identify dichotically presented tone pairs by stating which tone they heard in which ear, Wang, Jongman, and Sereno (2001) showed that Mandarin tones are predominantly processed in the left hemisphere by native Mandarin speakers. Moreover, native speakers of American English with no prior experience with a tone language did not show a dominance of either hemisphere, suggesting that left hemisphere dominance for the Mandarin speakers arose from the intrinsic linguistic significance of the F0 modulations. These results are consistent with recent neuroimaging studies that show that prefrontal cortex in the right hemisphere is involved in pitch judgment tasks (Zatorre et al., 1992), whereas left hemisphere prefrontal cortex is involved in the processing of lexical tone in native speakers of tone languages when the presented tones have linguistic relevance (Gandour et al., 2000; Klein et al., 2001; Wang et al., 2003).

Summary and future directions

The research on production and perception of Mandarin tones reviewed in this chapter demonstrates that fundamental frequency, amplitude, and temporal properties such as overall duration and Turning Point are effective phonetic correlates of tone. Evidence was also presented to illustrate how tones interact with other suprasegmental processes such as intonation and stress. In particular, research suggests that the surface F0 pattern in sentences is driven by the tones. Stress and intonation may stretch, contract or shift F0 range, but the tone is left distinct.

These findings suggest that tones in Mandarin Chinese are affected by some of the same phonetic processes as segments. Coarticulation, intrinsic pitch, and context effects demonstrate that tones are sensitive to the surrounding phonetic

environment. A complex array of acoustic properties may be used to distinguish tones, as in segments.

One distinction between suprasegmentals and segments resides in their relationship to context information. Descriptions of segments may be stated in terms of formant frequencies in the case of vowels, for example, or manner and place of articulation cues for consonants. These descriptions only include properties that are intrinsic to the segment. Tone descriptions, on the other hand, must always include a relative property, such as "high" or "low." Thus, as these terms imply, tone identity is assumed to be intimately linked to the surrounding context (e.g. Lehiste, 1970). More research is required to understand exactly how these acoustic properties are affected by context.

The increased availability of neuroimaging techniques will result in a more detailed understanding of the processing of both segmental and suprasegmental information, including the extent to which speakers of tone languages process F0 fluctuations in a way that is fundamentally different from speakers of non-tonal languages.