Chinese Language Learning Sciences

Huei-Mei Liu
Feng-Ming Tsao
Ping Li *Editors*

Speech Perception, Production and Acquisition

Multidisciplinary approaches in Chinese languages

Springer

Chinese Language Learning Sciences

This book series investigates several critical issues embedded in fundamental, technical, and applied research in the field of Chinese as second language (CSL) learning and teaching, including learning mechanism in the brain, technology application for teaching, learning and assessment. The book series discusses these issues from the perspectives of science (evidence-based approach) and technology. The studies in the book series use the methods from the fields of linguistics (such as corpus linguistics and computational linguistics), psychological and behavioural sciences (such as experimental design and statistical analyses), informational technology (such as information retrieval and natural language processing) and brain sciences (such as neuroimaging and neurolinguistics). The book series generally covers three main interdisciplinary themes: (1) fundamental investigation of Chinese as a first or second language acquisition, (2) development in Chinese language learning technology, and (3) applied research on Chinese language education.

More specifically, the book series involves seven research topics:

- language transfer mechanism in Chinese as a second language
- factors of Chinese as a second language acquisition in childhood
- cultural influence on Chinese acquisition
- information technology, corpus
- teaching material design
- teaching strategies and teacher training
- learning models
- assessment methods

Please contact Melody Zhang (e-mail: melodymiao.zhang@springer.com) for submitting book proposals for this series.

More information about this series at http://www.springer.com/series/13176

Huei-Mei Liu · Feng-Ming Tsao ·
Ping Li
Editors

Speech Perception, Production and Acquisition

Multidisciplinary approaches in Chinese languages

Springer

Editors
Huei-Mei Liu
National Taiwan Normal University
Taipei, Taiwan

Feng-Ming Tsao
National Taiwan University
Taipei, Taiwan

Ping Li
The Hong Kong Polytechnic University
Hong Kong, Hong Kong

ISSN 2520-1719 ISSN 2520-1727 (electronic)
Chinese Language Learning Sciences
ISBN 978-981-15-7605-8 ISBN 978-981-15-7606-5 (eBook)
https://doi.org/10.1007/978-981-15-7606-5

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

A significant amount of research has gone in the last several decades into the study of speech learning and processing in Chinese, especially with regard to the processing of tones in Chinese (including Mandarin, Cantonese, Southern Min, and other major Chinese dialects). Against the backdrop of this research, we are pleased to present to our readers this volume as a synthesis and a roadmap to research in speech perception, production, and learning in the Chinese language context, with multidisciplinary research topics, approaches, and methodologies.

The idea of this book originated from the discussions in the context of the book series *Chinese Language Learning Sciences*, an ambitious Springer project initiated at the National Taiwan Normal University (NTNU). When the book series was first launched, one of the editors of this volume (PL) was on sabbatical leave visiting NTNU and had a chance to frequently discuss common interests with the other two editors (FT and HL) with regard to the topics of speech perception in children and adults, in native and non-native languages, and in typical and atypical language development. We felt that a book that can cover these grounds, with special reference to lexical tone perception in Chinese, would significantly help researchers, in particular, young scholars, and junior faculty members at institutions both in Asia and other parts of the world to become familiar with the field, and to carry out further exciting work on the basis of the extant literature. Thus, we have invited leading scholars from linguistics, psychology, cognitive neuroscience, and communication disorders to contribute to this volume. We specifically highlighted four major domains of work, including basic cognitive processes, neural representations, domain-general transfer and cross-modal integration, and development of speech from infancy to adulthood. Although the domains are not meant to be exhaustive of the large literature, we hope that this book represents some of the most exciting ongoing work and serves its purpose as both an overview of major research questions and a roadmap for future research.

We are grateful to the book series editor Prof. Yao-Ting Sung at NTNU for his constant encouragement and support, which makes this volume possible in the first place. We would also like to thank Lawrence Liu, Lay Peng Ang, and Melody Zhang at Springer for their support and patience during the editing of this book.

We thank the *Institute for Research Excellence in Learning Sciences*, NTNU, for their support. We also thank reviewers, Yang Zhang (University of Minnesota), Linjun Zhang (Beijing Language and Culture University), and Christina Zhao (University of Washington), for providing their suggestions that helped improve this book. Finally, the book would not have been possible without the hard work and strong commitment from the many contributing authors, who have gone through their manuscripts many times to ensure readability and accuracy. Needless to say, there may still be errors or areas for improvement, and we welcome readers from the community to provide critical comments and evaluation.

Taipei, Taiwan — Huei-Mei Liu
State College, USA — Ping Li
Taipei, Taiwan — Feng-Ming Tsao

Contents

Chapter 1
Introduction

Huei-Mei Liu, Feng-Ming Tsao, and Ping Li

This handbook brings together 12 chapters written by leading scholars in the field who provide perspectives and syntheses of important issues in speech processing and language learning in Chinese, with particular reference to lexical tones. In this book, the inter-disciplinary nature of the field is reflected in the diverse expertise of authors and the fields that they represent, including linguistics, psychology, cognitive neuroscience, education, and communication disorders. The research designs and methodologies used by researchers in this field are multidimensional, including but not limited to paradigms and methods from humanities, social sciences, computational science, neuroscience, and genetics, as clearly illustrated by the many studies either conducted or reviewed by the authors of this volume. In this introductory chapter, we discuss general issues addressed in different chapters, present the organization of this book, and provide an overview of each chapter. We hope that the readers will take our introduction as a starting point to read the individual chapters and then consider to examine the specific research issues further.

The organization of this volume is divided into four major areas or approaches. First, the most basic topics in this field of study are the phonological features and perceptual representations of speech sounds in Chinese, and therefore, the first part of this volume presents analyses of acoustics, perception, and production of lexical tones in native Chinese speakers and in adult learners of Chinese as a second language (CSL). Second, in addition to behavioral approaches, the rapid development and use of non-invasive neuroimaging tools to study processes of lexical tones

H.-M. Liu
National Taiwan Normal University, Taipei, Taiwan

F.-M. Tsao (✉)
National Taiwan University, Taipei, Taiwan
e-mail: tsaosph@mail2000.com.tw

P. Li
The Hong Kong Polytechnic University, Hong Kong, China

H.-M. Liu et al. (eds.), *Speech Perception, Production and Acquisition*,
Chinese Language Learning Sciences,
https://doi.org/10.1007/978-981-15-7606-5_1

not only help to reveal neurocognitive mechanisms but also provide convergent evidence that is consistent with results of behavioral approaches to support models of speech processing. Thus, the second part of this volume reviews studies on neural and cognitive representations of lexical tones in adults and children, using various neuroimaging methods, e.g., fMRI, EEG/ERP, and MEG. Third, speech sounds are acoustic signals of spoken words and are mainly processed through auditory systems; however, visual cues during speech sound production also contribute to the perception of speech sounds (cf. the McGurk effect of phonetic perception). Additionally, domain-general learning mechanisms facilitate phonetic learning. Cross-modal integration of speech perception and domain-general transfer are also evident with lexical tones. Given these effects in the literature, the third part of this volume demonstrates cross-modal integration and domain-general transfer of lexical tone processing, for example, with studies that examine the effects of visual cues of lexical tone production and music training effects on the perception of lexical tones. Finally, the scope of this volume would not be complete, if reviews on development of speech perception and production were not included. Studies of the production and perception in infants and children not only demonstrate developmental trends but also reveal underlying mechanisms that facilitate perceptual learning and production in native and non-native speakers of Mandarin Chinese. Thus, the final part of this volume covers speech perception and production development of Chinese languages from infancy through childhood, including studies exploring both monolingual and bilingual children, and also studies of typically developing children and children with communication disorders (e.g., autistic spectrum disorder).

Each chapter in this volume reviews a subfield of studies with research methods, empirical data, and theoretical explanations to address essential issues of Chinese speech processing. There are multiple research questions that span the chapters of this book. For example, effects of learning a native language on the processing of lexical tones are extensive, as illustrated in lifespan development of tone perception and production from infants (Chaps. 10, 11), children (Chaps. 6, 12, 13), and adult native speakers of tonal and non-tonal languages in native and second language learners (Chaps. 2–5, 7–9). Besides the general issue of language-learning experience and its impact, several chapters focus on issues that are particularly of relevance to lexical tones through acoustic/prosodic features of the tones. Compared with phonological features of non-tonal languages, tone sandhi is a special feature of tonal languages, and several chapters discuss this effect for which the phonological context affects the production of lexical tones (Chaps. 2, 7). Additionally, effects of tone sandhi on neural representation (Chap. 7) and word perception development in children (Chap. 11) are reviewed and syntheses provided. The acoustic/prosodic features of lexical tones (i.e., fundamental frequency and pitch) are also clearly topics of research interest and their impact on both music and linguistic prosody are examined in several chapters of the volume: for example, some studies have investigated the transfer effects from music to language-learning in infants (Chap. 10), non-tonal language speaking adults with amusia or with music training (Chap. 8), and the interaction between lexical tone and linguistic prosody in children with ASD (autistic spectrum disorder) who produce inappropriate intonations (Chap. 13).

These examples indicate that the chapters in this volume showcase a variety of research topics, approaches, and methodologies. The scope of this book goes beyond traditional studies of language teaching and learning or developmental studies of first language acquisition. It highlights learning, perception, and production of speech, from preverbal infants learning their first language to adults learning Chinese as a second language. While we recognize that speech perception in the Chinese context is a large and rapidly growing field and therefore it is impossible to represent all the exciting research in this volume, we believe that our handbook has provided a balanced view of the most important issues currently under investigation. We hope that readers will appreciate the significance of the presented issues in this volume for both theory building and applications in science, technology, education, and practice.

Below we provide a quick overview of the chapters and the organization of the volume.

1.1 Part I. Acoustics, Perception, and Production of Lexical Tones (in Adults)

Fon (Chap. 2) introduces the phonetic features of vowels, consonants, and lexical tones in Mandarin Chinese. In addition to depicting acoustic features of individual phonemes, Fon also presents major allophonic rules. Besides the common phonetic features in Mandarin speakers across many geographic regions, several phonetic features of Mandarin would interact with other languages, and this dynamic language interaction results in local variants of consonants, vowels, and lexical tones across different regions. Readers can take this chapter as a first step to understand the linguistic features for studying the mechanisms of speech perception and production.

Lee and Wiener (Chap. 3) review both top-down and bottom-up processes that are essential to lexical tone and spoken word perception. Regarding bottom-up processing, acoustic features (e.g., F0) of lexical tones vary with phonetic contexts and native Mandarin speakers perceive F0 contours of lexical tones by considering tonal coarticulation. Regarding top-down processing, Mandarin speakers track multiple knowledge-based information, e.g., from syllable-tone co-occurrence probabilities to tonal neighborhood density, syllable-tone lexical frequency, and lexical tone categorization. In sum, both acoustic-based (i.e., bottom-up) and knowledge-based (i.e., top-down) processes are used to achieve speech perception and accurate spoken word recognition, similarly for lexical tones and for phonetic segments.

There is an increasing interest in studying non-native in addition to native speakers' processing of lexical tone, as seen in this book. Ingvalson and Wong (Chapter 4) review tone training studies in second-language learners. Their chapter provides the evidence that sources of individual differences on tone learning could be the results of neurophysiological, neuroanatomic, or even genetic differences. They also present how the individual differences of tone training effect could be optimized by adjusting acoustic variations of tone stimuli to match learners' aptitude before training. Studies discussed in this chapter show that being exposed to lexical tones in single-speaker condition during tone training program is effective for low aptitude learners, but multiple-speaker condition is optimal for high aptitude learners. These studies point to the important new and promising directions, including research focused on second-language learners who vary in their aptitude and sensitivity to pitch acuity, to musical pitches (e.g., people with congenital amusia), and those who have lost their auditory sensitivity due to aging.

1.2 Part II: Neural Representations

Yu, Wang, and Li (Chap. 5) provide a review of the role of acoustic versus phonological information of lexical tones and summarize fMRI and ERP results that indicate issues related to the processing of tones by native Chinese speakers and second-language learners of Chinese. They also discuss the issue of brain lateralization: for example, native processing of the acoustic information would show right hemispheric lateralization, whereas their processing of the phonological information shows left hemispheric lateralization. By contrast, non-native processing patterns are inconsistent at first. Additionally, it is hypothesized that hemispheric lateralization of non-native processing of lexical tones would shift as second-language learners (L2) improve their L2 proficiency. The time course of processing is still an open issue: for example, whether acoustic information of lexical tones is processed earlier than phonological information in native speakers' processing of tones. Further, the functional connectivity between brain regions for specific acoustic cues (pitch height and pitch contour) in both native Mandarin speakers and non-native L2 learners requires further investigation.

Lee and Cheng (Chap. 6) present neurophysiological studies that use the event-related potentials (ERP) as measures of perceptual discrimination to depict the developmental trends of lexical tone perception from infancy through middle childhood. The presence of mismatched negativity (MMN), one of pre-attentive ERP components, indicates children's ability to discriminate between lexical tones. In early infancy, MMNs are observed when infants distinguish acoustically distinct tone contrasts, but positive mismatch response (P-MMR) is observed in the discrimination of acoustically similar contrasts. In addition to typically developing children, this chapter also reviews ERP studies of lexical tone perception in children experiencing language-learning difficulty, such as late-talking preschoolers or school-aged children with reading impairment, and the authors propose that several ERP components would be the neural markers to identify children at risk for later language-development impairment.

Chang and Kuo (Chap. 7) focus on the neural representations of perceiving and producing tone sandhi in tone-language speaking adults. As discussed earlier, the rules of tone sandhi represent context-dependent phonological rules for producing lexical tones in multi-syllabic words. The majority of tone sandhi studies in the past have explored the rule of Mandarin Tone 3 sandhi in disyllables, i.e., Tone 3 of the first syllable should be produced as Tone 2 (i.e., 33 → 23) when it is followed by a second syllable that also has Tone 3. The authors show that when producing Tone 3 sandhi, neuroimaging data revealed that the right posterior inferior frontal gyrus (pIFG) and its interaction with other areas in the right hemisphere are associated with the overt production of tone sandhi. Additionally, right IFG is also involved in the perception of lexical tones. Several issues need further investigation, including how the brain represents the effects of linguistic context (e.g., learning experience with a tone language) on the processing of Tone 3 sandhi.

1.3 Part III: Domain-General Transfer and Cross-Modal Integration

Ong, Tan, Chan, and Wong (Chap. 8) highlight the transfer from music training to lexical tone perception and review behavioral and neuroimaging studies that compare three groups of participants: listeners with extensive musical training, listeners with musical disorders such as amusia, and naïve listeners. Their review shows clear domain-general transfer that formal musical training benefits non-tonal language speakers in perceiving lexical tones, as well as producing better lexical tones. Tone-language speaking amusics tend to show impairment in musical pitch processing as well as in tone perception. The explanations of music-to-lexical tone transfer could rely on shared neural mechanisms for pitch processing in both lexical tone and music and also domain-general cognitive improvements (e.g., auditory memory). Future directions in the music-to-lexical tone transfer include setting criteria to clearly define musicians and patients with amusia in normative and neurocognitive studies.

Wang, Sereno, and Jongman (Chap. 9) survey the cross-modal integration effects of visual cues on lexical tone production and perception. When producing lexical tones, head, jaw, eyebrow and lip movements are aligned with spatial and temporal movement trajectories of specific tones. Perceptual findings show that facial and hand gestures improve tone intelligibility when they correspond to producing tones, and these benefits can be augmented by linguistic experience. In addition, greater visual benefits are found for contour tones (e.g., Tone 3 in Mandarin) as compared with flat tones (e.g., Tone 1 in Mandarin). Such findings suggest language-specific mechanisms in cross-modal integration of tone production and perception. Future studies should explore specific visual cues that can benefit the production of individual tones and evaluate how visual tonal cues facilitate perceptual categorization of lexical tones.

1.4 Part IV: Development from Infancy Through Childhood

Tsao and Liu (Chap. 10) provide an overview of studies on tone perception development in infants learning a tonal language (e.g., Mandarin and Cantonese) or a non-tonal language (e.g., English and Dutch). Infants learning a non-tonal language are able to discriminate tone contrasts of foreign languages at the age of 4–6 months, but they lose this discrimination ability when they reach the age of 9–12 months. The non-tonal language learners regain the sensitivity to distinguish tone contrasts at 18 months of age. Tone-language-learning infants improve their ability to discriminate native contrasts around the first birthday. Developmental factors for the perceptual development of lexical tones include listening to a tone language, the acoustic salience of lexical tones, statistical learning, music-to-tone transfer, and referential word learning. The authors point to the significance of assessing the effect of short-term music exposure on the perception of lexical tones, which could be one important future direction to examine in the field of cross-domain tone learning in infancy.

Expanding the study of speech perception development from lexical tone perception to word perception, Singh (Chap. 11) reviews studies on how lexical tones influence fundamental processes in the development of the mental lexicon, i.e., word segmentation, word recognition, and word learning, in both bilingual and monolingual learners of Mandarin. Studies reveal developmental differences between the course of tone acquisition and phonetic segments in early lexical processes. To fully account for early lexical development in childhood, the theoretical models need to take into consideration not only the influences of consonants and vowels in lexical processes, but also the influence of lexical tones on how children represent phonological details of novel words from infancy through early childhood. Further studies on lexical development in both monolingual and bilingual children are needed to provide empirical evidence to evaluate these models.

Peng and Chen (Chap. 12) summarize phonological development of consonants, vowels, and lexical tones in Mandarin-speaking children from infancy through middle childhood. In addition to depicting the developmental time course, the authors discuss various factors, including phonological markedness, role of different phonetic units within a given language, child-directed speech, and the articulation complexity in phonological theories to account for orders of development. They point to several issues that could be addressed in future studies, including the need to evaluate the effects of factors contributing to developmental order on individual phonological acquisition, as well as the importance to assess how links between perception and production facilitate phonological development in Mandarin-speaking children.

In addition to these chapters that focus on typical language development, in the final chapter, Yu and Shafer (Chap. 13) review behavioral and neurological evidence of prosody and lexical tone processing in Chinese-speaking adults and children with ASD. Chinese-speaking individuals with ASD produce atypical speech prosody and form less accurate perception of lexical tones. They also exhibit atypical auditory and language processes, e.g., enhanced pitch processing and impaired linguistic prosody processing. Yu and Shafter further report preliminary neural data on bilingual English–Mandarin-speaking children with ASD. This chapter shows that future directions on lexical tone development of Chinese-speaking children with ASD should explore the developmental trajectories with regard to various aspects of linguistic prosody, evaluating music-to-lexical tone transfer effect and designing language intervention programs that are based on neuroplasticity and theoretical models in the study of theory of mind, a major cognitive theory that accounts for communication impairments in ASD with respect to the lack of ability to reason about other people's beliefs, intentions, and emotions.

Part I
Acoustics, Perception, and Production of Lexical Tones (in Adults)

Chapter 2
The Phonetic Realizations of the Mandarin Phoneme Inventory: The Canonical and the Variants

Janice Fon

Abstract This chapter provides an overview on the phonetic realizations of the Mandarin phoneme inventory. There are two major sections. The first is a general description of the phoneme inventory, which includes five vowels and 19 consonants, along with four lexical tones. In addition to acoustic properties of individual phonemes, major allophonic rules are also discussed. The second section covers some variations on consonants, vowels, and tones in three major Mandarin varieties of Taiwan, Singapore, and China. Some variations are fairly region-specific, while others are more commonly found across various dialects. The former includes the retroflexed vowel suffix in the Mainland variety, the qualitative difference in realizing the neutral tone between the Taiwan and the Mainland variety, and the so-called fifth tone in Singapore Mandarin. The latter includes the deretroflexion and hypercorrection of sibilants, both of which can be found in Taiwan and Singapore Mandarin, and the syllable-final nasal mergers, which can be found in all three major dialects. Interestingly, these cross-dialectal variations also show large within-region variabilities. Both the canonical and the variant realizations of the phonological system are discussed in light of child language acquisition.

Mandarin is by far the most widely spoken language in the world. As of 2013, it boasts an estimated L1 population close to 900 million (Lewis, Simons, & Fennig, 2016), accounting for more than 12% of the total world population at the time (cf. Population Reference Bureau, 2013). It is the sole official language of Taiwan and China, and is one of the four official languages in Singapore. This chapter contains two main sections. The first is devoted to an overview of the phonetic realizations of Mandarin phonology, and the second focuses on some variations in the three major Mandarin variants of Taiwan, Singapore, and China. All three main aspects of phonology, i.e., vowels, consonants, and tones, are discussed.

J. Fon (✉)
Graduate Institute of Linguistics, National Taiwan University, Taipei, Taiwan
e-mail: jfon@ntu.edu.tw

H.-M. Liu et al. (eds.), *Speech Perception, Production and Acquisition*, Chinese Language Learning Sciences,
https://doi.org/10.1007/978-981-15-7606-5_2

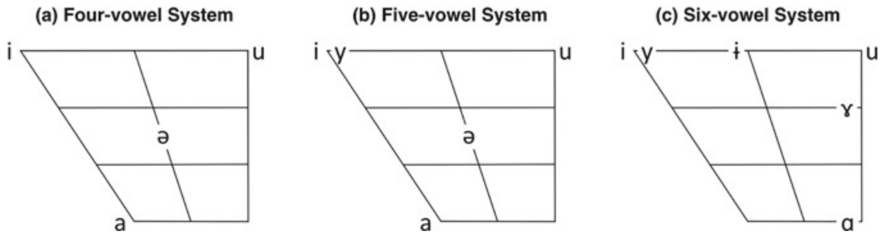

Fig. 2.1 Three different proposals for the Mandarin vowel system

2.1 The Phoneme Inventory

2.1.1 Vowels

Depending on the theoretical frameworks, Mandarin is said to have a system ranging from four to six single vowels. The four-vowel system is based on feature geometry and was proposed by Wu (1994) (Fig. 2.1a), while the six-vowel system is based on phonemic evidence and was proposed by C.-C. Cheng (1973) (Fig. 2.1c). Most other scholars argued for a five-vowel system, with an inventory of three high vowels, one mid-vowel, and one low vowel (Chao, 1968; R. L. Cheng, 1966; Duanmu, 2007; Y.-H. Lin, 1989; Wan & Jaeger, 2003), as shown in Fig. 2.1b. As all three proposals include largely overlapping sets of vowels, this chapter follows the perspective of the majority and reviews only the five-vowel system in detail.

The three high vowels in Mandarin could be nicely demonstrated by a minimal triplet, as shown in (1).[1] These three phonemes behave in a rather similar fashion and become a homorganic glide when they precede a non-high vowel, as shown in (2). [i] and [u] can also be the second part of a diphthong and act as an off-glide, as shown in (3).

(1)	*í* /i/	*yú* /y/	*wú* /u/
	‘aunt’	‘fish’	‘nil’

(2)	*yăn* /ian/→[jan]	*wăn* /uan/→[wan]	*yuăn* /yan/→[ɥan]
	‘to perform’	‘bowl’	‘faraway’

(3)	*năi* /nai/→[naɪ]	*năo* /nau/→[naʊ]
	‘milk’	‘brain’

[1] *Hanyu Pinyin* is adopted for Mandarin romanization throughout this chapter.

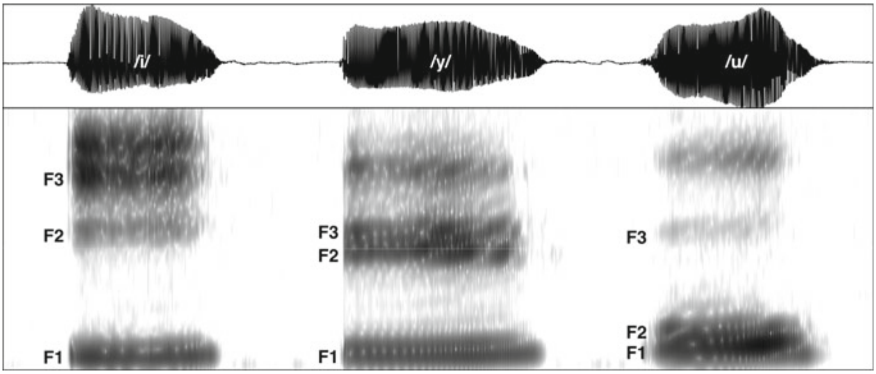

Fig. 2.2 Waveforms and spectrograms of the three high vowels, *yí* /i/ 'aunt', *yú* /y/ 'fish', and *wú* /u/ 'nil'. The first three formants are labeled accordingly

Figure 2.2 shows the waveforms and spectrograms of the three high vowels, *yí* /i/ 'aunt', *yú* /y/ 'fish', and *wú* /u/ 'nil', produced by a female speaker. Notice that the acoustic cues correspond nicely with the articulatory descriptions of the vowels. All three vowels are high, and therefore their F1s are unanimously low. /i/ and /y/ are both front vowels, while /u/ is a back vowel, and therefore the former two have higher F2s than the latter. /y/ and /u/ are both rounded vowels and therefore have relatively low F3s (and F2s), as compared to the unrounded /i/.

The high vowel /i/ is worth further mention. Besides being realized as a high front vowel and a glide, it also has a third allophone [ɨ] that occurs after dental and retroflex sibilants [ts ts^{h} s tʂ^{h} ʂ^{h} ʐ] (see below) (C.-C. Cheng, 1973). (4) contains a near-minimal triplet of [i] and [ɨ],[2] and their waveforms and the spectrograms are shown in Fig. 2.3. Compared to their counterpart [i] in *tì* [ti] 'ground', the F2s in *zì* [tsɨ] 'Chinese character' and *zhì* [tʂɨ] 'mole (face)' are much lower, indicating backing of the tongue position.

(4)	*tì* /ti/→[ti]	*zì* /tsi/→[tsɨ]	*zhì* /tʂi/→[tʂɨ]
	'ground'	'Chinese character'	'mole (face)'

[2]The phone has been problematic in Mandarin phonology and has invited much discussion and debate. Classical views usually argued for two phones instead of one. For example, Norman (1988) proposed two apical vowels of [ɿ] and [ʅ], Lee-Kim (2014) argued for two syllabic approximates of [ɹ̩] and [ɻ̩], and Duanmu (2007) suggested two syllabic fricatives of [z̩] and [ʐ̩]. The former ones of each pair are designated to occur after dental sibilants, while the latter ones are designated to occur after retroflexes. However, since the formants in Fig. 2.3 show little frication, implying that it is at least not always realized as a syllabic fricative, and since the major difference between the two renditions of the phone (*zì* [tsɨ] 'Chinese character' vs. *zhì* [tʂɨ] 'mole (face)') mainly lies in F3, not F2, implying that there is little change in tongue backness, this chapter adopts a unifying symbol [ɨ] instead, following C.-C. Cheng (1973), and views the phonetic variation as a product of coarticulation. There is a slight departure from C.-C. Cheng's (1973) original proposal, however, as this chapter argues for a nonphonemic status instead (see Fig. 2.1).

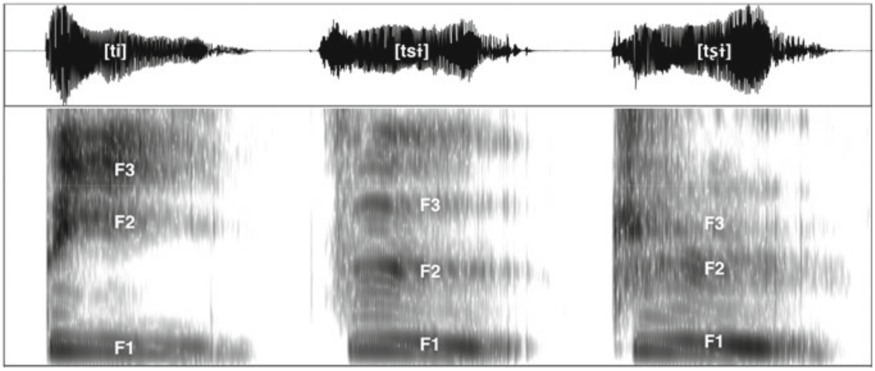

Fig. 2.3 Waveforms and spectrograms comparing the two allophonic variants of /i/ in *tì* [ti] 'ground', *zì* [tsɨ] 'Chinese character', and *zhì* [tʂɨ] 'mole (face)'. The first three formants are labeled in white

Mandarin only has one mid-vowel, but its frontness and rounding show environmentally dependent allophonic changes while its height stays the same. Most previous works agree that there are at least four variants, [ə], [e], [o], and [ɤ] (C.-C. Cheng, 1973; R. L. Cheng, 1966; Y.-H. Lin, 1989). [ə] occurs in CVN syllables and is considered the default (R. L. Cheng, 1966; Y.-H. Lin, 1989; Wan & Jaeger, 2003; Wu, 1994). [e] occurs in CjV and CɥV syllables, while [o] occurs in CwV syllables. Finally, [ɤ] occurs in CV syllables. (5) shows a set of examples.

(5)	*lèng* /ləŋ/→[ləŋ]	*liè* /liə/→[lje]	*luò* /luə/→[lwo]	*lè* /lə/→[lɤ]
	'absent-minded'	'to crack'	'to fall'	'happy'

Figure 2.4 shows the waveforms and spectrograms of the four allophonic variants of the mid-vowel /ə/ in *lèng* [ləŋ] 'absent-minded', *liè* [lje] 'to crack', *luò* [lwo]'

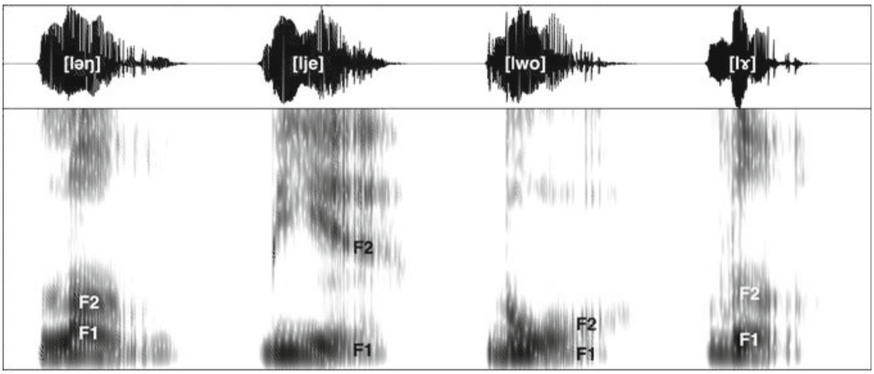

Fig. 2.4 Waveforms and spectrograms of the four allophonic variants of the mid-vowel /ə/ in *lèng* [ləŋ] 'absent-minded', *liè* [lje] 'to crack', *luò* [lwo] 'to fall', and *lè* [lɤ] 'happy'. The first two formants of the main vowels are labeled in black or white accordingly

to fall', and *lè* [lɤ] 'happy'. Notice that both [e] and [ɤ] are mid-high vowels and therefore have similar F1s. [e] has a higher F2 than [ɤ] because it is a front vowel. [o] and [ɤ] have similar F1s and F2s because they are both mid-high back vowels, but the former has a slightly lower F2 due to rounding. Compared with [e] and [o], [ə] has a higher F1, showing that it is relatively low in height, and its F2 value is intermediate between that of [e] and [o], showing that it is in the central position.

Similarly, Mandarin also only has one low vowel, and its frontness and height are context-dependent and show allophonic changes. However, its unrounded quality stays the same. Most studies agree that there are at least three allophonic variants, [a], [ɑ], and [ɛ] (C.-C. Cheng, 1973; R. L. Cheng, 1966; Y.-H. Lin, 1989; Wu, 1994). [a] is default and occurs in open and C(w)Vn syllables (R. L. Cheng, 1966; Y.-H. Lin, 1989; Wan & Jaeger, 2003). [ɑ] occurs in CVu and C(G)Vŋ, and [ɛ] occurs in CjVn and CɥVn syllables. (6) is a set of examples.

(6)	*án* /an/→[an]	*áng* /aŋ/→[ɑŋ]	*ián* /ian/→[jɛn]
	'1.SG (regional)'[4]	'to raise high'	'salt'

[4] 1.sg: first person singular pronoun

Figure 2.5 shows the waveforms and spectrograms of the three allophonic variants of the low vowel /a/ in *án* [an] '1.SG', *áng* [ɑŋ] 'to raise high', and *ián* [jɛn] 'salt'. Notice that both [a] and [ɛ] are front vowels and therefore have higher F2s than the back vowel [ɑ]. [ɛ] also has a lower F1 than [a] and [ɑ], indicating its higher vowel height.

There is an additional retroflex vowel /ɚ/ in Mandarin, which can only occur by itself in a bare V syllable (Duanmu, 2007). (7) shows a minimal pair of /ɚ/ versus /ə/. Figure 2.6 shows a spectrographic comparison between the two vowels. Notice that because of the rhotic quality of the vowel /ɚ/, the third formant is drastically lowered as compared to that of the mid-vowel /ə/.

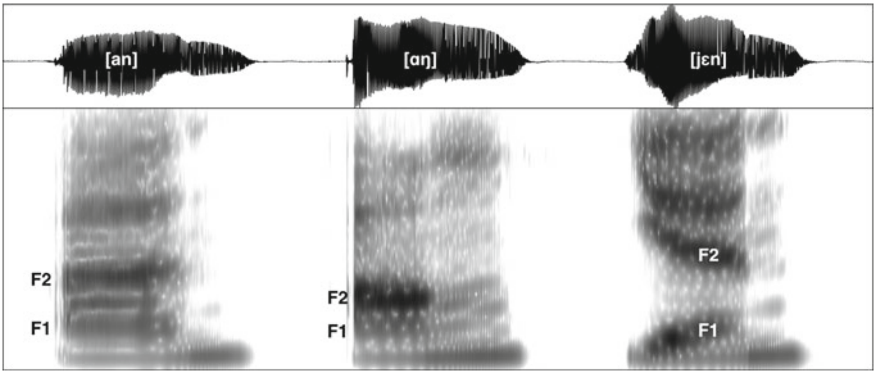

Fig. 2.5 Waveforms and spectrograms showing the three allophonic variants of the low vowel /a/, *án* [an] '1.SG', *áng* [ɑŋ] 'to raise high', and *ián* [jɛn] 'salt'. The first two formants of the main vowels are labeled in black or white accordingly

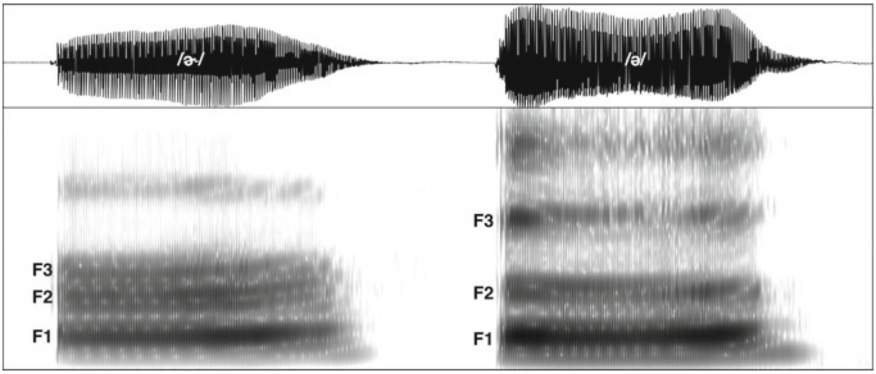

Fig. 2.6 Waveforms and spectrograms of the retroflex vowel *ér* /ɚ/ 'son' and the mid-vowel é /ə/ 'goose'. The first three formants of the main vowels are labeled accordingly

Table 2.1 Mandarin consonant chart

	Labial		Dental		Retroflex		Velar	
Stop	p	p^h	t	t^h			k	k^h
Affricate			ts	ts^h	tʂ	tʂ^h		
Fricative	f		s		ʂ	ʐ	x	
Nasal	m		n				ŋ	
Lateral			l					

(7) *ér* /ɚ/ *é* /ə/
'son' 'goose'

2.1.2 *Consonants*

Mandarin has 19 consonants, including 6 stops, 4 affricates, 5 fricatives, 3 nasals, and 1 lateral (Chao, 1968; R. L. Cheng, 1966; Duanmu, 2007) (Table 2.1).[3] The stops occupy three places of articulation, labial, dental, and velar, and have both the aspirated and the unaspirated series. (8) shows a minimal sextuple. The waveforms and the spectrograms of the sextuple are shown in Fig. 2.7.

[3]Some researchers contended that Mandarin consonant inventory also includes an additional set of three alveolo-palatal sibilants /tɕ tɕ^h ɕ/ [e.g., Luo (1993)]. However, since these three are in complementary distribution with the velar series /k k^h x/, the dental series /ts ts^h s/, and the retroflexes /tʂ tʂ^h ʂ/, this chapter views them as allophones of other phonemes and discusses them in a later section.

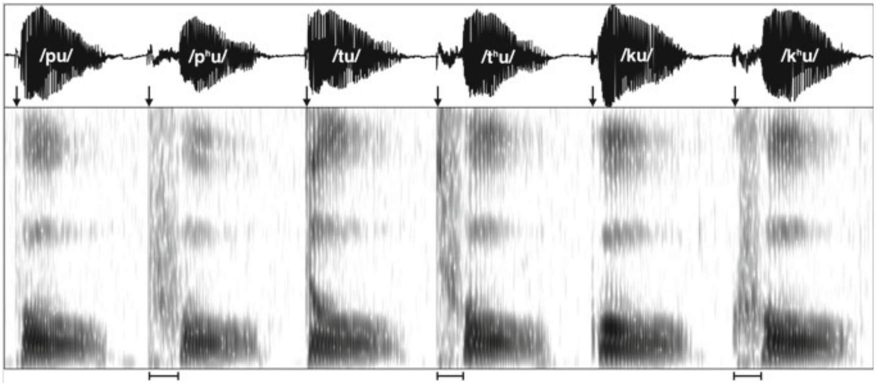

Fig. 2.7 Waveforms and spectrograms of the six stops, *bǔ* /pu/ 'to mend', *pǔ* /p^{h}u/ 'sheet music', *dǔ* /tu/ 'to gamble', *tǔ* /t^{h}u/ 'soil', *gǔ* /ku/ 'ancient', and *kǔ* /k^{h}u/ 'bitter'. The downward arrows above the spectrograms indicate the stop bursts, and the horizontal line segments beneath the spectrograms indicate the aspiration noise of aspirated stops

(8)	*bǔ* /pu/	*pǔ* /p^{h}u/
	'to mend'	'sheet music'
	dǔ /tu/	*tǔ* /t^{h}u/
	'to gamble'	'soil'
	gǔ /ku/	*kǔ* /k^{h}u/
	'ancient'	'bitter'

There are two sets of affricates in Mandarin, the dental and the retroflex. Like their stop counterparts, both the aspirated and the unaspirated series are included. (9) shows a minimal quadruple. The waveforms and the spectrograms of the quadruple are shown in Fig. 2.8. Notice that the lowest major energy concentration for the frication portion of the affricates is lower for the retroflexes than that for the dentals.

(9)	*zàn* /tsan/	*càn* /ts^{h}an/
	'to praise	'bright'
	zhàn /tʂan/	*chàn* /tʂ^{h}an/
	'to stand'	'to shiver'

Fricatives are the largest set of inventory in Mandarin. In total, there are three sibilants and two nonsibilants. Of the three sibilants, there is one voiceless dental /s/, one voiceless retroflex /ʂ/, and one voiced retroflex /ʐ/. (10) shows a minimal triplet. The waveforms and the spectrograms of the triplet are shown in Fig. 2.9. Like the affricates, the retroflexes have lower major energy concentration than the dental.

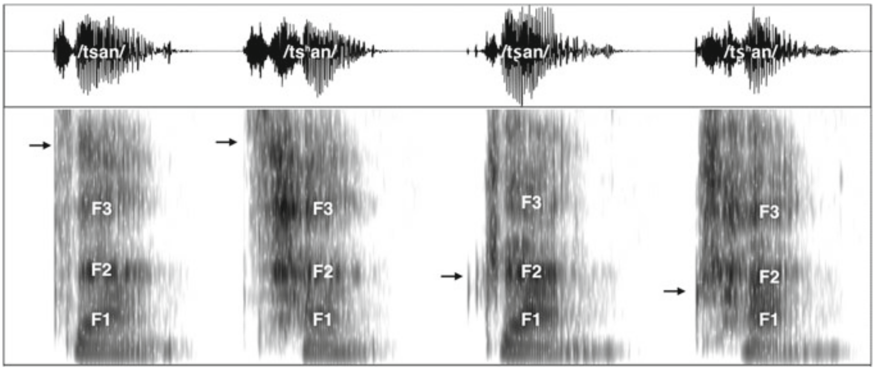

Fig. 2.8 Waveforms and spectrograms of the four affricates, *zàn* /tsan/ 'to praise', *càn* /tsʰan/ 'bright', *zhàn* /tʂan/ 'to stand', and *chàn* /tʂʰan/ 'to shiver'. The rightward arrows indicate the lowest major energy concentration for the noise burst of the affricates. The first three formants of the main vowels are labeled in white accordingly

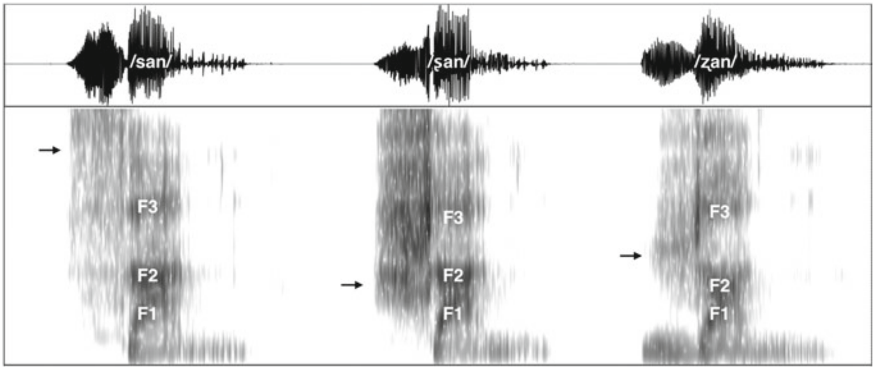

Fig. 2.9 Waveforms and spectrograms of the three Mandarin sibilant fricatives, *sǎn* /san/ 'umbrella', *ǎn* /ʂan/ 'to flash', and *rǎn* /ʐan/ 'to dye'. The rightward arrows indicate the lowest major energy concentration for the fricative noise. The first three formants of the main vowels are labeled in white accordingly

(10) *sǎn* /san/ *shǎn* /ʂan/ *rǎn* /ʐan/

'umbrella' 'to flash' 'to dye'

The two nonsibilant fricatives are labial /f/ and velar /x/. (11) shows a minimal pair. The waveforms and the spectrograms of the pair are shown in Fig. 2.10.

(11) *fú* /fu/ *hú* /xu/

'blessing' 'lake'

Mandarin has three nasals, /m n ŋ/, and one liquid /l/. /m n l/ can occur in the onset position, as indicated in (12). The waveforms and the spectrograms of the triplet are shown in Fig. 2.11. /n ŋ/ can occur in the coda position, as shown in (13).

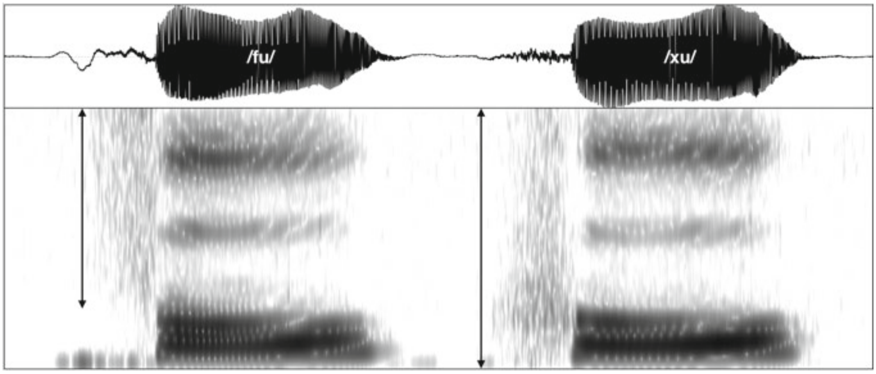

Fig. 2.10 Waveforms and spectrograms of the two nonsibilant fricatives, *fú* /fu/ 'blessing' and *hú* /xu/ 'lake'. The up–down arrows indicate the range of frequency noise for the two fricatives

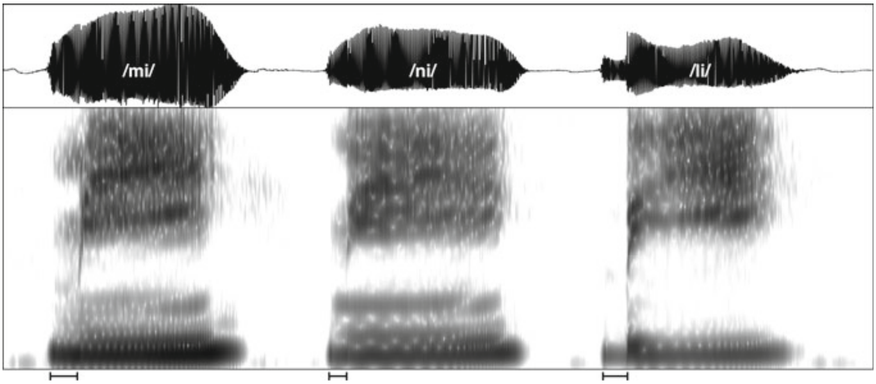

Fig. 2.11 Waveforms and spectrograms of the three onset sonorants, *mí* /mi/ 'mystery', *ní* /ni/ 'mud', and *lí* /li/ 'pear'. The horizontal line segments beneath the spectrograms indicate the sonorant sections

As illustrated in Fig. 2.12, the two nasals differ not only in nasal formants, but also in the F2 and F3 offsets of the preceding vowels. /n/ has both a falling F2 and a falling F3, while /ŋ/ has a rising F2 and a falling F3.

(12)	*mí* /mi/	*ní* /ni/	*lí* /li/
	'mystery'	'mud'	'pear'
(13)	*pín* /pʰin/	*píng* /pʰiŋ/	
	'poverty'	'level'	

There is also a special set of alveolo-palatal sibilants [tɕ tɕʰ ɕ], which are in complementary distribution with the velars /k kʰ x/, the dentals /ts tsʰ s/, and the retroflexes /tʂ tʂʰ ʂ/. The alveolo-palatal set only occurs before /i y/, while the other

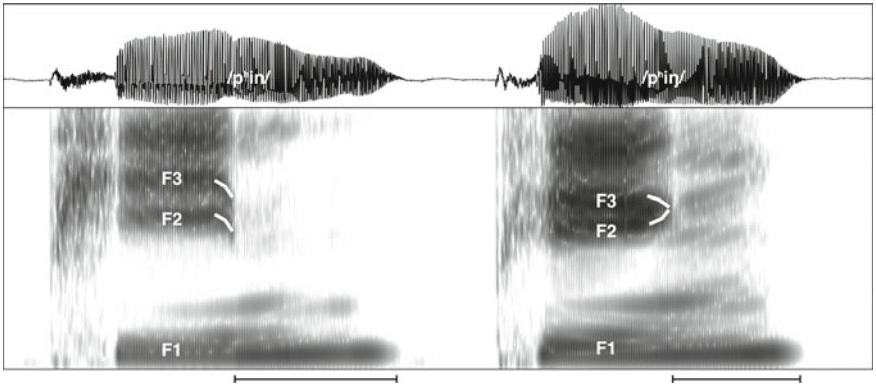

Fig. 2.12 Waveforms and spectrograms of the two nasal codas, *pín* /pʰin/ 'poverty' and *píng* /pʰiŋ/ 'level'. The first three formants of the main vowels are labeled in white. The F2 and F3 offset tracings of the main vowels are also shown in white. The horizontal line segments beneath the spectrograms indicate the nasal sections

three only occur elsewhere, as shown in (14). Diachronically, the alveolo-palatal set stems from two historical sources, the velars and the dentals. Synchronically, some scholars have identified it with the velar set [e.g., Chao (1968), R. L. Cheng (1966)], while others have identified it with the dental set [e.g., Hartman (1944)]. In reality, the evidence could go either way (Duanmu, 2007; Y.-H. Lin, 1989). Figure 2.13 shows the waveforms and the spectrograms of an alveolo-palatal triplet.

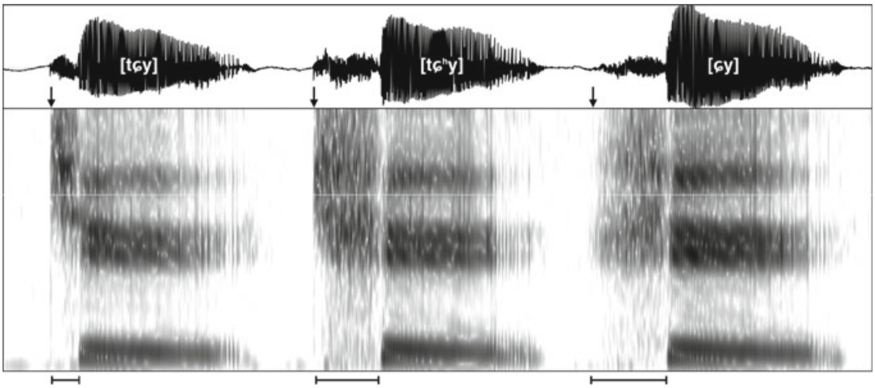

Fig. 2.13 Waveforms and spectrograms of the three alveolo-palatals, *jǔ* [tɕy] 'to lift', *qǔ* [tɕʰy] 'to take', and *xǔ* [ɕy] 'to promise'. The downward arrows above the spectrograms indicate the stop bursts, and the horizontal line segments beneath the spectrograms indicate frication noise

(14)	*jǐ* [tɕi]	*qǐ* [tɕʰi]	*xǐ* [ɕi]
	‘to squeeze’	‘to rise’	‘to wash’
	jǔ [tɕy]	*qǔ* [tɕʰy]	*xǔ* [ɕy]
	‘to lift’	‘to take’	‘to promise’
	*/ki-/ */ky-/ */tsi-/ */tsy-/ */tʂi-/ */tʂy-/	*/kʰi-/ */kʰy-/ */tsʰi-/ */tsʰy-/ */tʂʰi-/ */tʂʰy-/	*/xi-/ */xy-/ */si-/ */sy-/ */ʂi-/ */ʂy-/

2.1.3 Tones

There are four citation tones in Mandarin, traditionally characterized as Tone 1, Tone 2, Tone 3, and Tone 4. Tone 1 is a high-level tone, Tone 2 is a rising tone, Tone 3 is a dipping (i.e., falling–rising) tone, and Tone 4 is a falling tone (Chao, 1968). (15) shows an example of a minimal quartet, using a fairly common transcription convention of suffixing the tonal number to the IPA transcription of the syllable (Duanmu, 2007). Figure 2.14 shows an example of the F0 tracks of the four tones. Although tones are realized by continuous pitch contours, different sections of the contours seem to have different weightings for different tones. Lee & Wiener (Chap. 3 of this volume) and Tsai & Liu (Chap. 10 of this volume) give a good overview of

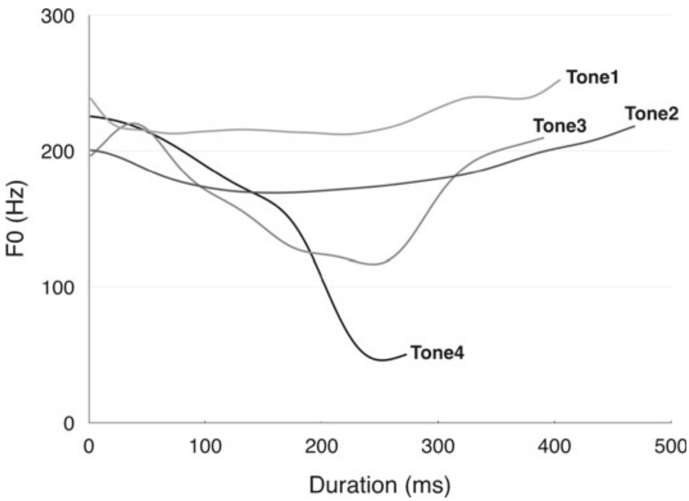

Fig. 2.14 F0 tracks of the four tones in Mandarin on the syllable *ying* /iŋ/