

”LEARNING FOREIGN LANGUAGES”

PHONETIC ANALYSES OF A TIĀNJĪN XIÀNGSHENG

Martin Overå Johnsen



M. A. thesis in East Asian Linguistics

Department of Culture Studies and Oriental Languages

Faculty of Humanities

UNIVERSITY OF OSLO

1 December, 2010

Table of contents

Preface	4
Notational Conventions	5
1 Introduction	7
1.1 Material.....	8
1.2 Phonetics.....	11
1.3 Phonology	13
1.4 Structure and presentation	14
2 Phonetic Theory	16
2.1 Articulatory phonetics	16
2.2 Acoustic phonetics.....	17
2.3 Consonants.....	19
2.4 Vowels	21
2.5 Intonation, tone and stress	24
3 An Overview of Mandarin Phonology	27
3.1 Segmental structure	27
3.2 Tones.....	31
3.3 Stress.....	34
4 Methods and Analytical Categories	36
4.1 Consonants.....	36
4.2 Vowels	39
4.3 Tone, stress and intonation	43
5 The Phonetic Basis of Syllable Stress	50
5.1 Theoretical background	51
5.2 Examining the data	52
5.3 Conclusions.....	59

6	Some Thoughts on Tone 4	61
6.1	Theoretical background	61
6.2	Examining the data	64
6.3	Conclusions.....	68
7	Tone 3 and Tone 3 Sandhi	70
7.1	Theoretical background	70
7.2	Examining the data	72
7.3	Conclusions.....	77
8	Phonetic Transcription	79
9	Concluding Remarks	129
	References	130

Preface

I would like to thank my academic supervisor professor Christoph Harbsmeier for helping me through this project. If not for him, this thesis would probably never have seen the light of day. My thanks also go out to Helga Eggebø and Berit Johannesen for commenting on my work as I went along, and to Bao Yue and Guan Hongbo for helping me decipher the most exotic Chinese colloquialisms in the material I have investigated.

Martin Overå Johnsen

30 November 2010

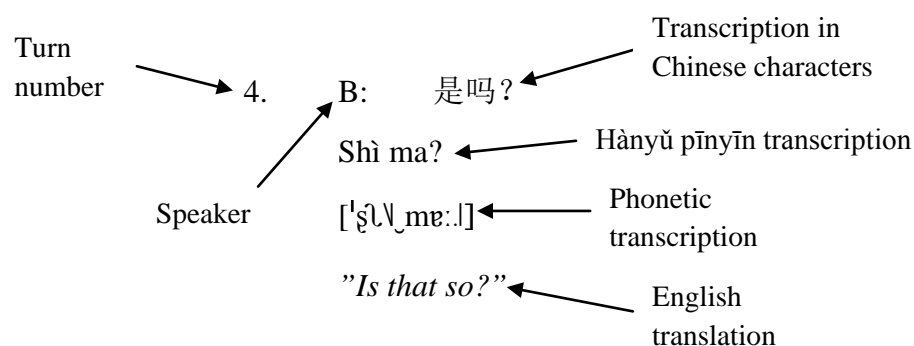
Notational conventions

Phonological notation	Definition
[xxx]	Phonetic transcription
/xxx/	Phonological transcription
/__xx/	Phonological environment: The underscore is used to make remarks about the environments in which certain speech sounds can or cannot appear. For example, the observation that “the only vowel that can appear between /x/ and /n/ is /i/” can be stated as: <div style="text-align: center;"> $\begin{array}{c} i \\ /x_n/ \end{array}$ </div>
→	Realized as: Indicates how a phonological form is pronounced. Often used in combination with slanting brackets and square brackets. For example, the observation that “the phoneme /j/ is pronounced as either [tɕ] or [dʒ] can be represented as: <div style="text-align: center;"> $\begin{array}{ccc} /j/ & \rightarrow & [tɕ] \\ & & [dʒ] \end{array}$ </div>
xxx xxx	Phrase boundary
C	Consonant
V	Vowel
¹ ma	Primary stress: the most strongly stressed syllable in a phrase
₁ ma	Secondary stress: any stressed syllable that is not the most strongly stressed in a phrase.
Abbreviations	Definition
NP	Noun phrase
VP	Verb phrase

Mod-N	A noun preceded by a modifier
Mod-V	A verb preceded by a modifier
OED	Oxford English Dictionary
XHC	Xiàndài Hànyǔ Cídiǎn
IPA	International Phonetic Association

Transcription of Xué Wàiyǔ

The transcription of Xué Wàiyǔ is given in four tiers:



When one particular phrase, word or syllable from the data set is referred to in discussions I always follow it with a parenthesis containing the number of the turn where it occurs. This will allow readers to easily look up that phrase, word or syllable in the phonetic transcription. For example, if I refer to “zàijiàn (117)” in the text, this signifies that I am referring especially to the word zàijiàn that is uttered by speaker A in turn 117 of Xué Wàiyǔ.

1 Introduction

Mandarin Chinese is the Lingua Franca of the People's Republic of China; the world's most populous nation, and also one of the most powerful. The last few years have seen a great increase in the demand for Europeans and Americans with proficiency in this language.

Unfortunately, a European or American student trying to learn to speak Mandarin Chinese is faced with considerable challenges. Mandarin is a tone language, which means that a phrase like “pingguo” can mean things as different as “apple” and “frying pan” depending on what kind of pitch accent you say it with. Mandarin also contains several consonants and vowels that are not only non-existent in languages like English, but that are hardly found in any other languages at all. And as if that weren't enough, the language has a rhythm and flow which is somehow very difficult to grasp for European speakers.

How, exactly, is Mandarin pronounced by native Chinese then? This thesis attempts in part to answer that question by looking at the pronunciation of two particular Chinese speakers, in one particular setting. Specifically, the aim of this thesis is to develop and present a phonetic transcription of a Mandarin comedic dialogue called a “xiàngsheng”.

The most common way of representing Chinese speech sounds in writing, and also the chief tool for teaching Chinese pronunciation to foreigners, is China's official alphabetical writing system; Hànyǔ Pīnyīn (commonly referred to as just Pīnyīn). Most modern textbooks use this alphabet to give students an introduction to the system of Mandarin speech sounds. The major dictionaries of modern Mandarin, the Xiàndài Hànyǔ Cídiǎn 现代汉语词典 (2007), Hànyǔ Dà cídiǎn 汉语大词典 (2000), Hànyǔ Dà Zìdiǎn 汉语大字典 (2001), Xīnhuá Dà Zìdiǎn 新华字典 (Zhou, 2004) and Cíhǎi 辞海 (1999), also use Pīnyīn to indicate the pronunciation of entry words. And yet, there is often a considerable distance between the way words are written in Pīnyīn and the way they are pronounced. The advantages of Pīnyīn as an orthographic writing system are great; it provides a simple and efficient way of typing Chinese on a computer, and has also made it possible to arrange words alphabetically in Chinese dictionaries. However, such uses require Pīnyīn to be as simple as possible, which makes it unsuited to representing the finer points of Mandarin pronunciation.

Every language has its own unique system of speech sounds. The greatest shortcoming of Pīnyīn as a phonetic writing system is a direct result of the way the Mandarin sound-system functions. The system of Mandarin speech sounds is very simple and predictable on the level

of each individual syllable, but becomes much more complex and irregular when the syllables are put together to form word and phrases. When two syllables are pronounced together, the pronunciation of each syllable may change considerably, while its meaning stays unchanged. The simplest and most practical spelling system for Mandarin Chinese is therefore one which has a fixed spelling for each syllable, and ignores the changes that occur when the syllables form larger units like words.

There are writing systems that are developed especially to give as accurate an account of speech sounds as possible. The most common of these is the International Phonetic Alphabet (IPA, 2005). Because writing systems like Pīnyīn benefit from having as few symbols as possible, they are made in such a way that each symbol implicitly reflects as much information as possible without leading to ambiguities. With systems like the Phonetic Alphabet the goal is often exactly the opposite; to represent as much information as possible with separate symbols. In other words, to be as explicit as possible about each detail of the pronunciation (Laver, 1997: 559). Unfortunately, phonetic transcriptions that represent this level of detail have the disadvantage of being somewhat difficult to read.

Using the symbols and conventions of the International Phonetic Alphabet, I aim to produce a transcription of a Mandarin dialogue that is more detailed than ordinary Pīnyīn writing. My transcription should therefore reveal what happens to the pronunciation of syllables when they are strung together to form words and phrases. And yet, it should be systematic and simple enough that it could be of practical use to advanced students of Mandarin. The central questions of my thesis could therefore be summarized as:

- How can I best represent the speech sounds of a Mandarin xiàngsheng through a phonetic transcription?
- What can this transcription reveal about the way Mandarin speech sounds behave when connected to form words and phrases?

1.1 Sources and material

The material I have chosen to transcribe phonetically is a so-called xiàngsheng 相声.

Xiàngsheng, often translated to English as “crosstalk”¹, is a traditional Chinese performing art with its roots in Běijīng. Xiàngsheng is comedic genre, whose primary goal is to be funny

¹ This translation, although it is by far the most widely used, is inaccurate. The term xiàngsheng actually means “resembling/mimicking sounds” (Kaikkonen, 1990: 14).

and entertaining. Each piece is relatively short, usually under 30 minutes, and can feature one, two or three actors. The number of actors participating determines what style the piece is classified as; pieces with one actor are called “single xiàngsheng”, pieces with two actors are called “double xiàngsheng”, and if there are three actors the piece is a “group xiàngsheng”.² Before 1949 these three forms were all common, but now the double xiàngsheng is far more popular than the others (Kaikkonen, 1990: 8-12).

The xiàngsheng that I will analyze here is a piece called Xué Wàiyǔ 学外语 “Learning Foreign Languages”. It was chosen because of its convenient length, and because its relatively simple and ordered structure fits the limited scope of my study. Xué Wàiyǔ is a double xiàngsheng lasting just over ten minutes, and it is performed by Mǎ Sānlì 马三立 and Zhāng Qìngsēn 张庆森. It is available as an mp3 download from www.tingcd.net (2010), and a cartoon video version of the same recording can be viewed on www.youtube.com (2009). Similar cartoon versions of the recording can be found on numerous Chinese language websites, but most have very poor sound quality.

Mǎ Sānlì is one of the most revered performers in the modern history of xiàngsheng. He is especially known for his ability to make his xiàngsheng appear natural and unrestrained, and thereby create a unique rapport with his audience. The son and grandson of famous xiàngsheng performers, Mǎ Sānlì was born in Běijīng in 1914 (Xue, 1985: 76). At the age of three, he moved to Tiānjīn, where he was to go to elementary and middle school as well as study xiàngsheng with the great master Zhōu Déshān. In 1926, when he was 12 years old, he dropped out of school and began his profession as a xiàngsheng performer in Tiānjīn. When in Běijīng in 1948, Mǎ Sānlì met Zhāng Qìngsēn, and the two began performing together (Ma, 2003). Zhāng Qìngsēn was born in 1912, two years earlier than his partner, and his family is of the Manchu nationality (Bǎidù Bǎikē, 2010). I have not been able to learn exactly when the two friends’ performance of Xué Wàiyǔ was recorded, but it was most likely sometime during the early nineteen fifties, which was the heyday of their partnership.

A xiàngsheng piece is traditionally described as consisting of four basic elements; “joking”, “speaking”, “imitating”, and “singing”.³ However, speaking is by far the most important of these elements and the art of xiàngsheng is therefore, in many ways, a linguistic art (Xue,

² In Chinese: dānkǒu xiàngsheng 单口相声, duìkǒu xiàngsheng 对口相声 and qúnkǒu xiàngsheng 群口相声, respectively.

³ dòu 逗, shuō 说, xué 学 and chàng 唱, respectively (Xue, 1985: 7-33).

1985: 17). In fact, the most important characteristic of a xiàngsheng performer is the way he speaks. The speaking in a xiàngsheng piece should not be theatrical and exaggerated, but natural and leisurely. An artist should talk as if he were having an ordinary, informal conversation, and the skill of making their stage speech sound as natural as possible is practiced rigorously by the performers (Kaikkonen, 1990: 245). From a linguistic point of view, this leads to something of a paradox: everything a xiàngsheng performer says is planned and practiced to sound as natural as possible, but the defining characteristic of natural speech is precisely the fact that it is unplanned and unpracticed (Halliday, 1994). One could say that in order to sound as natural and unplanned as possible, the xiangsheng artist goes to the extreme opposite of what he is trying to imitate.

What does this tell us about xiàngsheng as an object of linguistic study? On the one hand, xiàngsheng speech is extremely artificial. Because everything the artists say is self-monitored, one should expect both phonetic and grammatical deviations from the way people speak in real life. On the other hand, xiangsheng pieces are imitations of natural speech, which, although different in nature, contain many of the same elements. Expressions characteristic of colloquial speech are frequent, and the pronunciation plays on typical colloquial conventions. Even hesitations, stammering, interruptions and false starts are carried over from everyday speech into the performances.

Because recordings of entirely natural speech are difficult to obtain, many phoneticians have attempted to recreate natural sounding discourse in their laboratories. In order to research Mandarin intonation structure, Susan Shen (1989: 14-15) had her informants rehearse the sentences they were going to pronounce in advance, and often used several “takes” to ensure that the resulting utterances sounded as natural as possible. Using xiàngsheng for phonetic study is in some ways similar to this; the xiàngsheng performers also rehearse their dialogues carefully before performing them, and the goal is to make them sound natural.

In the end, the question that must be asked is whether or not the xiàngsheng Xué Wàiyǔ can be accepted as a representative of generic Mandarin Chinese. As I have come to realize gradually through my work on this thesis, the answer to that question is definitely no. For one thing, enacted dialogues are necessarily different from naturally occurring dialogues, and the importance of this difference should not be underestimated. Also, Standard Mandarin is based on the Beijing dialect, and while both Mǎ Sānlǐ and Zhāng Qìngsēn are born in that city, they lived and practiced their art primarily in Tiānjīn. This does not mean that Xué Wàiyǔ is

spoken in the very distinctive Tiānjīn dialect; it most certainly is not. The accent of Xué Wàiyǔ most of all seems like a kind of “average” North Chinese colloquial speech. It is, however, natural to assume that it is influenced both by Tiānjīn and Běijīng conventions, although it would take a specialist in Chinese dialectology to determine the exact nature of these influences.

On the other hand, there does not seem to be any exact definition of what constitutes “Standard Mandarin” (Mair, 1991: 10-12). The closest one could get to a normative standard for Modern Mandarin pronunciation is probably the Xiàndài Hànyǔ Cídiǎn (2007); but as I’ve already mentioned, that dictionary gives pronunciation only in Pīnyīn, which specifies a very low level of phonetic detail. In any case, I do not claim that the observations made in this thesis are valid as general statements about Mandarin Chinese. They are accounts of the speech sounds uttered in this one enacted dialogue, which is an artificial construct based on the conventions of naturally occurring Mandarin.

1.2 Phonetic theory

Phonetics is the study of speech sounds; it is the study of sound used as a medium to express language (Abercrombie, 1968: 1-4). This means that phoneticians are generally not interested in all kinds of sound, but specifically in the sounds that human beings can learn to pronounce (Laver, 1994: 28-29). Humans have a remarkable ability to unconsciously recognize and reproduce speech sounds, but after having spent years using that ability to learn our native languages we often lose it to some degree (Laver, 1994: 556). If we are to learn the pronunciation of new and unfamiliar languages, we therefore need to start consciously exploring the way in which speech sounds can be produced. The phonetician Henry Sweet (1877) describes the importance of training our speech organs to pronounce new and unfamiliar sounds:

“Those who try to learn new sounds by ear alone, without any systematic training in the use of their vocal organs, generally succeed only partially. Even in those exceptional cases in which a naturally quick ear combined with favourable occasions for practice enables a linguist to acquire an accurate pronunciation of foreign languages by imitation alone, his knowledge is little better than that of a parrot, for he is unable to record the sounds he has learnt, or to teach them to others, and the results of his labours perish with him” (1877: 21).

To aid in this “systematic training”, phonetics classify sounds according to how the human speech organs produce them. In a phonetic transcription, these categories are represented by symbols, so that each symbol stands for a specific action (or rather, a combination of actions) that the speech organs perform. To read and write phonetic symbols you therefore have to be somewhat familiar with the speech organs and how they work. This study of the speech organs is specifically called **articulatory phonetics**.

Another branch of the phonetic sciences, called **acoustic phonetics**, is dedicated to studying the actual physical properties of sounds. Acoustic phonetics will be just as important as articulatory phonetics in this thesis, because I have relied heavily on the use of instruments to measure the acoustical dimensions of speech sounds. This has been done to ensure that phonetic symbols are used consistently throughout the transcription, and ensure (as far as possible) that the analytical categories that I employ correspond to objective phonetic fact. All the instrumental analyses in this thesis are performed with the free computer software called “Praat” (Boersma & Weenink, 2010).

I have tried to place myself in the tradition of the so-called “English school of phonetics”⁴. One central idea in this tradition seems to be that “phonetics”, as Daniel Jones (cited in Laver, 1997: xxv) puts it, “is a means to an end”. In my case, that end is to arrive at a systematic categorization of the speech sounds in the *xiàngsheng Xué Wàiyǔ*. I have tried as best I can to steer away from theoretical perspectives that are not directly relevant to achieving this end, and my account of phonetic theory may therefore seem simplistic to those with knowledge of the discipline. The general information about phonetics that I recount in my thesis is mostly drawn from David Abercrombie’s (1968) “Elements of General Phonetics”, Henry Sweet’s (1877) “Handbook of Phonetics”, John Laver’s (1994) “Principles of Phonetics”, Peter Ladefoged’s (2001) “Vowels and Consonants: An Introduction to the Sounds of Languages” and the second edition of “The Handbook of Phonetic sciences” (Hardcastle, Laver & Gibbon, 2010).

1.3 Phonological theory

Phonology is defined in “A Glossary of Phonology” as “the study of the sound systems found in human languages” (Carr, 2008: 139). This is distinct from, but obviously closely related to,

⁴ A widely used term, probably stemming from Henry Sweet’s (1877) assertion that “England may now boast a flourishing phonetic school of its own” (1877: viii). Laver (1997: xxv) uses the term “the British school”.

the discipline of phonetics, which is defined as “the study of human speech sounds” (Carr, 2008: 127).

There are two fundamental notions of phonology that is important to this thesis. One is the **phoneme**. A phoneme is the smallest sound-unit in a language that can cause a difference in meaning, and differences in sounds that do not cause a change in meaning are therefore not phonemic (Laver, 1994: 41-45). In order to find out which sounds are to be classified as different phonemes, phonologists look for **minimal pairs** (Laver, 1994: 30). A minimal pair is a pair of words or morphemes that have different meanings, and that are identical except for a single feature. Consider the English words “teal” and “deal”. They mean different things, and are identical except for the first sound. This tells the phonologist that, in English, /t/ and /d/ are two different phonemes. Because the only difference between these two segments is that /d/ is voiced and /t/ is unvoiced, voicing is said to be a **distinctive feature** in English (Laver, 1994: 110-112).

The other central notion of phonology is that of **distribution**. Distribution refers to the possible combinations of speech sounds in a language. In English, it is possible for some consonants to follow directly after one another, while other consonants cannot. For example, /s/ and /t/ can be combined (as in “stop” or “stand”), but /p/ and /k/ cannot; there are no words in English that contain the sequence /pk/. Using the two concepts of phoneme and distribution, it is possible to chart the entire sound system of a language (Laver, 1997: 38).

While this thesis is primarily a study in phonetics, I have found it useful to include some discussions on phonology as well. One reason for this is that I believe it will make reading easier for any readers who are not already familiar with Mandarin Chinese. The phonetic details revealed by a phonetic transcription are often more interesting if one has some knowledge of the system that spawned them. Moreover, some of the choices I have made as far as phonetic notation is concerned were influenced by phonological theory as well as by phonetic theory. For example, my transcription uses so-called “tone-letters” to denote syllable tones. The reason for this is that tone has a phonemic status in Chinese. Were I to transcribe a language like English, where tones do not have the same status in the phonological system, I would most likely not use these symbols at all.

The general account of Mandarin phonology in this thesis is based primarily on Y. R. Chao’s (1948; 1968) “Mandarin Primer” and “Grammar of Spoken Chinese”, Paul Kratochvil’s (1968)

“The Chinese Language Today: Features of an Emerging Standard”, and Wang Li’s (1982) “Hànyǔ Yīnyùn Xué 汉语音韵学”.⁵

1.4 Structure and presentation

Chapters 2 through 4 of this thesis are devoted to reviewing the theory and methods that inform my study. Chapter 2 presents the most relevant concepts of phonetic theory, and chapter 3 gives a very brief account of the phonology of Mandarin. In chapter 4 I describe how I have applied phonetic theory in the analysis of my particular data set. I give an account of all the analytical categories that I have distinguished in the transcription of Xué Wàiyǔ, and explain the methods for arriving at these categories. Together, these chapters describe the process of developing my phonetic notation. They are meant to contain all the information one needs to read and make sense of the phonetic transcription. In chapter 4, the symbols used in the transcription are all listed and described using both phonetic terms and, where possible, comparisons to sounds of English and other languages.

The next three chapters each single out a specific topic of Mandarin phonology that has been deemed of special importance to the present data set. In chapter 5 I attempt to find any reliable phonetic correlates of syllable stress. The results of this investigation were crucial to developing a method for distinguishing stressed syllables in the transcription. In chapter 6 I examine the Mandarin tone 4, and attempt to determine whether the changes this tone displays are best analyzed as results of tone sandhi or varying syllable stress. The results of this investigation have great influence on how tones have been transcribed in the phonetic text. Chapter 7 is devoted to examining tone 3, and discussing its phonetic behavior in the light of phonological theory.

The rest of the thesis is devoted to the transcription and translation of the xiàngsheng Xué Wàiyǔ. The transcription has been annotated with footnotes to draw the readers’ attention to some salient features of the pronunciation. A transcription in Pīnyīn has also been included, in to allow the reader to compare the phonetic and orthographic writing systems.

To illustrate theoretical points along the way, I have often used examples from other languages than Chinese. Where possible, I have used examples from English, and where

⁵ I should also mention that there are many important works on Mandarin phonology that I have not used here, such as Alexis Rygaloff’s (1955) “La Phonologie du Pekinois”, and Charles F. Hockett’s (1947) “Peiping Phonology”.

English did not furnish any fitting examples I have used Norwegian. In the digital version of this thesis, many of the examples are accompanied by embedded audio files that can be played by clicking on them. Quotations from the xiàngsheng are also accompanied by audio files. Hopefully, this will make it easier for the reader to continuously judge the validity of my arguments.

2 Phonetic theory

2.1 Articulatory phonetics

Speech is driven by the stream of air that is pumped out from the lungs⁶. As the air travels through the windpipe, it reaches the “voice-box”, usually called by its Greek name **larynx**, and the **vocal cords**. The vocal cords can be thought of as a pair of lips or “curtains” made of thin membrane, that are stretched horizontally across the windpipe. The cords and the passage between them are collectively known as the **glottis**. The glottis can perform several functions. It can close off the lungs entirely (which is what we do when we hold our breath) and open up to let the air through (which is what we do when we breathe normally). It can also be partially closed, making the vocal cords vibrate when the air escapes past them. When the vocal cords vibrate, they cause the air to vibrate at the same frequency, producing a stronger, more resonant sound. Sounds spoken with vibrating vocal cords are said to be **voiced**, and sounds spoken without any vibrations are said to be **unvoiced**. When we sing a song, the melody is made up of voiced sounds. Anyone who has tried to whisper a song will know that it is very difficult to make the melody heard at all. This is because it is mainly voiced sounds that carry tones. For the most part, our speech is voiced, with unvoiced parts forming breaks or junctures in the speech flow (Abercrombie, 1967: 20-28).

After passing through the glottis, the air-stream reaches the mouth or the nose, where it escapes the body. The oral and nasal cavities are the most important in the production of **vowels** and **consonants**, collectively called **segments** (Laver, 1994: 112-113), or sometimes simply “speech sounds” (Abercrombie, 1965; Ladefoged, 2001: 171-173). Segments are strung together to form **syllables**. Usually, the core, or **nucleus**, of the syllable will be a vowel, with consonants appearing as a kind of “modifiers to the vowels” (Sweet, 1877: 87). It is possible, however, for a consonant to form a syllable on its own. The reason why vowel-less syllables are quite rare in most languages is probably that consonants are generally less resonant than vowels, and therefore less audible. A language made up of significantly more consonants than vowels would therefore be quite impractical as a tool for communication since it would generally only be audible at short distances (Sweet, 1877: 87).

⁶ This is known as the egressive air-stream. The lungs also pull air in through the same route, which is called the ingressive air-stream. Usually, speech is produced with the egressive air-stream, but it is entirely possible to speak while breathing in as well (Abercrombie, 1968). For example, it is not uncommon for Norwegians to pronounce the word “ja” (meaning “yes”) with an ingressive air-stream.

Syllables are generally acknowledged to be an important unit of organization in all the world's languages, but they are also notoriously hard to define (Laver, 1994: 113-115). H. R. Stetson (1957) developed a very influential theory that syllables are connected to the so-called "chest pulses", the periodic contractions of the muscles that are responsible for pumping air out of the lungs. One such pulse produces one burst of air, which in turn results in one syllable of speech. This theory long remained popular among some scholars (Abercrombie, 1968: 34-36), but was rejected by others (Luo, 1957: 104-108).

Later studies have shown that one chest pulse not necessarily corresponds to one syllable, and that the chest-pulse theory is therefore inaccurate. However, the fact remains that the syllable is very significant to the way humans perceive and organize speech (Ladefoged & Maddieson, 1996: 282). If asked to do so, most people seem able to divide any given utterance into syllables (Abercrombie, 1968: 34). For example, most speakers of English will be aware that the word "refrigerator" is made up of five syllables: "re-frig-e-ra-tor". Similarly, most Norwegians will know that the number of syllables in the corresponding Norwegian word, "kjøleskap", is three: "kjø-le-skap".

Syllables are strung together to form larger units. Exactly what these larger units are, there is no general agreement on. Researchers have suggested complex hierarchies of units to describe the organization of speech above syllable level⁷. It seems that the units one chooses to divide an utterance into are mostly dictated by the needs of each specific analysis. I will assume that utterances can be divided into intonation phrases, or **phrases** for short. An intonation phrase is the domain of a single pattern of speech "melody". One phrase corresponds to an articulatory gesture, because it is usually (although not always) the sequence of speech that we can utter in a single breath (Abercrombie, 1968: 108; Laver, 1997: 492; Cao, 2003). The phrase is also a useful analytical unit in phonology, because it usually (but not always) corresponds to one grammatical sentence (Abercrombie, 1968: 108; Halliday & Matthiessen, 2004: 11-17; Laver, 1997: 492).

2.2 Acoustic phonetics

How speech sounds are articulated is not the only concern of phonetics. While analytical categories are generally defined according to the actions performed by the speech organs, it will normally not be possible for the phonetician to observe the speech organs directly. He

⁷ See Fletcher (2010) for an up to date summary of this field.

must be able to classify sounds indirectly; first he hears them, and then he deduces how they are pronounced. What the phonetician observes is therefore the physical properties of sound that the human ear is capable of picking up. These properties can be summed up as **pitch**, **duration**, **loudness** and **quality** (Laver, 1997: 27; Luo, 1957: 33-37).⁸

Physically, sounds are vibrations of the air, or **sound waves** (Ladefoged, 2001: 5). The most basic of the graphic representations of sound used in acoustic studies is the so-called **waveform** (figure 2.1). This is a two dimensional representation of sound waves, where the air is (rather abstractly) represented as a thin black line. A completely straight line represents air which is absolutely still, and when the line deviates from the middle the air is vibrating. The greater the deviation from the middle, the louder the sound; which means that a straight line represents absolute silence (Ladefoged, 2001: 5-8; Luo: 32).

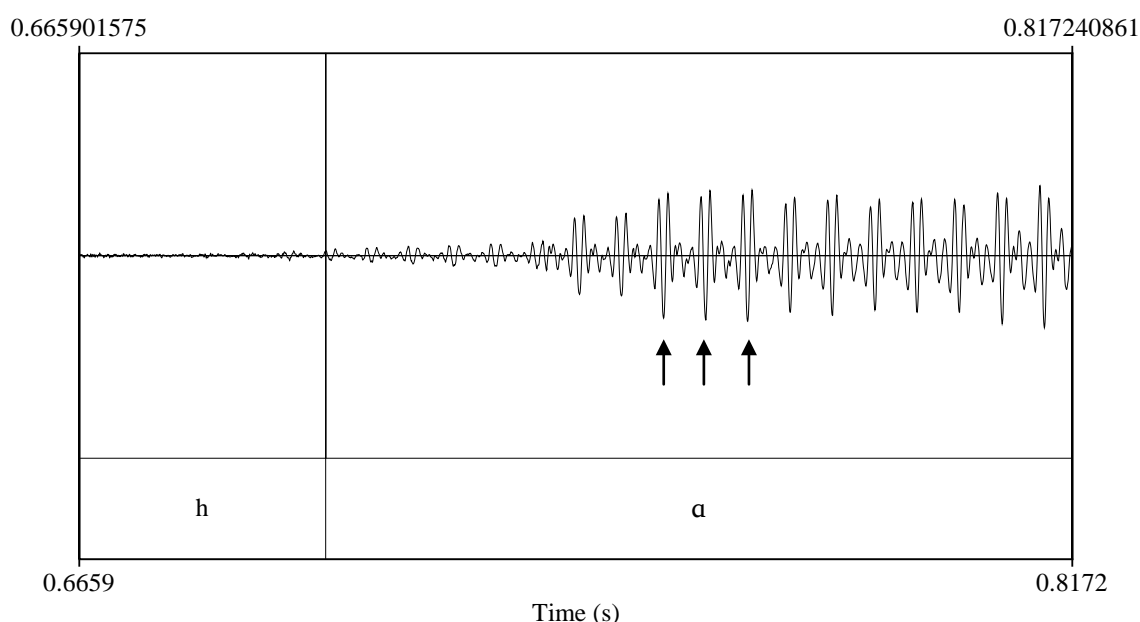


Figure 2.1: Waveform of [hɑ:]

The waveform in figure 2.1 is produced from a recording of me saying [hɑ:]. I described above how the the vocal cords could be opened to let the air-stream rush past noiselessly, and constricted and made to vibrate in order to create noise. In figure 1.1, [h] is really just the air rushing out of the lungs with both the glottis and the mouth wide open. This only produces an unvoiced sound, barely audible in itself. At around the point where the [ɑ:] begins, the glottis starts to contract and the vocal cords begin to vibrate. A stronger, more resonant sound

⁸ Luo (1957: pp. 33-37) uses the terms yīngāo 音高 “sound height”, yīncháng 音长 “sound duration”, yīnshì 音势 “sound amplitude” and yīnsè 音色 “sound colour”.

emerges as the vibrations from the cords cause more powerful ripples in the air. The waveform therefore shows that [h] is an unvoiced sound and that [ɑ:] is an unvoiced sound. The duration of the segments can be read from the horizontal axis, and the loudness can be seen from the size of the sound waves.

Pitch is what we, in everyday speech, often call “notes” or “tones”. In other words, in the C minor scale, a D has higher pitch than a C; an E^b has higher pitch than a D, etc. The physical quality that makes a sound high-pitched or low-pitched is the speed, or **frequency**, at which the air vibrates. In phonetic literature, pitch is often referred to as **fundamental frequency** (abbreviated to **F0**), and the unit most commonly used to measure F0 is called **Hertz** (abbreviated to **Hz**). Hertz stands for the number of vibrational cycles that take place in the space of one second, so that a sound with 200 cycles in one second has a fundamental frequency of 200 Hertz. The arrows in figure 1.1 mark the peaks that begin each vibrational cycle. The number of such peaks in one second is equal to the frequency, measured in Hz (Ladefoged, 2001: 5-6).

In each main cycle there are a number of lesser cycles, visible in the waveform as smaller ripples between the vibrational peaks. These can tell the viewer some things about the quality of the sound (Ladefoged, 2001: 7). Though higher resolution is required for close inspection, figure 1.1 shows the cycles of [ɑ:] to be relatively stable and regular. The cycles in [h] on the other hand, are erratic and uneven. The sound quality is what distinguishes vowels and consonants from each other.

2.3 Consonants

A consonant is produced when the airstream from the lungs encounter some kind of “obstacle” on its way out of the body. The obstacle is formed by one of the speech organs blocking or tightening the passage of the airstream. This process is called **articulation**. Consonants are classified according to two main parameters; **how** the airstream is obstructed, or **manner of articulation**; and **where** the airstream is obstructed, or **place of articulation** (Abercrombie, 1968: 47-50). There is a wide range of possible articulations that are used in human languages. Here I will only summarize those that are relevant to this thesis. A complete list of manners and places of articulation can be found on the web pages of the International Phonetic Association (IPA, 2005).⁹

⁹ All articulations mentioned below are stated as specified by the IPA (2005), and Ladefoged (2001: 99)

The relevant manners of articulation are as follows: **Plosives** are made by blocking off the air-stream entirely for a very short time, and then releasing it with a sudden burst of air, the “c” in “coffee” is a plosive, and so are both the “k” and “t” in “kitty”. **Fricatives** are produced by constricting the articulators to such a degree of closure that the air can barely escape. This gives fricatives a kind of “whistling” quality, like the “s” in “snake”. The “hushing” sound we make when we want someone to be quiet (shhhh) is really a drawn out fricative. **Affricates** are combinations of a plosive and a fricative that are pronounced as a single sound, such as the “ch” in “chance”. These are not so common in English, but very common in, for example, Chinese. **Nasals** are made by opening the velum (see above), allowing the air to escape through the nose and at the same time closing off the mouth with the tongue or lips. This produces sounds like the “n” in “no”. **Laterals** are sounds made when the tongue obstructs the central passage through the mouth, forcing the air to flow out along the sides. A typical example of a lateral is the “l” in “live”. **Approximants** are the least strictly defined category, as they are segments that are not conveniently described as either vowel or consonant. Sounds that have all the physical characteristics of vowels but function like consonants, for example vowels that do not form the nucleus of a syllable, will often be classified as approximants (Laver, 1994: 148). Typical examples of this would be the “y” in English “yes” and the “j” in Norwegian “jeg” (“I”).

The relevant places of articulation are as follows: When constriction is caused by the lips pressing against each other, the consonant is **bilabial** (“p” in “peat”). When the front teeth are pressed against the lips, the consonant is **labio-dental** (“f” in “front”). When the tip of the tongue is pressed against the gums, right behind the front teeth, the sound is **alveolar** (“s” in “sin”). When the ridge of the tongue is used to press against the area just behind the gums, the consonant is **alveolo-palatal**¹⁰ (there are no alveolo-palatal sounds in English, but the “kj” in western Norwegian “kjeller” is a close analogue). When the tongue is curled up and back so that the tip touches the roof of the mouth, the consonant is **retroflex** (“r” in “ready”). If the root of the tongue touches the soft, rearmost part of the mouth’s roof, the resulting sound is **velar** (“c” in cold). When the object causing the constriction is the uvula¹¹, the consonant is **uvular** (English has no uvular consonants, but the “r” in French “roue” is typically uvular). If the constriction does not occur in the mouth at all, but in the glottis, the sound is called a

¹⁰ The term “alveolo-palatal” is used specifically by Ladefoged & Maddieson (1996: pp. 150-151) to describe certain fricatives of Mandarin Chinese.

¹¹ The fleshy appendage that hangs from the back of the palate.

glottal (creates a “coughing sound” which is usually ignored in written language, but sometimes represented by an inverted comma, as in “Hawai’i”).

Most consonants, regardless of how and where they are articulated, can be either **voiced** or **unvoiced**. Voiced sounds, as I mentioned above, are produced when the vocal cords vibrate. Consider for example the labio-dental fricatives [f], and [v]. They are articulated in precisely the same place and manner, but the [v] has a much stronger, more resonant sound. This is because the [v] is voiced, while the [f] is unvoiced (Abercrombie, 1968: 27). Some consonants can also be **aspirated**, in which case they are pronounced with a strong burst of air accompanying the release, making them sound “harder” or more forceful. When a consonant sound is classified, it is usually specified for all four of these parameters. For example, the “t” in “terrier” is an “aspirated unvoiced alveolar plosive”.

When they are connected and made into words and sentences, consonants sometimes absorb articulatory features of their neighbors. This process is known as **assimilation**. The features that are most typically assimilated are the place of articulation, voice and nasality. An example of assimilation can be found in the English phrase “I’m very tired”. The “m” in “I’m” is normally a bilabial nasal, but in rapid speech it may assimilate the labio-dental place of articulation from “v” and become a labio-dental nasal [m]. If a segment influences the segment that is spoken directly before it, as in this example, assimilation is said to be **regressive**. If the assimilation works “forwards”, that is, if a segment influences the following segment, the assimilation is said to be **progressive** (Abercrombie, 1968: 133-139).

2.4 Vowels

Vowels not have a specific place of articulation, because they are made when the air-stream is allowed to escape unhindered through the mouth. And yet there is clearly more than just one kind of vowel. The characteristics of a vowel, which make up the **vowel quality**, are classified along three dimensions. The first two have to do with the position of the tongue in the oral cavity. While the tongue does not constrict the airflow as is the case when pronouncing consonants, it is still possible for it to assume different positions in the mouth. The important factors for the sound of a vowel concern both the tongue’s horizontal and vertical movements. A vowel pronounced with the tongue in a high position sounds different from one produced with the tongue in a low position, etc. When the tongue is in a high position, the vowel is said to be **close**, and when the tongue is in a low position the vowel is **open**. When the tongue is in a forward position, the vowel is **front**, and when the tongue is in

a retracted position the vowel is **back**. The third dimension has to do with the lips. A vowel that is pronounced with rounded lips (a **rounded** vowel) will sound differently from a vowel pronounced with spread lips (an **unrounded** vowel) (Abercrombie, 1968: 55-58; Ladefoged, 2001: 106-117).

The acoustic effects caused by the position of the tongue and the lips are not as obvious as those caused by the articulation of consonants. While the consonants break off the airstream or direct it through narrow paths between the articulators, vowels simply change the way the sounds resonate in our bodies. The total acoustic makeup of vowel quality is fairly complex, and is still not completely understood by scientists¹². Simply put however, the tongue divides the vocal tract into two distinct “chambers of resonance”; the throat and the mouth. When the vocal cords begin to vibrate, they will make the air in the throat vibrate at one frequency and the air in the mouth vibrate at a different frequency (Ladefoged, 2001: 32-34). A fronted tongue will make the forward resonance chamber smaller, which means that it will resonate at a higher frequency. In the same way, a raised tongue will make the air in the rear resonance chamber vibrate faster. Peter Ladefoged (2001: 33) likens this to blowing across the top of a glass bottle; if you blow at an empty bottle the sound will be low pitched, but if you fill some water in the bottle the resonance chamber will be smaller and the pitch will be higher.

The waveform in figure 1.1 showed cycles of vibrations, with each cycle ending in a vibrational peak. The interval between the peaks corresponds to the fundamental frequency of the sound. In between the peaks there were smaller cycles of vibrations that repeat themselves many more times. These represent the frequencies of the various resonance chambers in the vocal tract. The quality of a vowel can be more accurately illustrated by a spectrograph. The spectrograph is an instrument which reads the amount of acoustic energy that is released in different areas of the spectrum, and “paints” energy intensive frequencies in dark shades; the greater the energy, the darker the shade. The resulting graphic representation is called a **Spectrogram**. Figure 2.2 shows spectrograms of the vowels [a] and [i].

¹² See Harrington (2010) for an up-to-date summary of scientific progress in this area.

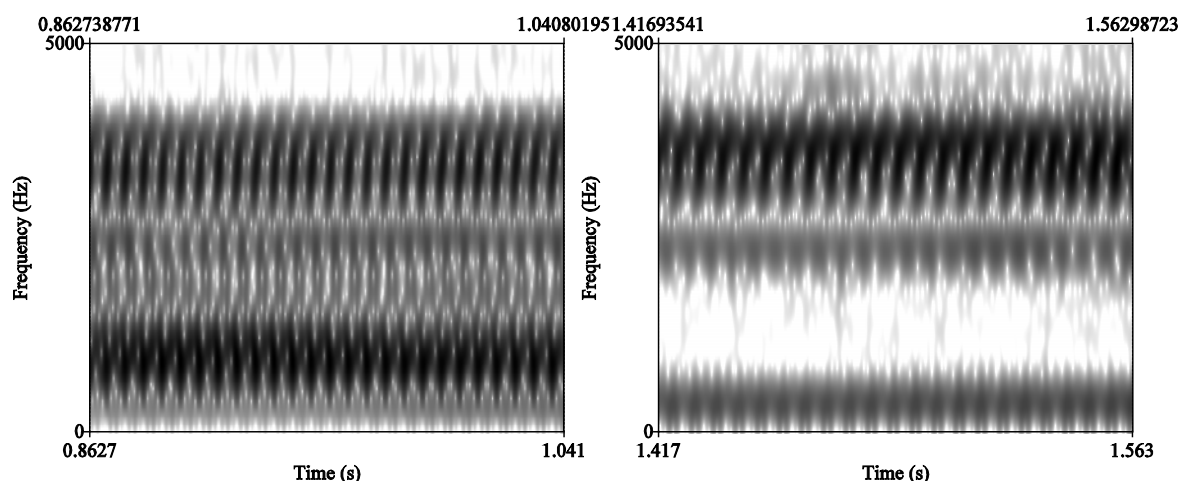


Figure 2.2: Spectrograms of vowels [a] (left) and [i] (right)

The concentrations of energy that can be seen as dark bands across the spectrograms are called **formants**. Remember that the fundamental frequency is often abbreviated to F0. Since the formants are harmonies, superimposed on the fundamental frequency, they are abbreviated as F#, where # stands for the number of harmonies above F0. This means that the lowest formant will be called **F1**, the number two formant from the bottom will be called **F2**, etc. In the case of vowels, F1 corresponds to the resonance in the throat, and F2 corresponds to the resonance in the mouth (Ladefoged, 2001: 32-34). In figure 1.3, the F1 and F2 of [a], which is an open back vowel, are so close that they can be difficult to separate. F1 lies around 800 Hz, and F2 around 1000 Hz. F1 (the throat resonance) lies at a high frequency because the tongue is in a low position, and since F2 (the mouth resonance) lies at a low frequency because the tongue is far back. In [i], the situation is completely opposite; because [i] is a close front vowel, F1 is about 350 Hz and F2 about 2200 Hz.

When a vowel is spoken alone or surrounded by consonants, it is called a **monophthong**. When two vowels are strung together without any intervening consonants they are called a **diphthong**, and when three vowels are strung together they are called a **triphthong** (Laver, 1994: p. 143). In rapid speech, vowels are often pronounced with very little articulatory force. When this happens, the tongue has a tendency to gravitate towards the center of the mouth, where it needs to expend no effort to maintain its position. Because of this, vowels in particularly short or unstressed syllables have a tendency to be pronounced as the mid-central vowel [ə].

2.5 Intonation, tone and stress

While they have some inherent characteristics of pitch and loudness, consonants and vowels are largely distinguished on the basis of quality. Any combination of vowels and consonant can be spoken both softly and loudly, and at both high and low pitch. An [a] will still be an [a], regardless of how loudly you say it. Because of this, features of pitch, loudness and duration are often described as being located “above” the segments. Consequently, they are known as **supra-segmental** features (Carr 2008: p. 170).

Sentences or phrases are spoken with varying pitch, much like a musical melody. In many languages, the melody of a sentence, which in phonetics is referred to as **intonation**, will influence its meaning. In English, the phrase “you’re going to the bathroom” could be a question if it is pronounced with a rising intonation (“you’re going to the bathroom?”) and a statement if it is pronounced with a falling intonation (“you’re going to the bathroom!”). However, intonation will not cause the phrase to mean something entirely different; no matter what intonation it is pronounced with, it will still concern the participant “you” performing the action of “going to the bathroom” (Abercrombie, 1968: 104-105).

Some languages, most famously Chinese, assign distinct melodies to each word or syllable. In these languages, words are distinguished on the basis of pitch alone. This kind of “word melody”, as opposed to sentence melody, is often referred to as **tone**. One language that has word tones is Norwegian. Take, for example, the two Norwegian words “tanken” (“the thought”) and “tanken” (“the tank”). These are written the same way, and the vowels and consonants all sound the same; the difference lies in the tone.

To display pitch as conveniently as possible, Praat and other computer programs can automatically calculate the number of cycles per second, and then represent the pitch of an utterance as a graph where the vertical axis represents frequency (Hz) and the horizontal axis represents time (s). From this, we get a line, or contour, illustrating the melody of the utterance. This is known as the **pitch contour** (Laver, 1994: p. 155). The pitch contours of the two Norwegian¹³ words mentioned above are displayed in figure 2.3:

¹³ The speaker of these two particular words is me. My native tongue is a Middle-Norwegian dialect spoken just south of the city of Trondheim, which means that the pitch contour displayed here probably differs from what one would hear from a speaker of Standard Norwegian.

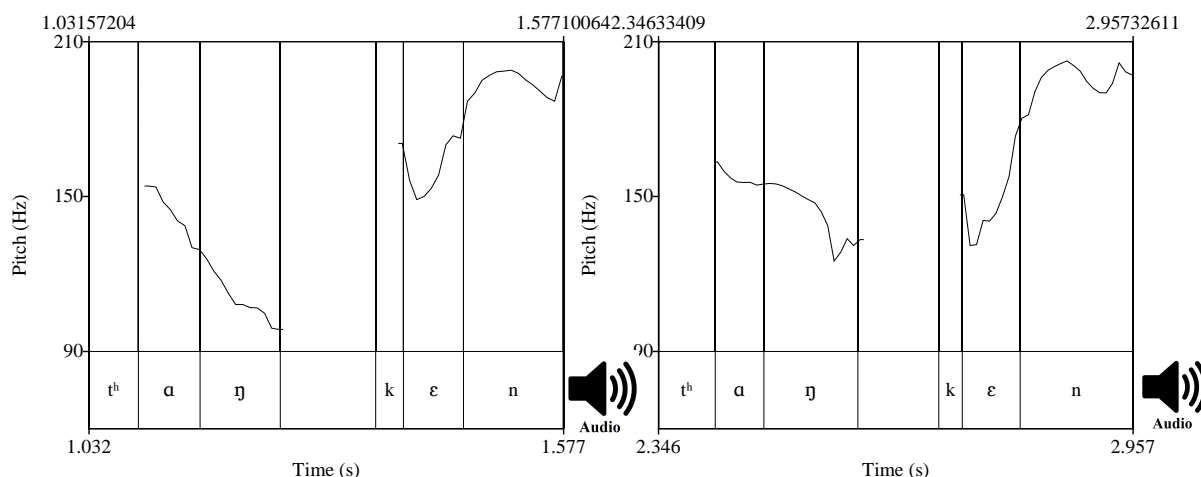


Figure 2.3: Pitch contours of tanken (“the tank”) (left) and tanken (“the thought”) (right)

In a phonetic transcription, tones can be represented in a number of ways (Abercrombie, 1968: pp. 132). Within Chinese linguistics, the most common notation for tones is a system invented by Y. R. Chao (2006a) in 1930. This system uses a scale of five tiers, where [1] represents a low pitch, [5] represents a high pitch, and [2] to [4] are intermediate levels. These levels can be represented either with numbers like these, or with a “tone-letter” which is a miniature graph set to the right of a vertical bar. The graph moves from one to another of the five levels, giving an image of the pitch similar to that of the pitch contour. Judging from figure 2.3, the two syllables in “tanken” (“the tank”) can be assigned the values [31] and [45], respectively. In tanken (“the tank”), the second syllable is very similar, but the first syllable has a higher tone; [43]. Using tone letters, the words can be transcribed phonetically as [tʰaŋ˥˥.kən˨˨] (“the thought”) and [tʰaŋ˥˥.kən˨˨] (“the thought”).

The final supra-segmental feature that I will touch upon here is that of syllable **stress**. A **stressed syllable** is a syllable that is somehow emphasized and made to stand out from the surrounding syllables. Often a stressed syllable will stand out because it is longer in duration, higher in pitch and greater in loudness than other syllables, but some languages systematically distinguish stressed syllables through articulation (Laver, 1997: 450, 511-517).

In the chest-pulse theory of H. R. Stetson (1954), one chest-pulse was seen as corresponding to one syllable. This theory explained stress through the observation that some contractions of the breathing muscles are stronger than the rest. These so-called stress-pulses produce a more powerful burst of air and result in syllables that are emphasized over the surrounding ones. While the chest-pulse theory remained influential for a long time, it was eventually shown

that one chest pulse not necessarily corresponds to one syllable. It has later been said of Stetson's (1954) theory that it would be "the nicest available definition of stress, if only it were true" (Hayes, 1995: 5). Today, there does not seem to be any general agreement on the articulatory basis for syllable stress (Fletcher, 2010).

While tone and intonation create a speech melody, stress is closely linked to speech rhythm. All human languages have a rhythm. According to David Abercrombie (1968), "rhythm [...] arises out of the periodic recurrence of some sort of movement, producing an expectation that the regularity of succession will continue" (Abercrombie, 1968: 96). Based on this definition, he classified the speech rhythms of different languages into two distinct categories. In **stress-timed** languages, the stressed syllables form the basis of the rhythm. In these languages, of which English is an example, syllables occur at uneven intervals, while stressed syllables are evenly spaced. The rhythm arises from the regularity of stressed syllables. In **syllable-timed languages**, for example French, the opposite is the case. Here, each syllable is equally spaced, while stressed syllables are scattered around at unequal intervals (Abercrombie, 1968: 96-98). The general validity of this theory has been questioned in later years, but the typology of stress-timed and syllable-timed languages still remains in widespread use (Fletcher, 2010; Laver, 1997: 527-533).

3 An overview of Mandarin phonology

The most important unit in the study of Mandarin phonology is the syllable. As Paul Kratochvil (1968 p. 23) notes, when asked to divide an utterance into smaller parts, a speaker of English is most likely to divide it into words. An educated Chinese speaker asked to do the same thing is likely to divide the utterance into syllables. What constitutes a word in Chinese is a question that even linguists are pressed to answer, but Chinese morphemes are in almost all cases monosyllabic. The writing system of Chinese also reflects this, as all but a very few characters stand for a single syllable. This has led Kratochvil (1968) to call the syllable “the basic natural unit in [Modern Standard Chinese]” (p. 23).

In traditional Chinese phonological theory, the syllable is said to be made up three parts; the **initial**, the **final** and the **tone**¹⁴. The initial is a consonant or semivowel, while the final is a vowel, a group of vowels or a vowel followed by a consonant (Wang, 1982: 40-43). Because the ancient Chinese did not have an alphabetic writing system, a method was developed sometime in the late Han dynasty (ca. 180-220 AD) to transcribe the pronunciation of a character using two other characters. This system, called *fǎnqiè* 反切 “turning and cutting”, denoted the structure of a syllable by taking the initial of one character and combining it with the final of another character. The tone of the syllable was denoted by a certain combination of tones in the two characters used, though the exact method varied somewhat from era to era. (Wang, 1982: pp. 108-112). For example:

东 德 红 切¹⁵

dōng, dé hóng qiè

“‘Dōng’ has the initial of ‘dé’ and the final of ‘hóng’”

In modern times, a great variety of systems have been developed for writing the sounds of Chinese in roman letters. The system which is now the most common is *Hànyǔ Pīnyīn*, which I have already mentioned several times. While alphabetic, *Pīnyīn* is also based on the notion that syllables are divided into initials and finals (Lin, 2007: p. 123).

3.1 Segmental structure

The segmental structure of a Mandarin syllable is (C)V(C). The (C) stands for an optional consonant and the V for an obligatory vowel. This means that a syllable may consist of only a

¹⁴ In Chinese 声母(or 纽 niǔ), 韵母 yùnmǔ and 声调 shēngdiào respectively (Wang, 1982: 40-81).

¹⁵ Example quoted from Wang (1982: p. 109)

vowel, a consonant followed by a vowel, a vowel followed by a consonant, or a consonant followed by a vowel followed by a consonant (Kratochvil, 1968: p. 24). The initial is defined as the consonant that precedes the vowel. If there is no initial consonant, the initial may be a vowel or the syllable may have the so-called “zero-initial” (Chao, 1968: p. 18). Phonetically, syllables with the zero-initial will have a variety of non-phonemic initial consonants, sometimes reported to be glottal plosives, glottal fricatives (Duanmu, 2000: p 83), velar fricatives or uvular fricatives (Chao, 1968: 20-21). Syllables belonging to this type include /an/ and /ao/.

The initial consonants of Mandarin are given in table 3.1, with Pīnyīn spelling in slanting brackets and typical phonetic realization in square brackets. Shaded cells represent articulations that do not exist in the phonemic system. All phonetic values given here and throughout this chapter are meant to illustrate the typical way of pronouncing a phoneme, and it should be noted that in connected speech they seem to vary a great deal. The phonetic values are based on the material I have studied and transcribed, and may therefore differ from those found in the more comprehensive accounts of Mandarin phonology.

	Bilabial	Labio-dental	Alveolar	Retroflex	Alveolo-palatal	Velar
Plosive	/p/ [p ^h]		/t/ [t ^h]			/k/ [k ^h]
	/b/ [p]		/d/ [t]			/g/ [k]
Fricative		/f/ [f]	/s/ [s]	/sh/ [ʂ]	/x/ [ɕ]	/h/ [x]
Affricate			/c/ [ts ^h]	/ch/ [ts ^h]	/q/ [tɕ ^h]	
			/z/ [ts]	/zh/ [ts]	/j/ [tɕ]	
Lateral			/l/ [l]			
Nasal	/m/ [m]		/n/ [n]			
Approximant				/r/ [ɻ]		

Table 3.1: Mandarin initial consonants

Perhaps the most striking feature of this system of consonants is that there is no opposition of voiced and unvoiced.¹⁶ Phonemes like /p/ and /b/ are not distinguished by voicing, but by aspiration. This is different from English, where voicing is a distinctive feature. Thus there are minimal pairs in Mandarin that are distinguished by aspiration alone, such as:

¹⁶ As noted by Kratochvil (1968: p. 26). My analysis is somewhat different from Kratochvil's, however, as he treats the /r/ phoneme as a voiced retroflex fricative, while I treat it as an approximant. Kratochvil therefore counts 神 shén and 人 rén as a minimal pair distinguished by voicing.

攀 pān "to climb" → [p^han̚]

搬 bān "to move" → [pan̚]

If the initial consists of the first C in a syllable, it follows that the final will consist of V(C). The V element can be a close vowel, a mid vowel or an open vowel. It can also be a monophthong, diphthong or a triphthong. Diphthongs are based around one nuclear vowel, which is either preceded or followed by a shorter vowel segment. I will refer to the short vowel as an **on-glide** if it occurs before the nuclear vowel, and an **off-glide** if it occurs after. Triphthongs consist of a syllabic vowel along with both an on-glide and an off-glide. In Mandarin diphthongs and triphthongs, some of the vowels are much shorter than others. In these cases very short vowels are written with a superscript character (following Howie, 1973).

The optional final element of the Mandarin syllable, (C), is either one of the following:

/n/ → [n]

/ng/ → [ŋ]

No other consonants are possible in this position. This means no syllable can end on anything but these two consonants, or a vowel.

The finals of Mandarin that end on a vowel (called open finals) are given in table 3.2. The rows give the phonetic height of the syllabic vowel and the columns show whether it occurs as alone or with an on-glide, with an off-glide or with both. Table 3.3 displays the finals that end with a consonant. Since no consonant can follow a triphthong or an off-glide, the rows give the height of the nuclear vowels and specify whether or not it is preceded by an on-glide. The columns give the final consonant.

When the final has an on-glide and there is no initial consonant, Pinyin spelling conventions dictate that the syllable be written with an initial "w", as in 问 wèn [ʷən] "to ask", or "y" as in 眼 yǎn [ʰɛn] "eye" (Lin, 2007: p. 126). In the tables, these forms are given to the right of the normal spelling.

	Monophthong	On-glide	Off-glide	Triphthong
Close	/i/ [i], [ɿ] or [ʅ]			
	/u/ [u]			
	/ü/ [y]			
Mid	/e/ [ɤ]	/o/, /wo/ [ʊɔ]	/ei/ [eɪ]	/iu/ [ʰioʊ]
		/yo/ [ʊɔ]		
		/ie/ [ʰɛ]	/ou/ [oʊ]	/ui/ [ʰəɪ]
		/üe/, /yue/ [ʰɛ]		
Open	/a/ [ə]	/ia/, /ya/ [ʰə]	/ai/ [aɪ]	/uai/, /wai/ [ʰəɪ]
		/ua/ [ʰə]	/ao/ [ɑo]	/iao/, /yao/ [ʰɑo]

Table 3.2: Mandarin open finals

		Final /n/	Final /ng/
close	monophthong	/in/ [in]	/ong/ [ʊŋ]
		/ing/ [i̯ŋ]	
		/yun/ [y̯n]	
	on-glide		/iong/, /yong/ [ʰiŋ]
mid	monophthong	/en/ [ən]	/eng/ [əŋ]
	on-glide	/ian/, /yan/ [ʰɛn]	
		/un/, /wen/ [ʰən]	/weng/ [ʰəŋ]
open	monophthong	/an/ [an]	/ang/ [aŋ]
	on-glide	/uan/, /wan/ [ʰan]	/iang/, /yang/ [ʰaŋ]
		/yuan/ [ʰən]	/uang/, /wang/ [ʰaŋ]

Table 3.3: Mandarin closed finals

The Close unrounded vowel phoneme /i/ has three different realizations, depending on which initial precedes it. It is pronounced [ʅ] after /zh/, /ch/, /sh/ and /r/; [ɿ] after /z/, /c/ and /s/; and [i] after all other initials. Pronunciation of the other phonemes may also vary, but not in as systematic a way as /i/.

There are restrictions on which initials and finals that can be combined. The most important are the following: The alveolo-palatal initials can only occur with finals that begin with /i/ or /y/. The velar initials, on the other hand, cannot occur with any final that starts with /i/ or /y/. The retroflex initials can also not occur with /i/ or /y/, except with the monophthong /i/. The vowel /y/, as a monophthong or as part of a diphthong, cannot occur with any initials but the alveolo-palatals, /l/, and /n/.

In addition to the syllable finals given in the tables above, Mandarin has a special final which occurs only with certain words, but which may occur with all types of syllables. This final is indicated in Pīnyīn by the addition of an /r/ at the end of the syllable. Phonetically, the final /r/ is realized by dropping the final nasal consonant (if there is one to begin with) and pronouncing the vowel with the tongue in a retroflex position (Chao, 1948: 30). This kind of vowel articulation is called **r-coloring** (Ladefoged, 2001: 27).

3.2 Tones

All Mandarin tonic syllables have one of four tones (Chao 1968, p. 25). The phonetic shapes of these tones can be represented with a graph like that in figure 3.1. This graph corresponds to the tone letters I introduced in section 2.5:

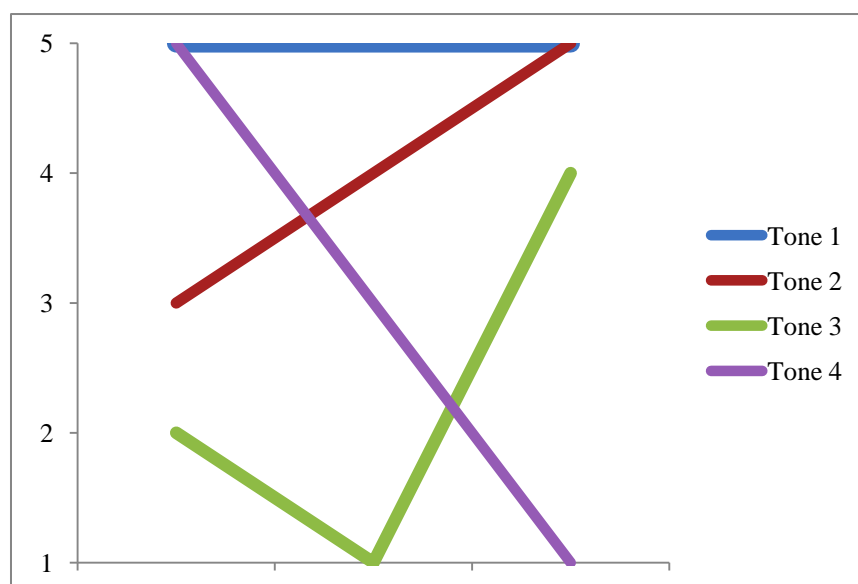


Figure 3.1: The four tones of Mandarin.

Tone is a distinctive feature in Mandarin. This can be illustrated through these four contrasting words, often cited in textbooks on Chinese language. The Pinyin spelling conventions for the four tones are given next to the characters:

mā 妈	→	[mæ˥]	”mother”
má 麻	→	[mæ˥˥]	“hemp”
mǎ 马	→	[mæ˨˥]	“horse”
mà 骂	→	[mæ˨˩]	“to curse (someone)”

I say that all “tonic” syllables have one of these four tones, and this of course implies that Mandarin also has “atonic” syllables (Kratochvil, 1968: p. 35). Although nominally atonic, these syllables also contain voiced elements, and as such they do carry pitch. Because of this I will use the term **neutral tone** syllable, after Y. R. Chao (1968: p. 35). The defining characteristics of neutral tone syllables, according to Chao (1968: p. 35), are that they are very short in duration and that there is no definite pitch contour. The pitch level of the neutral tone is commonly thought to be determined by the preceding tone. Chao (1968: p. 36) gives the following rules for the behavior of the neutral tone. The pitch level is given as a single point on the five-tier scale, and represented with an alternative tone letter consisting of a dot against a vertical bar:

After T1:	[2]	tā de 他的 “his” [tʰæ˥˥̣ t̚˥˥]
After T2:	[3]	shéi de 谁的 “whose?” [ʂɛi˥˥̣ t̚˥˥]
After T3:	[4]	nǐ de 你的 “your(s)” [ni˥˥̣ t̚˥˥]
After T4:	[1]	dà de 大的 “big one(s)” [t̚˥˥̣ t̚˥˥]

Some neutral tones may gain a falling or rising contour as an effect of intonation, particularly if they occur at the end of a phrase (Chao, 1948: p. 27; Shen, 1989: p. 38-48).

In connected speech tones sometimes exert a certain influence each other, so that some tones may change their shape due to influence of adjacent tones. Some of these phenomena have been extensively reported and discussed by linguists for many years. It has become common practice to use a Sanskrit word, **sandhi**, to denote tone change¹⁷ in Chinese. The term, which was first used in ancient Indian linguistics, means “putting together”. It is often used to cover any phonetic process that operates across the boundaries of linguistic units such as words or

¹⁷ The term “tone change” is sometimes used to denote a specific process, reminiscent of inflection, whereby words signal a change in grammatical function by adapting a different tone (Chen, 2000: 31). Since this phenomenon is not relevant to the present study, I will use “tone change” more freely to refer to any kind of phenomenon that involves tones changing.

syllables (Carr, 2008: 153). In Chinese linguistics, however, the term sandhi is mostly used to describe the behavior of tones, and it is therefore common to talk of **tone sandhi**. The most generally acknowledged and frequently discussed tone sandhi are those reported by Y. R. Chao (1968).¹⁸

The most extensively researched tone sandhi phenomenon is the so-called **tone 3 sandhi**, which was reported by western linguists as early as 1682 (Shih, 1997: 82). Chao (1968) describes T3 sandhi in the following way: “when a [tone 3] is followed by another [tone 3], the first one changes into a [tone 2]” (Chao, 1968: 27). Because of this, certain words that are normally pronounced differently, such as zhǎo 找 “to look for” and zháo 着 “to be affected by”, can become identical in certain contexts. According to Chao (1968: 27) 找火 zhǎo huǒ “to look for fire” becomes homophonous with 着火 zháo huǒ “to catch fire”:

zháo huǒ 着火 “to catch fire”: [1 3] → [1 3]

zhǎo huǒ 找火 “to look for fire”: [3 3] → [1 3]

Tone 3 also displays another kind of sandhi, which occurs when it is followed by any tone but another tone 3. When this happens, the tone 3 is cut in half, and only the first, low falling part is pronounced (Chao 1968: p. 27). This phenomenon is often called **half tone 3 sandhi** (Chen, 2000: p. 21). For example:

mǎimài 买卖 “to buy and sell”: [3 3] → [3 3]

The final tone sandhi reported by Chao (1968: pp. 27-28) is **tone 2 sandhi**. This sandhi causes a tone 2 to change from its normal rising [35] to a level [55], when is preceded by a T2 or a T1 and followed by any tone except the neutral tone. Chao (1968: 28) gives the following examples (among others):

sānniánjí 三年级 “third-year class”: [1 1 1] → [1 1 1]

cōngyóubǐng 葱油饼 “onion oil cake”: [1 1 3] → [1 1 3]

¹⁸ The terms are actually separated in Chinese, where sandhi is called 连音 liányīn “connected sounds”, and tone sandhi is called 变调 biàndiào “tone transformation”.

3.3 Stress

Stress in Mandarin Chinese is a somewhat controversial topic. The most common view seems to be the one proposed by Y. R. Chao. He finds that Mandarin has three levels of stress; contrasting stress, normal stress and weak stress. Contrasting stress is used to signal that one element of a phrase is emphasized or contrasted against another element, by stressing it more heavily than would otherwise be the case. Syllables with normal stress include all tonic syllables, and weak stress includes all neutral tone syllables (Chao, 1968: 35-37). The distinction between normal stress (i.e. tonic) and weak stress (i.e. neutral tone) is clearly phonemic, as evidenced by the following minimal pair:

dōngxī 东西 “east and west”

dōngxi 东西 “things”

There is a competing theory which states that there, in addition to contrasting stress, are three phonemic stress levels in Mandarin; strong stress, weak stress and no stress (i.e. neutral tone) (Kratochvil, 1967; 1968; Hoa, 1983). This view states that there are minimal pairs of disyllabic words consisting of two tonic syllables and distinguished by stress alone. For example (Kratochvil, 1967: 158):

'yàoshì 要是 “if”

yào'shì 要事 “important matter”

Chao (1968) admits that even tonic syllables are pronounced with different emphasis, but that there is no phonemic distinction between these levels. The disagreement on this topic stems from the fact that even native speakers of Mandarin have difficulties judging different levels of stress among tonic syllables. This is a debate I will not attempt to enter into here, as it falls far outside the scope of my study. Besides, an M.A. thesis is currently being written about this very topic by one of my fellow student at the Institute of Oriental Languages, which will probably deal with the debate much more thoroughly than I could ever hope to do here.

Although he does not recognize the distinction as phonemic, Chao (1968: 35) does state that some tonic syllables are pronounced with slightly more stress than others. In strings of tonic syllables, the last syllable gets primary stress, the first gets secondary stress, and those in between are pronounced with slightly weaker stress. For example in the phrase “rénrén dōu xiǎng qù 人人都想去 ‘everybody wants to go’” (Chao, 1968: 35):

˩Rén rén dōu xiǎng ˩'qù

This means that strings which consist of only tonic syllables will get a speech rhythm that is basically syllable timed. In strings that also contain neutral tone syllables, the regular succession of equally stressed syllables will be punctuated by points of weak stress, and the phrase will get a rhythm which is more stress timed (Chao, 1948: 26).

A different approach is taken by Monique Hoa (1983), who assumes that also tonic syllables may have strong and weak stress. Moreover, she points to a tendency towards regular alternation between strongly and weakly stressed syllables. According to her, the allocation of stress is governed by a “eurhythmic principle”,¹⁹ which states that stressed syllables must be timed at equal intervals, and a “principle of non-contiguity”²⁰ which states that no two stressed syllables may follow one another. The presence of neutral tones may actually disrupt this system, and cause longer strings of unstressed syllables than would otherwise have been the case (Hoa, 1983: 28-47). Hoa’s (1983) view is in practice the exact opposite of Chao’s (1968), because it assumes that the speech rhythm of Mandarin is basically stress-timed.

¹⁹ “Principe d’Eurythmie” (Hoa, 1983:33)

²⁰ “principe de Non-contiguïté” (Hoa, 1983: 31-33)

4 Methods and analytical categories

4.1 Consonants

Consonants were mostly detected without the aid of instruments, as they are easier to differentiate by ear alone than vowels. The number of consonant symbols used in the transcription therefore corresponds to the number of different articulations that I have been able to distinguish. In a few borderline cases where a consonant could be classified as either a plosive or a fricative, I have chosen the interpretation that results in the smallest total number of symbols. One issue still deserves some extra mention; the use of approximant symbols.

Approximants were defined in section 2.3 as “sounds that have all the physical characteristics of vowels, but function like consonants”. Since my phonetic transcription is contrasted with phonological orthography (pinyin, which describes the function of sounds in the phonological system of Mandarin), I have chosen to transcribe sounds according to their physical characteristics and not their function. This means that sounds that have all the physical characteristics of vowels will be transcribed as vowels. There are, however, two cases that are not clear cut in this respect. First, the initial /y/ in syllables like /yin/ and /ying/ is sometimes (but not always) pronounced with more friction than the following consonant. Phonetically, it appears to be somewhere in between [i] and [ʒ]. To show this distinction (which might be significant, as it appears systematically in strongly stressed syllables), I have transcribed this sound as [j], following Karlgren (1918: 7). The final approximant symbol that is used in this material is [ʋ] which denotes a labio-dental approximant. This segment has markedly less friction than the labio-dental fricative [v], and it occurs systematically in certain contexts. The approximant symbol has been used to mark this distinction.

Here follows a list of all the consonant symbols used, with references to similar consonants in other languages. The symbols follow the conventions of the IPA (2005). Where possible, examples from English have been used for comparison, but for sounds that have no analogue in English, examples from Norwegian and other languages have been used. The examples of pronunciation from all languages except Norwegian are based on those given by Ladefoged and Maddieson (1996), unless otherwise stated. I have also tried to limit myself to citing examples from languages that I have studied personally, or at least had the opportunity to listen to and verify. All aspirated consonants listed are unvoiced, and all consonants not

explicitly specified as aspirated are unaspirated. All nasals, laterals and approximants are voiced:

Plosives

- [p^h] **Aspirated bilabial**; like the “p” in English “pelvis”
- [p] **Unvoiced bilabial**; like the “p” in American “stopper”
- [b] **Voiced bilabial**; like the “b” in English “battle”
- [t^h] **Aspirated alveolar**; like the “t” in English “tell”
- [t] **Unvoiced alveolar**; like the “t” in English “nuts”
- [d] **Voiced alveolar**; like the “d” in English “double”
- [k^h] **Aspirated velar**; like the “c” in English “cold”
- [k] **Unvoiced velar**; like the “c” in English “reluctant”
- [g] **Voiced velar**; like the “g” in English “gap”
- [ʈ] **Unvoiced retroflex**; like the “rt” Eastern Norwegian pronunciation of the name “Martin”, and the final “t” of American “hurt”
- [ɖ] **Voiced retroflex**; like the “rd” in the Eastern Norwegian pronunciation of “Hardanger” (place name), and the final “d” in American “bird”
- [ʔ] **Unvoiced glottal**; like the “coughing” sound at the beginning of each syllable in English “uh-oh”

Fricatives

- [f] **Unvoiced labio-dental**; like “f” in English “fish”
- [v] **Voiced labio-dental**; like “v” in English “voice”
- [s] **Unvoiced alveolar**; like “s” in English “snake”
- [z] **Voiced alveolar**; like “z” in English “his”
- [ʃ] **Unvoiced retroflex**; somewhat like the “sh” in English “shape” and the “sk” in Norwegian “skip” (“ship”), but with the tongue flexed much further backwards

- [ʀ] **Voiced retroflex**; See previous entry. Reminiscent of the “r” sounds in the word “røre” (“to stir”) in the Northern Norwegian dialect spoken in the city of Tromsø.
- [ɕ] **Unvoiced alveolo-palatal**; like the “sh” in Japanese 新宿 Shinjuku (place name). Somewhat like “kj” in the Norwegian word “kjeller” (“basement”), spoken in the dialect of the city of Bergen.
- [ʐ] **Voiced alveolo-palatal**; See previous entry
- [x] **Unvoiced velar**; Like the “j” in Spanish “jugar” (“to play”)
- [χ] **Unvoiced uvular**; Like the “ch” in German “achtung” (“attention”)
- [ʁ] **Voiced uvular**; like the “r” in French “roue”
- [h] **Unvoiced glottal**; like the “h” in English “help”

Affricates

- [tsʰ] **Aspirated alveolar**; like the “ts” in Japanese (and English) 津波 tsunami, only more strongly aspirated
- [ts] **Unvoiced alveolar**; like the “zh” in German “zhen” (“ten”)
- [dz] **Voiced alveolar**; like the “z” in Japanese “ずっと zutto” (“by far”). See two previous entries.
- [tʂʰ] **Aspirated retroflex**; very rare consonant found in only a few languages. Similar to the “ch” in English “chance”, but with the tongue curled up and back.
- [tʂ] **Unvoiced retroflex**; see previous entry
- [dʐ] **Voiced retroflex**; see previous entry
- [tɕʰ] **Aspirated alveolo-palatal**; like the “Ch” in Japanese “千葉 Chiba” (place name). See alveolo-palatal fricatives
- [tɕ] **Unvoiced alveolo-palatal**; See previous entry
- [dʑ] **Voiced alveolo-palatal**; Like the “ji” in Japanese “実 jitsu” (“truth”)

Nasals

- [m] **Bilabial**; like the “m” in English “man”
- [ɱ] **Labio-dental**; like the “m” in Norwegian “komfyr” (“stove”)
- [n] **Alveolar**; like the “n” in English “man”
- [ɳ] **Retroflex**; like the “n” in Hindi “हिन्दी Hindī” (“the Hindi language”), but more retroflex
- [ŋ] **Velar**; like the “ng” in English “sling”

Laterals

- [l] **Alveolar**; like the “l” in English “left”

Approximants

- [ʋ] **Labio-dental**; like the “v” in Danish “lave” (“to make or do something”)
- [j] **Palatal**; like the “y” in English “yes” and the “j” in Norwegian “jippi” (“hurray”), when the words are spoken with emphasis.
- [ɭ] **Retroflex**; Like the “r” in American “rare”, but often slightly more retroflex.

Secondary articulations

- [t^w] **Rounded**; superscript symbol signifies a rounding of the lips that occurs simultaneously with the consonant being pronounced. The “p” in English “port” is rounded.

4.2 Vowels

My notational conventions for vowels are based on those of the IPA (2005), but are not identical. Some symbols have been omitted, others have been added, and some symbols are used a little differently. In this section I explain my methodological approach to classifying vowels, as well as the reasoning behind my decision to depart from the standards of the IPA.

My classification of vowels is based on instrumental analyses of vowel quality. The horizontal and vertical positions of the tongue determine the values of the first two formants, F1 and F2. A convenient and commonly used method of charting the vowels of a particular language or speaker is to plot the values of F1 and F2 in a graph, where the horizontal axis represents F2 and the vertical axis represents F1. If the chart has the proper dimensions, the positions of each vowel will correspond remarkably well to that vowel’s place of articulation

in the oral cavity (Ladefoged, 2001: pp. 39-46). However, the measuring of vowel formants is no exact science. The tongue is constantly moving around, and usually does not hold one position for the entire duration of a vowel segment. Any vowel is therefore more of a flow, or movement, than a fixed state, and formant frequencies towards the end of a vowel may be considerably different from those towards the beginning. A common way of going about charting the formants is to measure them at the so-called **vowel target**. The vowel target is usually at the centre of the segment, and is therefore assumed to be the point in time where the vowel is least influenced by surrounding segments (Harrington, 2010).

I have charted the vowels of both speakers in the way shown in figure 4.1. All vowels were measured at the vowel target, using the automatic formant measuring tool in the Praat software. The charts were based on measurements of carefully pronounced vowels, whose articulatory qualities were clearly distinguishable by ear. The formants frequencies of all other vowels in the material were then compared to the chart, and classified on the basis of this comparison. This was done to ensure a certain consistency in the use of the vowel symbols, and not to make general statements about the nature of vowel formants of Mandarin. However, it is still interesting to note that the findings presented in figure 4.1 correspond well to those made by J. M. Howie (1973), who's comprehensive study is still one of the most important accounts of the acoustic qualities of Mandarin Chinese.

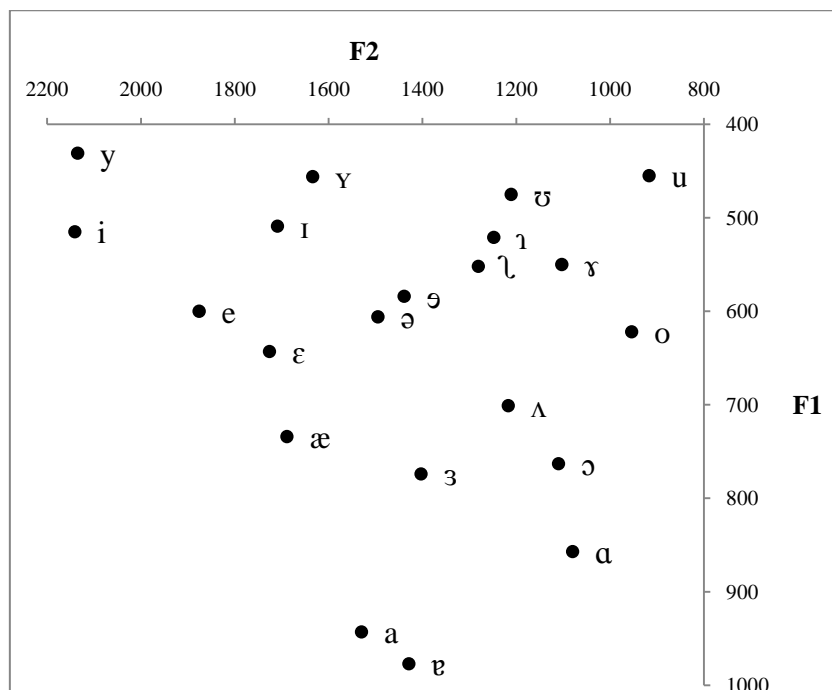


Figure 4.1: The vowels of speaker A

In practice, the formant frequencies of vowels are never the exact same as those in figure 4.1, and it is often difficult to classify a vowel as belonging to one category instead of another. The number of categories distinguished is also a purely practical concern. The IPA (2005) distinguishes a total of 28 symbols, of which there are nine that I do not use. They have been omitted either because the corresponding vowels do not occur in my material (there are, for example, no non-close front rounded vowels like [ø] or [œ]), or because their inclusion would not have led to any greater analytical clarity. As an example of the latter, the symbol [u] could have been used to transcribe one (or possibly two) instances of a very close back vowel. Because this vowel occurs so infrequently, and because it is acoustically so close to [ɤ] that the difference is practically inaudible, I decided that to classify it as a separate category would lead to confusion rather than clarity.

Two of the categories I use are not taken from the IPA at all, but are phonetic symbols intended especially for the transcription of Mandarin Chinese. These two symbols, [ɿ] and [ʉ], were introduced by Bernhard Karlgren (1918) and stand for what he calls the **apical** vowels. Because there is no general agreement on how to transcribe these vowels using the normal conventions of the IPA, Karlgren's symbols have become a common staple in Chinese linguistics. The symbol [ʉ] is used in my transcription to refer specifically to a rounded vowel, as opposed to the standard IPA usage which represents an unrounded one. Finally, the symbol [ɐ] is used to denote a fully open vowel, where the IPA symbol is used for a near-open one.

Based on the acoustic data from figure 4.1, the system of vowel symbols used in my transcription can be described by the diagram in figure 4.2. Symbols to the right of each bullet point represent rounded vowels, and symbols to the left represent unrounded vowels. The exceptions are the apical vowels [ɿ] and [ʉ]; they are both unrounded, and the symbol to the right represents an r-colored articulation. Note that this diagram is slightly different from the corresponding one used by the IPA.

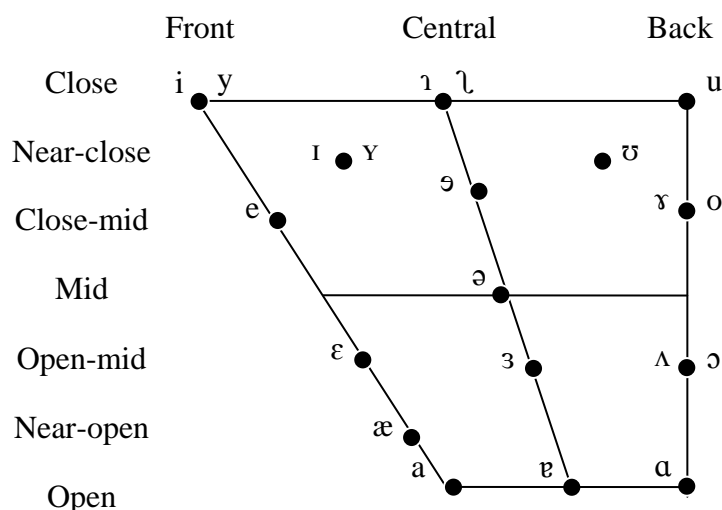


Figure 4.2: Diagram of the vowel space

Now follows a list of the vowel symbols used, each compared to similar vowels in English and other languages. Once again, the descriptions of the speech sounds of all languages except Norwegian are based on Ladefoged & Maddieson (1996).

Close vowels

- [i]; Like English “beat”
- [ɪ]; Like English “bit”
- [y]; Like German “müde” (“sleepy”) and French “Chute” (“fall”)
- [ʏ]; Somewhat like the first vowel in Swedish “nytta” (“to use”)
- [ɪ]; Pronounced like a [z], only with the space between the tongue and the gums expanded just enough that it is not a fricative consonant. Chao (1968: xxiii) gives the phonetic value of this vowel as [z]
- [ʌ]; Is similar to [ɪ], only with the tongue in a retroflex position. Chao (1968: xxiii) gives the phonetic value of this vowel as [z]
- [u]; Like Norwegian “sto” (“stood”) and the first vowel in Spanish “pujar” (“to struggle”)
- [ʊ]; Like English “good”

Mid vowels

- [ε] Like the first vowel in English “elemental”

- [ə] Like the last vowel in English “Texas”
- [ɘ] Like [ə], only with the tongue more tense and in a higher position.
- [ʊ] Like an “o”, only with spread lips
- [ɜ] Like English “bird”
- [ʌ] Like English “but” or “stub”
- [e] Like Norwegian “Emil” (given name)
- [o] Like Norwegian “stå” (“to stand”)
- [ɔ] Like English “caught”

Open vowels

- [ɑ] Like English “bard”
- [æ] Like the syllabic vowel in Norwegian “jeg” (“I”)
- [a] Like English “bad”
- [ɐ] Like the first vowel in the English diphthong “bite”

Special vowel articulations

- [̃] Diacritic signifies that the vowel is nasalized. Like French “sans” (“without”)
- [ɹ] Diacritic signifies that the vowel is r-colored. Like American “bird”
- [ɜ:] Diacritic signifies that the vowel is particularly long: like English “meal” when compared to “mill”
- [³] Vowels in superscript denote very short vowel segments that are part of a diphthong or triphthong. They are more like temporary states or points of transition than fully formed vowels.

4.3 Tone, stress and intonation

When spoken in isolation, the four tones of Mandarin can be conveniently transcribed using tone letters, as I did in section 3.2. However, transcribing tones in connected speech is a somewhat different matter. Because tone, stress and intonation exert great influence on each

other, I will argue that the entire complex of supra-segmental features in a phrase must be represented through the simultaneous use of stress marks and tone letters.

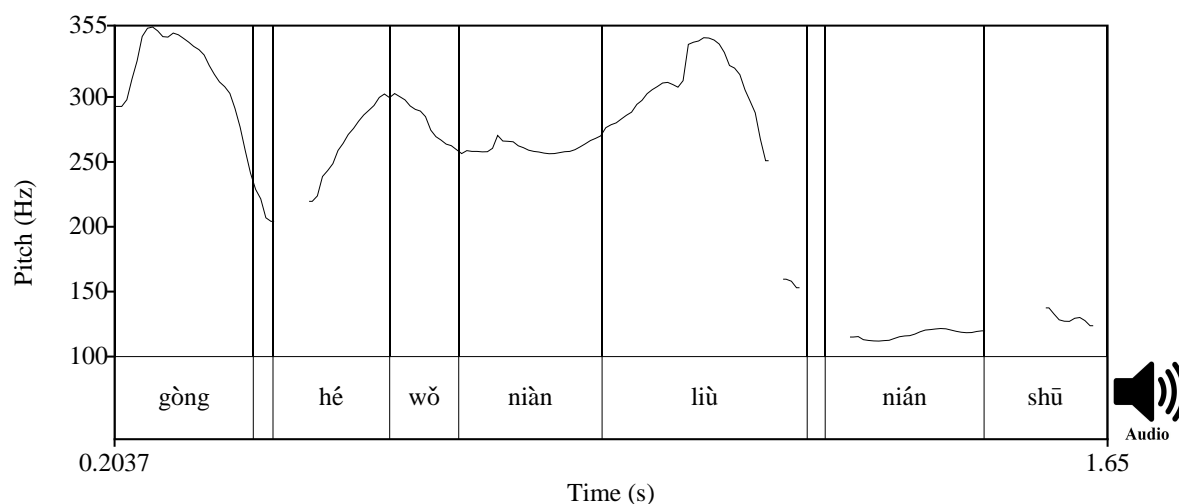
The first thing that must be made clear is what the tone letters' five levels actually stand for. We know that [5] is a high tone and [1] is a low tone, but high and low in comparison to what? Y. R. Chao (2006b), who invented the tone letters, addressed this particular question himself. The tone letters, he proposed, do not denote absolute values, but represent relative pitch height only. More specifically, the tone letters denote movement inside the individual voice range of any single person. Different voices cover different ranges of the tonal spectrum, and the levels [5] and [1] represent the highest and lowest pitch levels that an individual speaker uses in normal speech. In his own words, "what would be the low tone for a soprano is actually higher in pitch than the high tone of a tenor" (Chao, 2006b: 204).

However, this description is insufficient to account for the behavior of tones in the present material; the range of speaker A's voice stretches from about 50 Hz to about 400 Hz, yet no individual tone covers a range of more than 150 Hz. Thus, it is obvious that any individual tone only covers a fraction of the speaker's entire register. How, then, can one use the tone letters in a meaningful way to transcribe connected speech? One option I considered in an earlier phase of my work is that the tone letters' levels [5] and [1] may be used to stand for the highest and lowest pitch values of a single intonation phrase. In this way, a high tone would be high in the context of the phrase it is part of, but might well be lower than a low tone of another phrase. However, closer inspection of pitch contours made this method also seem unsatisfactory. I will illustrate the problem with an example:

8. B: 共合我念六年书
Gònghé wǒ niàn liù nián shū
"All in all I've gone to school six years."

This phrase has a variety of different tones. If one assumes that the levels [1] and [5] denote the roof and floor of the pitch range in this particular phrase, it should mean that the tone of /wǒ/ dips lower than those of /hé/, /nián/ and /shū/, and that /gòng/, /niàn/ and /liù/ extend all the way from roof to floor. When one takes a look at the actual intonation contour (figure 4.3),

however, it quickly becomes apparent that this is not the case.



Apparently, /gòng/ is higher pitched than /liù/, but does not cover as wide a register. /hé/ and /nián/, which are supposed to be the same tone, are in completely different keys. And, perhaps most surprising of all, the nominally low tone of /wǒ/ is much higher in pitch than that of /shū/, which is supposed to be a high tone. Apparently, if I were to use the tone letters as representations of the ceiling and floor of the pitch in any given phrase, most of the tones would not even register on the scale at all. To give an accurate representation of tones in this way would require a tone-letter with much more than just five tiers.

To try and solve this problem and find a practically satisfying way to use the tone letters, I follow John Laver (1997: 456-460) in separating between the two concepts of **pitch range** and **pitch span**. Pitch range is the total range of one speaker's voice in normal conversation, and varies from person to person. Thus, what Chao (2006b: 204) talks about when he describes the difference between a tenor and a soprano, is a difference of pitch range. The pitch span, on the other hand, is the range of the speaker's voice at any given moment of the utterance. A phrase not only utilizes a small part of the speakers pitch range, but also has a pitch span that expands and contracts with each syllable.

What causes the changes in the pitch span? One thing which can be observed in figure 4.3 is that the pitch goes down towards the end of the phrase. This is a result of the sentence melody, or intonation. The tendency of the intonation contour to fall towards the end of a phrase is a well known and documented phenomenon, not only in Chinese but in many languages (Laver 1994: 459; Shen, 1989: 9-27). In practice, it means that tones at the end of a phrase will

generally be pronounced in a lower key, and that their contours will be less defined (Wang & Lin, 2003).

But the influence of intonation does not explain the whole behavior of the tones in figure 4.3. It also seems that the pitch span is greatly widened at the points where the syllables /gòng/ and /liù/ occur. This can be explained by the fact that these two syllables have strong stress. Stressed syllables are pronounced with greater pitch movement and more exaggerated tone contours than unstressed ones. Syllables with high tones are pronounced in a higher average pitch when stressed, and syllables with low tones are pronounced with a correspondingly lower average pitch. In other words, syllable stress causes the pitch span to expand and contract; strong stress expands it, while weak stress contracts it.²¹ This means that stress and tone are mutually dependent on each other; a syllable is stressed by exaggerating the contours of the tone on that syllable, and tones on stressed syllables are pronounced with more exaggerated contours.

If we accept that syllable tones are influenced by stress and intonation, it is possible to see that the actual shapes of the tones are more or less as described in section 3.2. The exception is /niàn/, which has apparently been reduced to a neutral tone. Using phonetic symbols, the complex of supra-segmental features can then be represented in the following way. The syllable /gòng/ has been assigned a secondary stress mark which shows that the “size” of its tone is relatively larger than the other syllables, and /liù/ has been assigned a primary stress mark which indicates that its syllable tone is even more exaggerated and that it marks the turning point of the global intonation contour. To represent rhythm and timing a lower ligature tie has been used to connect syllables that are pronounced particularly close together:

Gònghé wǒ niàn liù nián shū

[₁kuŋ˥˥ xɑː˥˥ ˥˥ə˥˥ n'ɛn˥˥ ˥˥lɿəʊ˥˥ n'ɛŋ˥˥ ʃuː˥˥]

In practice, primary stress will determine the highest point of the intonation contour, with one notable exception: If primary stress falls to the end of a phrase it will not mark the highest pitch in the phrase, but will instead negate the general trend of falling intonation.²² There are therefore two typical kinds of intonation patterns that appear in the present material. The two

²¹ The phonetic characteristics of stressed syllables are discussed in greater depth in chapter 5. The description given here is based on the conclusions arrived at in that chapter.

²² This is the case regardless of whether the stressed tone has a high tone (tone 1, tone 2, tone 4) or a low tone (tone 3). When tone 3 is this strongly stressed, it usually retains its ideal falling-rising contour, and therefore in practice becomes a high tone.

kinds of patterns are illustrated through the use of two example phrases in figure 4.4. In the transcription, intonation patterns can be identified as belonging to either one or the other of these categories from the placement of the primary stress mark.

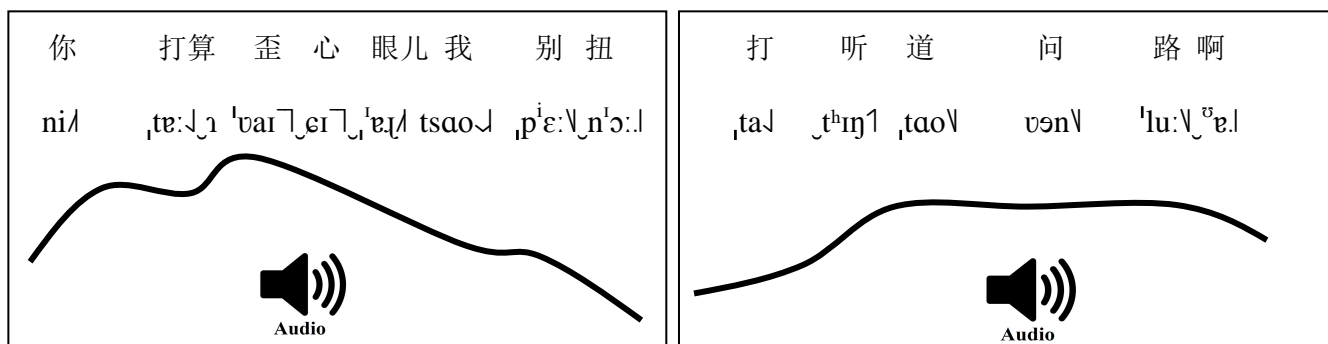


Figure 4.4: Simlified representation of two kinds of intonation contours

Because the tone letters must be small enough to be used in conjunction with other written symbols, and because they are made up entirely of straight lines, it follows that they still represent an abstraction and a simplification of the tonal contours shown in a pitch analysis. Because of this, I will argue that the tone letters are poorly suited to an impressionistic transcription. Instead, I have found that a more practical way of using is to give each of the four Mandarin tones a set of variants. The variants represent general trends that occur throughout the material, rather than the specific tonal contour of each individual syllable. For example, Mandarin tone 1 is generally transcribed [55]. Under certain circumstances, it has a tendency to change into a rising tone, which has been transcribed [45]. This reflects the fact that it is (usually) higher and less sharply rising than Mandarin tone 2, which is transcribed [35]. Though the [45] variant may sometimes rise just as sharply as tone 2, it is still transcribed as [45]. In a similar vein, I have chosen to transcribe the rising variant of tone 3 (described in section 3.2 as being homophonous with tone 2) as [14] instead of [35]. This allows the reader to immediately see whether the tone is a tone 2 or a changed tone 3:

着火 zháo huǒ "to catch fire": [1 4] → [1 4]

找火 zhǎo huǒ "to look for fire": [4 4] → [1 4]

Neutral tones are represented using the variant tone letter I introduced in section 3.3. The pitch height is determined relative to the preceding syllable, or to the following syllable if the neutral tone occurs phrase-initially. In the material I have examined, syllables with the neutral

Now follows a list of the tone letters used in my transcription of Xué Wàiyǔ:

[55] [ʔ]	Typical form of tone 1. High and level; sometimes slightly convex.
[45] [1]	High and slightly rising variant of tone 1; usually not as sharply rising as tone 2.
[35] [ʌ]	Typical form of tone 2. High and quite sharply rising. The rise becomes less sharp in environments of falling intonation, especially towards the end of an intonation phrase.
[214] [↘]	The ideal form of tone 3. Low-falling and then high-rising; in the present material, it appears only in syllables of relatively long duration.
[21] [ɿ]	Low falling variant of tone 3; in practice it is the first half of [214].
[14] [ʌ]	Rising variant of tone 3; in this material it is usually, but not always, lower than the [35] of tone 2.
[212] [↘]	An intermediate variant of [214] and [21]; it rises towards the end, but not as high as [214]
[51] [ʌ]	Typical form of tone 4. High-low falling tone; falls very sharply.

[1] [˩]	Very low, short and level tone.
[2] [˨˩]	Low, short and level, but not as low as [1]
[3] [˧]	Middle, short and level.
[4] [˥]	As high, or almost as high as level [5] in the preceding tone.

[5] [l]	Higher than level [5] in the preceding tone.
[42] [N]	Falling neutral tone; not as sharply falling as the [51] of tone 4.
[24] [4]	Rising neutral tone; usually only very moderately rising, and is not as pronounced as either of the rising tones that occur in tonic syllables.

Other supra-segmental features

[^h kuŋ]	Primary stress: the most heavily stressed syllable in a phrase, and the turning point of the intonation contour
[_h kuŋ]	Secondary stress: Any syllable which has more stress than all adjacent syllables, but which is not the most heavily stressed syllable in a phrase
[kuŋ_kuŋ]	Lower ligature tie signifies that two syllables are pronounced close together.

5 The phonetic basis of stress

While there is considerable disagreement on the status of stress in Mandarin phonology, most scholars seem to agree that it is difficult to judge which syllables are stressed and which are unstressed (Chen, 2000: p. 288; Duanmu, 2000: pp. 142-144; Hoa, 1983: p. 7; Chao, 1968: p. 38). Because even native speakers have this problem it was obvious that I, not a native speaker at all, would have difficulties giving an accurate account of stress in my transcription of Xué Wàiyǔ. To make matters more difficult, there does not seem to ever have been any general agreement on exactly how stress is reflected in phonetic parameters. Thus, even with the aid of instruments like Praat, there is no simple way to separate a stressed syllable from an unstressed one.

There have been several views on the acoustic basis of stress. Some languages, for example English, mark stress by vowel reduction and other segmental features, but for the most part stress is seen as related to the three acoustic parameters of loudness, pitch and duration (Laver, 1997: 450, 511-517). Which of these is the most important correlate of stress is not agreed upon. It seems that earlier works on phonetics (Abercrombie, 1968: p. 66; Luo, 1957: p. 136) mostly assume loudness to be of greatest importance, but later experimental studies have shown that this might not be the case. H. A. Spešnev (1959) argues that loudness is of no consequence to perceived stress in Mandarin, and that duration is the most reliable correlate. Paul Kratochvil (1974) takes a different approach, and suggests that judgment on stress corresponds best to a compound value of all three parameters.

It has also been proposed that stress has no definite acoustic basis. Because duration, loudness and pitch are all parameters whose primary function lies in other parts of the language (in Chinese languages, for example, pitch is primarily used to produce tones), stress is a kind of “parasitic” phenomenon. Stress, therefore, has no universal acoustic correlate, but changes between situations and languages (Hayes, 1995: pp. 5-8). In this chapter I attempt to resolve some of the difficulties this poses to my own work, and ask the question:

What are the acoustic correlates of stress for the speakers in Xué Wàiyǔ?

5.1 Theoretical background

In the nineteen fifties, experiments on syllable stress, both in English (Fry, 1955) and in Mandarin Chinese (Spesnev, 1959), showed that there was little correlation between stress perception and loudness.²³ This went contrary to the common assumption in earlier literature²⁴, which figured stressed syllables as being primarily cued by loudness. Spesnev's (1959) experiments with Mandarin speakers showed that, while stressed syllables were not necessarily louder than unstressed ones, they were uniformly longer in duration. This led him to conclude that "Word stress actually influences the duration of the stressed syllable, and not its loudness" (p. 34).²⁵

That duration is the primary acoustic correlate of stress is not, however, commonly accepted as a fact. In English specifically, it is now commonly claimed that stress is differentiated mainly on the basis of pitch height. This is because stressed syllables are pronounced with greater respiratory forces; with more air pushed out of the lungs. Greater respiratory force may lead to an increase in both loudness and duration, but its primary effect is to raise the pitch level (Ladefoged, 2001: pp. 22-24). This position has become well documented through a long line of acoustic experiments (Chen, 2000: p. 291). It has been argued that the same is the case in Mandarin, and that this accounts for native speakers' difficulties in perceiving stress. Duanmu (2000: pp. 142-144), suggests that pitch height cannot be freely used to mark stress in Mandarin, because it is already "used up" by the lexical tones. In English, on the other hand, pitch height serves no other phonemic purpose, and stressed syllables can therefore be assigned a special pitch contour to distinguish them from other syllables.

The validity of this argument partially depends on whether or not one accepts the notion that pitch height is the most important correlate of stress in all languages. Hayes (1995: pp. 5-8) argues that stress has no specific correlate, but that it may be realized differently from language to language. He proposes that languages where vowel duration is in itself phonemic (like English, which distinguishes long and short vowels) will not use duration to cue stress because that parameter is already used up for another phonemic feature. Similarly, languages where pitch is used for other phonemic features will not use pitch to cue stress. This would confirm Spesnev's (1959) view, in that it makes it entirely possible for Mandarin to mark syllable stress through an increase in duration.

²³ Actually, Spesnev (1959) stated that there was no correlation what so ever between loudness and stress: "元音的音强实际上跟重音没有任何关系" (Spesnev, 1959: p. 34)

²⁴ For example in Abercrombie (1968) and Luo (1957)

²⁵ "词的重音实际上是以重读音节的音长为转移, 而不是以重读音节的音强为转移" (Spesnev, 1959: p. 34)

A different approach to finding the acoustic correlate of perceived stress was made by Paul Kratochvil (1967; 1975; 1998). Through examinations of large numbers of disyllabic expressions in corpora of naturally occurring speech, and subsequent perceptual tests carried out on native speakers, he has concluded that the closest acoustic correlate to syllable stress is something he calls “syllabic volume”. Syllabic volume takes into account both pitch height, duration and loudness; more specifically it is the equivalent of an equation (Kratochvil 1974: p. 436):

$$\frac{\text{duration} \times \text{loudness} \times F_0}{100\,000} = \text{Syllable volume.}$$

Duration is here defined as the length of the tone carrying part of the syllable measured in milliseconds (ms), and loudness is the mean Db value of the tone carrying segments. F0 is defined as the mean F0 of the phrase to which the syllable belongs, plus the deviation of the mean F0 of the syllable from that baseline. In the case of tones 1, 2 and 4, this is simply the equivalent of the mean F0 of the syllable, because they deviate upwards from the baseline. With tone 3, however, the F0 value usually deviates downwards. In this case, the F0 value is calculated in the following way, where “a” stands for the mean F0 of the phrase and “b” stands for the mean F0 of the syllable (Kratochvil 1974: p. 436):

$$(a - b) + a$$

This way of calculating F0 reflects a specific view on how pitch height can be used to mark both stress and tones simultaneously. In Kratochvil’s (1974) model, stressed syllables exaggerate their tonal contours and F0 values. Thus, a syllable with a low tone becomes stressed, not by raising the overall F0 value, but by lowering it. As Susan Shen (1989) puts it, “the rising tones rise higher and falling tones drift lower. The falling-rising tones dip lower and the level tones become longer; the upward and downward slopes are steeper” (p. 60).

5.2 Examining the data

Since I have had no opportunity to conduct perception tests on native speakers of Chinese, I have instead measured different acoustic properties that have been reported as correlates of stress, and attempted to find configurations which coincide with expected stress patterns. As I mentioned in section 3.3, there does not seem to be any disagreement that Mandarin neutral tone syllables are unstressed. A word or an expression that consists of a tonic syllable followed by a neutral tone syllable is therefore generally agreed to have strong stress on the

first syllable and weak stress on the second. To investigate the acoustic correlates of stress, I have measured the duration, loudness and F0 of a number of expressions containing the neutral tone. The aim of this was to identify any systematic differences between the stressed and unstressed syllables.

Both the duration and pitch contours of neutral tone syllables are sometimes affected in phrase-final position. Also, what is to be classified as a neutral tone is sometimes ambiguous, since a neutral tone has no readily identifiable, inherent characteristics. Since my aim is to study expressions that have entirely unambiguous stress patterns, I attempted to pick out only the most clear cut examples from my material. To ensure that the study was broad enough to be applicable to the rest of my material I wanted all four tones represented in the selection. To make comparison easier, I also wanted the segmental structure of the expressions to be as similar as possible. To fulfill these requirements, expressions were chosen from both speakers of Xué Wàiyǔ based on the following criteria:

1. Background noise must be limited, and sound must be clear enough for the computer to perform accurate measurements.
2. The expression must be disyllabic; consisting of one tonic and one atonic syllable.
3. All of the four tones must be represented in the final selection.
4. The expression must contain a lexically determined neutral tone, (i.e. it must be an expression like *dōngxi* 东西 “thing” or *dìdao* 地道 “genuine”).
5. None of the syllables in the expression can have a final consonant.
6. The expression cannot occur in phrase-final position

As it turned out, not all of the criteria could be satisfied. Neutral tone expressions containing tones 1 and 2 were relatively rare, and therefore some syllables were chosen despite the fact that they contain a final consonant. One expression is very heavily represented in the set, because it occurs particularly frequently in the material.

The selected expressions were segmented in Praat, and the vowel segments were measured. In the syllables that had final consonants, these were not measured. This to ensure that they would not get an “unfair advantage” as far as duration is concerned. The values “F0 mean” and “Db mean” are the average F0 and Db in the vowel segments, and “duration” is the total duration of the vowels in milliseconds. F0 is measured as the deviation from the baseline F0 of the phrase, using the method devised by Kratochvil (1974) and described in the previous section.

If Mandarin syllables are indeed stressed by exaggerating contours of the syllable tone, as was suggested in the previous section, then it seems that stress should be more closely related to the absolute high (or low) point of F0 in the syllable than to its average. It is possible for a tone to rise sharply to a very high tone and still have an only moderate average pitch height. For example, if a tone 2 is followed by a tone 1, and the tone 2 begins lower and rises higher, the tone 1 might well end up with the highest mean F0. To account for this, I have measured the single point in each syllable that shows the greatest F0 deviation from the baseline of the phrase. I have classified the results of these measurements as “F0 peak”, to distinguish them from F0 mean. The results of my measurements are given in table 5.1.

Looking at the table, the first thing to note is that in every case duration is greater in the stressed syllable. This corresponds to what Spesnev (1959) found in his experiments. In most cases, this difference in duration is considerable. It would seem that other factors than stress can cause an increase in duration; diphthongs are generally longer than monophthongs, and syllables with a final consonant have shorter vowels than those with an open final. The apical vowel [ɿ] also seems to be shorter than the other vowels. However, none of these factors actually cause the stressed syllable to become shorter than the unstressed ones.

As far as loudness is concerned, the results are also very similar to Spesnev's. In the expression *màozi*, loudness is consistently lower in the unstressed syllables. However, this might be a result of other factors than stress. It seems very likely that the apical vowel [ɿ] in [mao tsɿ] is inherently less loud than other vowels, because it is very close to being a consonant. This becomes apparent if one tries to shout [ɿ] off the top of one's lungs, and then tries to shout [a] with the same force. The [a] will invariably be louder than the [ɿ]. It is not a stretch to assume that this same mechanism is at work in the data I have analyzed here. In the other seven words on the list, two show the unstressed syllable to be louder than the stressed syllable, and two show the syllables to be equally loud. If we discount *màozi*, this means that greater loudness corresponds with stronger stress in less than half of the cases.

Speaker	Expression	Syllable	Duration	F0 mean	F0 peak	Db mean
A	màozǐ 帽子 (171)	[mao˥]	188	327	358	79
		[dzɿ˨]	42	210	214	70
A	màozǐ 帽子 (173)	[mao˥]	170	126	285	78
		[dzɿ˨]	47	213	255	68
A	màozǐ 帽子 (175)	[mao˥]	187	205	296	83
		[tsə˨]	46	253	256	78
B	màozǐ 帽子 (154)	[mao˥]	280	260	357	82
		[dzɿ˨]	68	149	151	68
B	màozǐ 帽子 (156)	[mao˥]	270	247	330	78
		[dzɿ˨]	54	136	136	62
B	màozǐ 帽子 (168)	[mao˥]	203	278	356	85
		[dzɿ˨]	45	230	232	80
B	màozǐ 帽子 (174)	[maː˥]	129	226	232	83
		[dzɿ˨]	51	191	194	77
A	zhuōzi 桌子 (115)	[tʂuo˥˥]	160	225	234	82
		[tsə˨]	43	193	198	79
A	tāmen 他们 (3)	[tə˥˥]	114	239	251	77
		[mən˨]	61	246	258	77
B	zhēnde 真的 (14)	[tʂən˥˥]	78	365	394	72
		[də˨]	77	244	324	80
A	hánchen 寒碛 (205)	[xæn˥]	79	179	187	78
		[tʂɛ˨]	44	193	194	65
A	shénme 什么 (35)	[ʂəm˥]	63	181	200	74
		[mə˨]	46	243	256	75
B	shénme 什么 (140)	[ʂəm˥]	53	265	304	82
		[mə˨]	21	390	391	80
A	xiěde 写的 (15)	[ɕe˥˥]	112	297	307	77
		[də˨]	31	240	251	77

Table 5.1: Acoustic measurements of neutral tone expressions

Pitch is a slightly more complicated matter. Generally speaking, the stressed syllables have higher values of both F0 mean and F0 peak. The exception to this lies in the words where the first syllable has a rising tone. In the word tāmen (3), which in this case has the rising variant [45] of tone 1, the difference is small. The answer to why the unstressed syllable gets a higher pitch may lay in the global intonation contour of the phrase from which the expression is

taken. The [45] variant usually occurs in an environment where the global pitch is rising, and looking at this particular phrase, we see that this is indeed the case here. Tāmen (3) is placed in the upward intonation slope that occurs at the beginning of a phrase. In these cases it is actually expected that a neutral tone is higher than the preceding syllable.

With màozi (173) and màozi (175) Mean F0 is also higher in the unstressed syllable. These words also occur at the beginning of a phrase, where the global intonation contour is rising. Because of this the unstressed syllables of these two expressions both gain a pitch level of [4], instead of the [1] that we would expect after a tone 4²⁶. On the other hand, Peak F0 is consistently higher in the stressed syllables, regardless of mean F0. This seems to confirm the suggestion that a syllable is stressed by exaggerating its tonal contour, and not by pronouncing it at a generally elevated pitch.

The words shénme and hánchen present a somewhat different problem. For one, the difference in pitch is much greater; the unstressed syllable has a much higher pitch than the stressed one. As in tāmen, even F0 peak is higher in the unstressed syllable, which runs contrary to the tendency displayed by the rest of the data. The answer to why these two expressions distinguish themselves from the rest seems to lie in the fact that much of the tone 2 pitch movement falls on the final consonant (see figure 5.1).

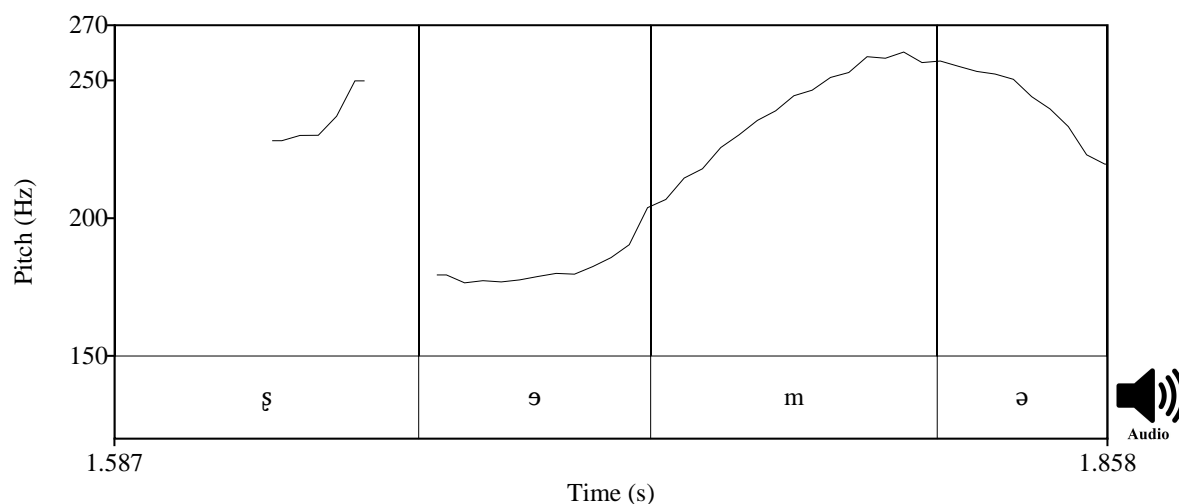


Figure 5.1: Pitch contour of shénme (35)

²⁶ See section 3.2.

A similar case is the expression *zhēnde*, where the pitch contour of the tone also continues onto the final consonant. The difference is that, since the tone of [tʂən̄] is level, the syllable's F0 values are not negatively affected by excluding the consonant from the measurements. A rising tone on the other hand, reaches its apex after the onset of the consonant, and this greatly diminishes both F0 mean and F0 peak. The latter of these two values is somewhat closer to the expected result, but it does not remedy the problem entirely. This may suggest a weakness in the method of measuring vowel segments only, as final consonants also seem to contribute a significant part of the tone. On the background of a comprehensive acoustical study of Mandarin tones, J. M. Howie (1973) defined the tone carrying segments as "the syllabic vowel and any voiced segment that may follow it in the syllable" (p. 219). Based on the results obtained here, that definition may also be the best suited for measuring acoustic correlates of stress.

From what the findings so far seem to indicate, duration is the only acoustic quality that is consistently greater in stressed syllables. Pitch height appears to be closely related to stress, but is also subject to modification by various other factors such as the global intonation contour. In addition, it seems that tone is sometimes carried by a syllable-final consonant as well as by the vowel. This means that the final consonants should be included when measuring F0. F0 peak proved more consistent with the expected stress patterns than F0 mean. In some cases, it corresponded to the expectations where F0 mean did not, and in all other cases it articulated the difference more clearly. Loudness seems, from this data, to be less directly related to stress, and more susceptible to influence by other factors such as vowel quality.

To test whether these results may be used to distinguish between syllables that are not in the neutral tone, I have conducted the same measurements on a very modest selection of disyllabic expressions consisting of two tonic syllables. As I mentioned in chapter 3, the distribution of stress among tonic syllables is a controversial topic in Mandarin phonology. It is therefore not possible to find any expressions without the neutral tone that have completely unambiguous stress patterns. It is possible, however, to find some examples where at least most scholars seem to be in agreement. Chao (1968: p 35) proposes that all disyllabic expressions have stronger stress on the last syllable, a view which several others have also adopted (Chen, 2000: p. 289). Monique Hoa (1983: pp. 194-195, pp. 226-227) proposes a kind of modification to this view. She acknowledges that most disyllabic expressions have

stronger stress on the last syllable, but that certain constructions are reversed to have stronger stress on the first syllable. To test the above findings on as unambiguous a material as possible, I have chosen only expressions that would be expected to have stronger stress on the second syllable according to both Hoa (1983) and Chao (1968). The findings are displayed in table 1.2.

Speaker	Compound	Syllable	Duration	F0 mean	F0 peak	Db mean
B	chàdiǎnr 差点儿 (32)	[tʂʰe:ʋ]	132	212	249	82
		[tiǎŋ]	173	208	229	82
A	bàodǔr 爆肚 (143)	[pɑʊ]	148	222	243	85
		[tuǎŋ]	203	240	283	81
A	zàijiàn 再见 (115)	[tsæiʋ]	134	182	214	82
		[tɕiɛnʋ]	153	206	244	82
A	duìhuà 对话 (109)	[tʰɛiʋ]	113	204	224	78
		[χʰe:ʋ]	112	206	221	78
B	duìhuà 对话 (110)	[tʰɛiʋ]	100	234	248	84
		[χʰe:ʋ]	146	233	260	82

Table 1.2: Acoustic measurements of words two tonic syllables

By and large, this investigation shows the same results as that of the neutral tone expressions, but the distinctions are less clear cut. First of all, loudness seems to be even more unpredictable here than in expressions containing the neutral tone. Not in a single case is the mean Db count higher in the strongly stressed syllable. This strongly supports the conclusion that the connection between loudness and stress is, at best, very loose. Duration is also here longer in the strongly stressed syllables. Only in one case is the weakly stressed syllable longer in duration, and then only by one millisecond. This particular case, duìhuà (109), deserves some closer inspection in regard to the other measurements as well. In this expression, peak F0 is also greater in the first syllable than in the second, defying the pre-stated expectations. One could argue that, in this case, the expectations do not hold and the first syllable actually has stronger stress. However, it could also be argued that the measurements made here are not sensitive enough to uncover the real stress pattern.

On closer inspection of the pitch contour, the tone of the second syllable in duìhuà (109) falls lower, and therefore covers a greater range (see figure 5.2). Measuring only the deviation of F0 from the baseline is not sufficient for detecting this factor. Moreover, The F0 peak output

is different depending on where the measurements are made. Remember Howie's (1973: 219) definition of the tone carrying segments as the nuclear vowel and any segments following it. This definition excludes on-glide vowels, such as the [ʊ] in [tʰʊɪ]. If that vowel is excluded from the measurements, the F0 peak of the first syllable is lower than that of the second. All in all, it seems that these factors make *duìhuà* (109) a borderline case, where stress is not easily perceived or measured.

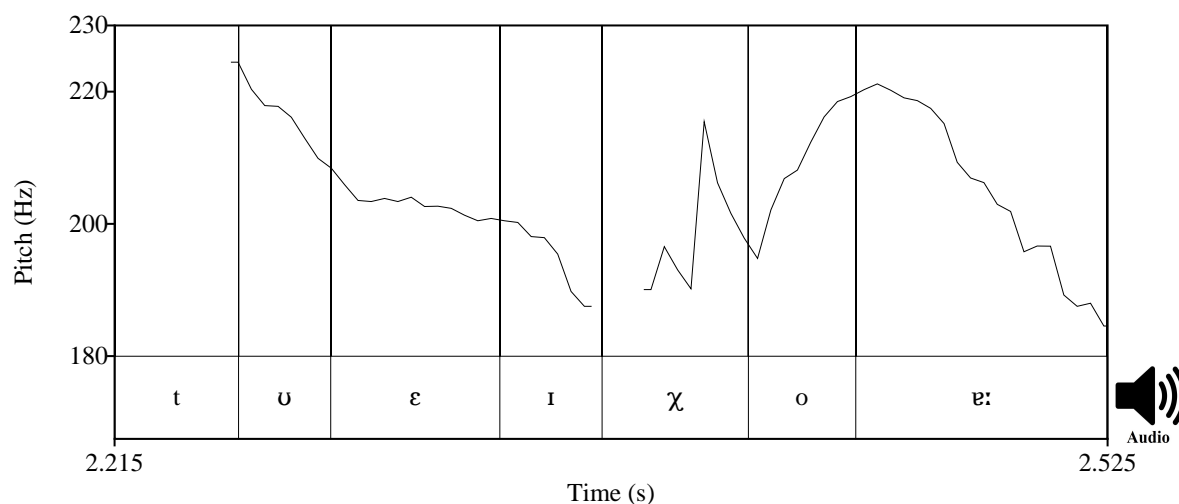


Figure 5.2: Segmented pitch contour of *duìhuà* (109)

5.3 Conclusions

In this chapter I have selected some words and expressions uttered by both speakers in *Xué Wàiyǔ*, and measured them for the acoustic qualities most commonly associated with linguistic stress. Expressions of the kind chosen are unanimously treated in the phonological literature as having a structure of stressed-unstressed, and the aim of this investigation was to see which of the measured acoustic parameters corresponded best to that expectation. The results showed the duration of the vowel segments to always be greater in the stressed syllables. Loudness, on the other hand, could not be shown to have any clear connection to stress.

F0 seems to be related to stress, but is more difficult to measure accurately and consistently. On the one hand, the F0 mean of stressed syllables is mostly higher than that of unstressed syllables, meaning that they are pronounced at a generally elevated pitch (or lowered pitch, in the case of tone 3). On the other hand, F0 peak corresponds even better to stress than F0 mean. This would seem to indicate that stress is marked primarily by increasing the range of the

tones, rather than modifying their overall pitch level. A few of the measured syllables deviated greatly from the expectations because much of the tone contour was carried by the final consonant. This indicates that final consonants should be taken into consideration when attempting to judge stress.

The same measurements were made on a selection of words and expressions with less clearly defined stress patterns. For the most part, the results corresponded to those obtained above, but the results were less clear cut. There was also one borderline case where stress was difficult to distinguish clearly.

The results I have obtained here are similar to those of H. R. Spesnev (1959). Kratochvil (1974) has claimed that loudness is related to stress, but his formula for calculating “syllabic volume” is arrived at on a basis of a much more extensive material than mine. It may therefore be more valid for Mandarin in general, but does not seem to be as relevant for the specific dialogue that I study here. This may be because the material studied by Kratochvil (1974) is specifically in the Beijing dialect, while mine is not (see chapter 1).

The results of this investigation have allowed me to arrive at a method for measuring stress in Xué Wàiyǔ. I have defined a strongly stressed syllable as having greater duration, a wider pitch span, and a higher (or lower) mean F0 than a weakly stressed one. A syllable can “compensate” for the lack of one of these qualities by increasing the others, e.g. a syllable with relatively short duration can still be stressed if it has an especially wide pitch span or high mean F0. A syllable has been defined as stressed, and assigned a stress mark, if it possesses more of these qualities than all adjacent syllables.

6 Some thoughts on tone 4

In section 3.2 I defined tone sandhi as “tones changing their shape as a result of the influence of adjacent tones”. I mentioned three kinds of tone sandhi; tone 3 sandhi, half tone 3 sandhi and tone 2 sandhi. These are the forms of tone sandhi described in Y. R. Chao’s (1968) “Grammar of Spoken Chinese”, and appear to be the ones most frequently discussed in Mandarin phonology. Some studies also mention a fourth kind of tone sandhi that I did not discuss in chapter 4, namely **tone 4 sandhi**. Among the works mentioning tone 4 sandhi, we find Chao (1948), Shen (1989), Chen (2000) and Yin (2003). The tone 4 sandhi rule states that a tone 4 loses its latter half, and therefore changes from [51] to [53] when followed by another tone 4 (Chao, 1948: 26). For example:

再见 zàijiàn ”good bye”: [ʋ ʋ] → [ʋ ʋ]

There exists an alternative approach to this same phenomenon, which notes that the range of tone 4 may be reduced in some cases, but which specifically states that this reduction is not a kind of tone sandhi. Instead, it describes the reduction of tone 4 as a result of weak stress (Chao, 1968; 1997).

In my phonetic transcription of Xué Wàiyǔ, tone sandhi is noted by a change in the tone letter, as in the example above. Degrees of strong and weak stress, on the other hand, are noted by the use of stress marks. To mark a kind of tone change that is entirely a result of stress with changes in the tone letter would therefore be redundant, as that change will already be accounted for through the use of the stress marks. Because of this, I needed to decide on the most satisfying approach for describing the behavior of tone 4 in the material I have transcribed. To this end, I will ask the question:

Is there any evidence of tone 4 sandhi, as described by Chao (1948), Shen (1989), Chen (2000) and Yin (2003), in the present material?

6.1 Theoretical background

As Yin (2003: 297) points out, the tone 4 sandhi rule is very similar to the tone 3 sandhi rule. When a tone 3 is followed by another tone 3, it loses its first (low-falling) half. A similar phenomenon apparently occurs with tone 4; when it is followed by another tone four it loses its low-falling half. In the case of ton 4, that low-falling part is the second half and not the

first. It seems from this that the mechanisms at work in tone 4 sandhi are the same as in tone 3 sandhi, which would mean that it is a systemically more sound approach to treat them both as variants of the same phenomenon.

The phonetic motivation for tone 4 sandhi has been described by Shih Chilin (1988). She does not explicitly name this phenomenon tone sandhi, but rather calls it “tonal coarticulation”. While there have been attempts to define tonal coarticulation as distinct from tone sandhi,²⁷ this distinction makes no difference for the present study; they are both caused by tones’ influence on each other. Moreover, Chen (2000: 23-28) argues that tonal coarticulation is not principally different from tone sandhi. There is only a practical difference, in that the traditionally recognized tone sandhi phenomena (tone 2 sandhi, half tone 3 sandhi and tone 3 sandhi) are detectable by ear alone, while tonal coarticulation might be detectable only by instruments. In the case of tone 4 sandhi, this does not even appear to be the case, as it is mentioned in studies that are not based on instrumental analyses of phonetic data (e.g in Chao, 1948).

Shih’s (1988) description of tonal coarticulation is as follows. Tones have inherent characteristics of high middle and low pitch (usually abbreviated to H, M and L). For example, a tone 4, with a phonetic value of [51], begins as H and ends as L. If two tones are pronounced in succession, it is not always possible for the pitch to move instantaneously from the end point of one tone to the starting point of the next. Some kind of time interval or tonal modification might be required to bridge the gap (Shih, 1988). Figure 6.1, borrowed from Shih (1988: 86), illustrates this effect in the word 海鸥 *hǎiōu* “seagull”. The transition from the L end point of the first syllable to the H starting point of the next syllable requires some modification of the ideal pitch contour. Note that a tone 4 following another tone 4 must undergo the same transition from L to H.

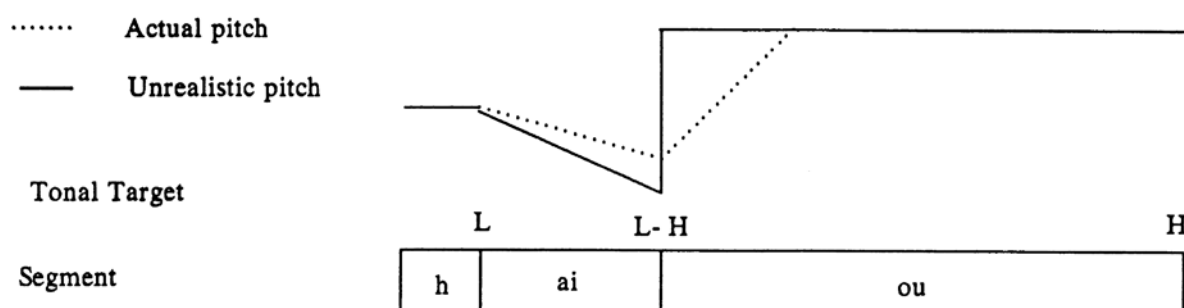


Figure 6.1: Tonal coarticulation

²⁷ See Chen (2000: 23-28) for a summary of this discussion.

Shih (1988) supports this description through an investigation of phonetic data. Her study consists in recording 16 disyllabic expressions with identical segmental features and different tones. The expressions were all made up of the two syllables /fu ji/, and all had primary stress on the second syllable. All possible combinations of the four tones were covered. Her most prominent findings were that tone 4 ends higher when followed any tone, tone 2 ends lower when followed by tone 1 or tone 4, and tone 3 starts higher when preceded by tone 2 or tone 1 (Shih, 1988: 88). If one accepts the argument that tonal coarticulation is to be treated as identical to tone sandhi, this means that Mandarin has the following additional tone sandhi rules: tone 3 changes from [214] to [314] when preceded by tone 2; tone 2 changes from [35] to [25] when preceded by tone 3, and changes to [34] when followed by tone 1 or tone 4; and tone 4 changes to [53] when followed by any tone. The potential effect of these tone sandhi rules on my phonetic notation is illustrated in these examples. The tone 4 sandhi rule is found in the first example:

附记 fùjì “supplement”:	[ʋ ʋ]	→	[ʋ ʋ]
扶戟 fújǐ “to hold a halberd”:	[1 ʌ]	→	[1 ʌ]
打熬 dǎ’áo “To suffer”	[ɿ 1]	→	[ɿ 1]
扶乩 fújī “planchette writing”	[1 ɿ]	→	[1 ɿ]

I will focus on tone 4 sandhi and ignore the other sandhi rules presented here for the time being, although I will return to them briefly in the conclusion.

Tone 4 sandhi is often described as a tone 4 losing its latter half when followed by another tone 4 (Shen 1989: 33; Chao, 1948: 26; Yin, 2003: 297). But if the reason for this change is the need to affect a transition between the final L of the first tone and the initial H of the second tone, why is it the first tone and not the second one that changes? For one thing, the tone sandhi processes of both Mandarin and other Chinese dialects are generally regressive; tones are usually affected by the following syllable and not the preceding syllable (Chao, 1968: p. 27). In addition, Yin (2003) formulates a rule which states that a “lengthened syllable keeps its lexical full tone” (2003: 300). Because the second syllable of a disyllabic expression is generally longer than the first, this explains why it is the first syllable which normally changes. Note also that all the disyllabic expressions investigated by Shih (1988) have stronger stress on the second syllable, which usually means that this syllable will be longer in duration.

In “The Mandarin Primer”, Y. R. Chao (1948) was the first to describe T4 sandhi. However, he later changes his mind about this. In the “Grammar of Spoken Chinese” (Chao, 1968), he states that his previous description of the phenomenon was inaccurate:

“I used to set up another rule of tone sandhi to the effect that a [tone 4] followed by another [tone 4] does not fall quite to the bottom [...]. But since stress will enlarge the range and length of a tone and since a two-syllable compound or phrase will have a slightly greater stress on the second syllable, unless it is in the neutral tone, a succession of two [tone 4 syllables] is more accurately represented as a small graph “V” followed by a bigger one than as represented above.” (Chao, 1968: 28-29)

He also suggests that if the reduction of tone 4 due to stress were to be represented with a different tone letter, there is no reason that its values should be [53]. In fact, it would be more accurate to use a [42], as the tone is not necessarily cut off at the bottom (Chao, 1997: 68). Recall that T4 sandhi has been described as similar to T3 sandhi because both processes consist of a tone losing its second half when followed by another tone of the same kind (Yin, 2003: 297). This generalization does not hold if one assumes that the shape of the reduced tone is more akin to [42] than to [53].

As Chao (1968: 29) points out, his recommended approach of treating the reduced tone 4 as a stress-related phenomenon applies just as much to the other tones. If a tone 4 increases its range when stressed, then so will any tone.

6.2 Examining the data

In section 4.3, I detailed my approach to transcribing Mandarin tones, and argued that it was necessary to see the tones as relative to the pitch span of each individual syllable. Because the pitch span expands and contracts as a result of stress, it should be presumed that a strongly stressed tone 4 will cover a greater range than a weakly stressed tone 4. This is in line with Chao’s (1968; 1997) view that tone 4 reduction is a result of weak stress. Nonetheless, it is still possible that tonal coarticulation works in combination with stress, so that the tones of weakly stressed syllables are reduced according to a predictable pattern. If so, one would expect the influence of tone 4 sandhi to be superimposed on top of the effects of stress, so that a weakly stressed T4 becomes a “small [53]” and not a “small [51]”. In this case, it would be important to mark this effect in the transcription, because it takes place independently of the changes represented by the stress marks.

In order to arrive at a phonetic notation for transcribing Xué Wàiyǔ I decided to test the patterns of tone 4 change in that specific material. I have examined in detail some disyllabic expressions with two consecutive tone 4 syllables. This has allowed me to compare the two tones in each individual expression, and it has also allowed me to compare the expressions to each other. Strings of two tone 4 syllables is, by chance, the most frequently occurring kind of disyllabic construction with two tonic syllables in my material. Analyzing these has helped me to determine which of the two views described above is best suited to describing tone 4 change in the present material.

Disyllabic expressions of two T4 syllables were chosen from Xué Wàiyǔ, based on the following criteria:

1. Background noise must be limited, and sound must be clear enough for the computer to perform an accurate pitch analysis.
2. The expression must be uttered by speaker A.
3. Both tones must be realized as a typical tone 4 (i.e. not reduced to the neutral tone).
4. The expression must be part of a single NP or VP containing no other tonic syllables, with a grammatical structure of either (V-N), (N-N), (Mod-N) or (Adv-V).

The expressions that fulfilled these criteria were segmented in Praat, and the tone carrying segments were assumed to be the syllabic vowel and any voiced segments that follow it before the syllable boundary²⁸. The highest and lowest F0 values within this domain were then assumed to be the starting point and end point of the tone, and were measured. The results are presented in table 6.1, along with the stress pattern of each expression and the general trend of the global intonation contour at the point where the expressions occur. Intonation was identified as either falling or rising. In cases where the expression carries phrasal stress, the intonation contour has been treated as rising when stress is on the second syllable, and falling when stress is on the first syllable.

Of the expressions with strong stress on the second syllable, only bìyè (7) has a structure that corresponds more or less to the one predicted by the tone 4 sandhi rule. Here, the tones on the first and second syllables peak on almost exactly the same pitch levels, but the second syllable falls much further. From this, it could be argued that the tones be transcribed [53 51].

²⁸ As proposed by Howie (1973: 219), and discussed in the previous chapter.

However, the rule predicts that the first tone will fall to the middle of its normal range, while in this case it covers less than a fourth of the range covered by the second tone. Duihuà (111) is the complete opposite, in that the two tones fall to almost the same level but start at different levels. Here the range covered by the weakly stressed syllable is very close to half the range covered by the strongly stressed one.

Stress	Expression	Syllable	F0 high	F0 low	intonation
weak-strong	bìyè 毕业 (7)	[piŋ]	240	208	Falling
		[ie:ŋ]	239	94	
weak-strong	wèn lù 问路 (87)	[vənŋ]	301	270	Rising
		[lu:ŋ]	337	190	
weak-strong	duìhuà 对话 (109)	[tʰɛiŋ]	208	188	Falling
		[χʰɛ:ŋ]	221	184	
weak-strong	zàijiàn 再见(115)	[tsæiŋ]	214	170	Rising
		[tɕiɛnŋ]	244	154	
strong - weak	tài shuài 太帅 (15)	[tʰaɛŋ]	267	214	Falling
		[ʂʰɜɛŋ]	163	125	
strong - weak	tài gòu 太够(23)	[tʰɐɛŋ]	210	169	Falling
		[koulŋ]	163	138	
strong - weak	tài huì le 太会了 (37)	[tʰaɛŋ]	304	191	Falling
		[xʰɛ:ŋ]	218	186	
strong - weak	tài huì le 太会了 (39)	[tʰaɛŋ]	322	177	Falling
		[xʰɛiŋ]	215	179	
strong - weak	tài huì le 太会了 (73)	[tʰaiŋ]	418	228	Falling
		[xʰɛ:ŋ]	178	135	
strong - weak	zuì kuài 最快 (89)	[tsʰɛiŋ]	372	273	Falling
		[kʰʰaɛŋ]	336	238	

Table 6.1: Pitch values of tone 4 disyllabic expressions

Wèn lù (89) and zàijiàn (117) are very similar, in that the tones of the weakly stressed first syllables in both cases start lower and finish higher than that of the strongly stressed syllables. If this phenomenon were to be distinguished using tone letters, it would seem best to transcribe the expressions [42 51], representing the weakly stressed tone with a graph that moves from [4] to [2], instead of from [5] to [3]. It should be noted that both Wèn lù (89) and zàijiàn (117) occur in a position where the global intonation contour is rising, although it is difficult to say from this limited data set whether the intonation actually affects the relationship between the tones.

Among the syllables with a stress pattern of 1-2, the results are more uniform. The tone of the strongly stressed first syllable always begins at a much higher pitch than the tone of the weakly stressed syllable, and it covers a greater range. In three out of six cases, the second tone 4 actually begins at a lower pitch than the first tone 4 ends. This is interesting, because it would seem to eliminate the need for “bridging the gap” between the final L of the first tone and the following H of the second tone. This strongly supports the argument I made in section 4.3, where I showed how a (nominally “L”) tone 3 can actually be higher than a (nominally “H”) tone 1 in the same phrase. In the material I have transcribed, the intonation contour of a phrase fluctuates so much that phonological values like H, M and L (or [1] [3] and [5], on the five point scale) are highly relative. In such an environment, it is possible that there is no need for the phonetically motivated coarticulation effect described by Shih (1988: 86-88) and illustrated in figure 2.2.

It should be noted that all the 1-2 expressions I have analyzed occur in an environment of falling intonation, and that an analysis of similar expressions with a rising intonation might yield different results altogether.

Figures 6.2 and 6.3 show the tonal contours of two disyllabic expressions with contrasting stress patterns.

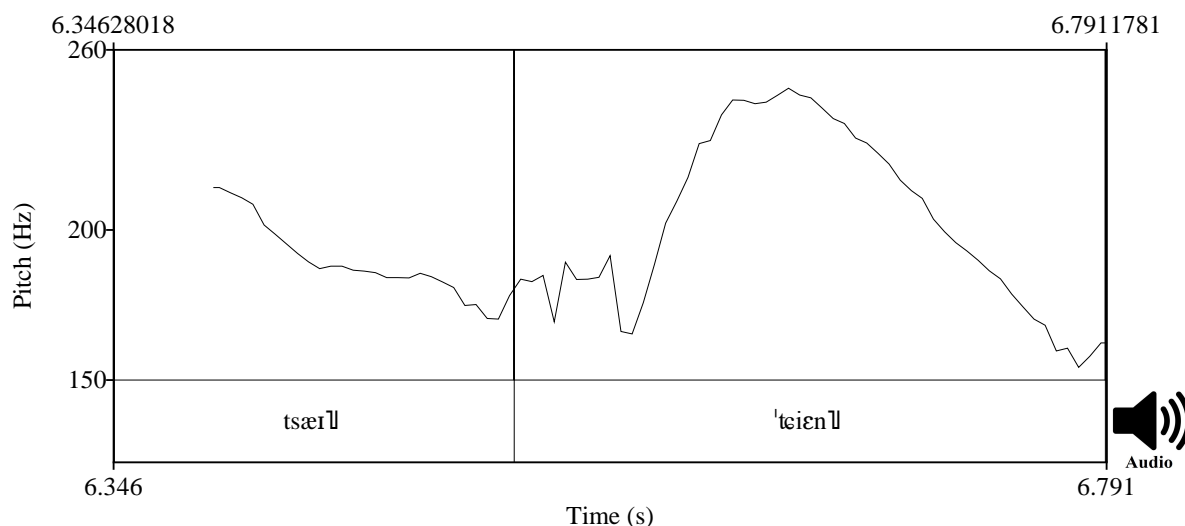


Figure 6.2: Pitch contour of zàijiàn 再见 (115)

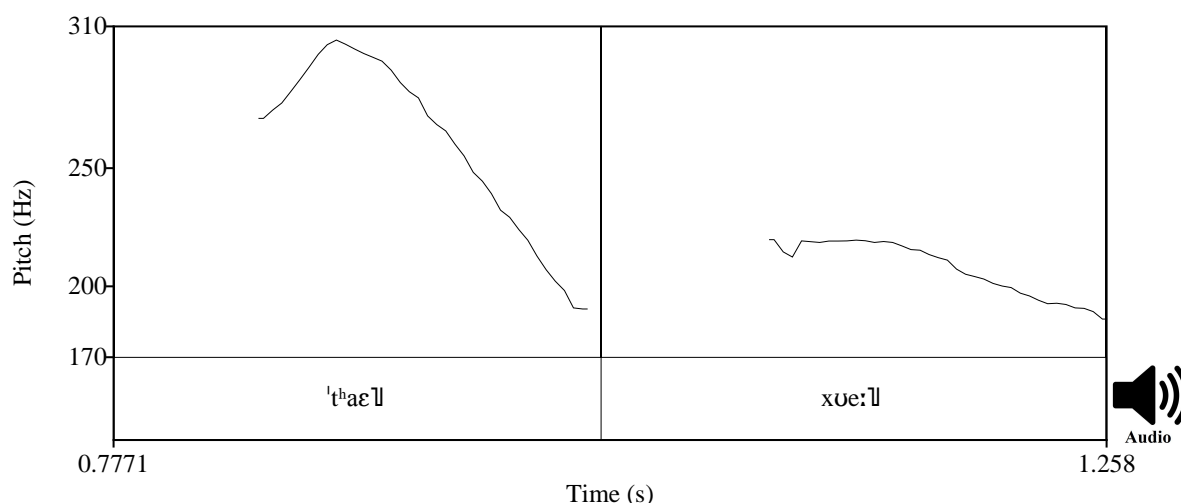


Figure 6.3: Pitch contour of tài huì 太会 (37)

6.3 Conclusions

All told, the data examined here clearly shows that tone 4 syllables with strong stress display greater pitch movement than similar syllables with weak stress. This is entirely in line with the acoustic correlates of stress that were defined in chapter 5.

The results of this study provide no evidence that tone 4 sandhi, as described by Chao (1948), Shen (1989), Chen (2000) and Yin (2003), occurs in Xué Wàiyǔ. There does not appear to be any regularity in the behavior of tone 4, other than that caused by stress and intonation. Moreover, the variation of the pitch span that can occur inside a single intonation phrase means that values like “H” and “L” are highly relative. It is possible, therefore, that the phonetic motivation for tone 4 sandhi, as described by Shih (1988: 86-88), has no greater relevance to the present material.

These observations have led me to choose the following notation for transcribing strings of two tone 4 syllables:

再见 zàijiàn → [tsæi˥˩ ˈtɕiɛn˥˩]

Here, the greater pitch movement of the second syllable is represented by a stress mark and not by a different tone letter. This notation is, in my opinion, a systemically more satisfying solution to transcribing tones in the material I have worked with, as it accounts for the variations in all tones and not only tone 4. Because the tonal coarticulation effects described

by Shih (1988: 86-89) seem to be largely irrelevant to my material, I have also seen no need to attempt using the tone letters to account for the various other tone sandhi phenomena she describes.

7 Tone 3 and tone 3 sandhi

In chapter 6, I mentioned the phonological rules that govern the behavior of tone 3; namely tone 3 sandhi, and half tone 3 sandhi. Of these, especially tone 3 sandhi has received much attention in studies of Mandarin phonology. This phenomenon has been reported by phonologists as far back as the fifteenth century (shih, 1997: 82), and recent studies that discuss it include Chen (2000), Duanmu (2000), Shih (1997) and Shen (1989). Duanmu (2000) in fact suggests that tone 3 sandhi is “perhaps the best known phonological process in [Standard Chinese]” (237). Despite this, there have been claims that tone sandhi is a mostly theoretical concept, and that it occurs only very infrequently in connected speech (Shen, 1989: 49). This objection may also be relevant to the present material; in a study of tonal processes in Běijīng xiàngshēng, Paul Kratochvil (1984) concludes that “the evidence for the existence of [tone sandhi] phenomena in [Beijing Dialect] stage speech is very weak” (1984: 146). There have also been studies which note that, although tone sandhi does occur in connected speech, the tonal variants commonly associated with sandhi may also occur in other contexts (Shen, 1989: 53-54). This chapter will summarize the occurrences and variations of tone 3 in Xué Wàiyǔ, and compare the results to the commonly acknowledged phonological rules of tone sandhi. On the basis of this comparison, I ask:

Can the tone sandhi rules proposed by Y. R. Chao (1968) account for the variations of tone 3 in the present material?

7.1 Theoretical background

The basic rules of tone sandhi as proposed by Y. R. Chao (1968: 27-28) were described in section 3.2. Tone 3 is presumed to have an ideal, low-falling high-rising form of [214], which it retains in isolation and in phrase final positions. When followed by another tone, tone 3 changes to a rising [14], and when followed by any tone but a tone 3 it changes to a low-falling [21]. This means that tone 3 has three variants, and that these variants can only occur in the specific positions referenced here. In other environments, the variants are systematically absent (Chen, 2000: 21).

Most recent studies of tone 3 sandhi agree that there is an exception to this rule; in long strings of tone 3 syllables, tone sandhi may (or may not) be blocked and the tones retain their ideal form. For example, a sentence like Lǎo Lǐ mǎi hǎojiǔ 老李买好酒 “old Li buys good

wine” can take a number of different forms, depending on the tempo at which it is read (Shih, 1997: 85):

老李买好酒	lǎo Lǐ mǎi hǎojiǔ	→	[\uparrow \uparrow \downarrow \uparrow \downarrow]
		→	[\uparrow \uparrow \downarrow \uparrow \downarrow]
		→	[\uparrow \downarrow \uparrow \uparrow \downarrow]
		→	[\uparrow \uparrow \uparrow \uparrow \downarrow]

In situations like this, both [21] and [214] can occur in an environment where they would otherwise be absent. Explaining how tone sandhi applies in these lengthy expressions is the goal of a great many studies, and is described by Duanmu (2000) as “the greatest challenge in the analysis of [tone 3 sandhi]” (238). However, I will not go into greater detail about the various solutions proposed to the problem as they are not relevant to this discussion.

Further exceptions to the rules of tone sandhi have been noted by Susan Shen (1989). In her empirical study of Mandarin connected speech, she found two instances of tone 3 variants systematically occurring outside the environments prescribed by phonological theory. First, she found that stressed tone 3 syllables kept their ideal shape of [214]. This also coincided with the blocking of tone sandhi in long strings of tone 3 syllables, leading her to propose that tone sandhi is an entirely stress related phenomenon (Shen, 1989: 60-61).

The second exception to the sandhi rules that was noted by Shen (1989) is the tendency of tone 3 to change into its rising variant even when it is not followed by another tone 3 (Shen, 1989: 53-54). The reason for this, she argues, lies in the influence of intonation on the lexical tones. If tones occur in an environment where the global intonation contour displays a particularly strong rising tendency, the need to follow that tendency may override the tone sandhi rules. Because the intonation contour is most often rising at the beginning of an intonation phrase, a tone 3 that occurs phrase-initially may be realized as a rising tone, even though it is not followed by another tone 3 (Shen, 1989: 52-57). However, Shen does not propose this as a new phonological rule for Mandarin tone change. Instead, the phenomenon is explained as a possible case where pragmatic concerns override the phonology, allowing native speakers to disregard some of its conventions (Shen, 1989: 53).

7.2 Examining the data

As mentioned in section 4.3, my transcription separates four different variants of tone 3, instead of just the regular three. The reason for this is that some occurrences of tone 3 are not easy to classify as either [21] or [214]. These “borderline cases” generally display the falling contour [21], and then begin rising after the low point has been reached. On average, they rise more or less to the level of their starting point, and have therefore been transcribed as [212]. Sometimes they may rise slightly higher, and sometimes not quite as high.

In my notation, the [212] variant is therefore a kind of wastebasket category; it is defined primarily as being neither [21] nor [214]. The occurrence of syllables that were not easily classified as one or the other of these two is relatively frequent. It is also quite systematic, as the [212] variant occurs with certain specific syllable types and stress patterns. However, it is worth noting that I have not carried out any tests on whether [212] is actually perceived by native speakers to be distinct from [21] and [214]. The inclusion of this category is therefore to be regarded primarily as a kind of analytical tool. I do not intend the inclusion of an extra variant of tone 3 as a criticism of phonological theory, but rather as a reflection of the fact that phonetic reality is more like shades of gray than black and white. Because ambiguous cases were numerous, I deemed it useful to include an intermediary analytical category.

The typical difference in tonal contour between [212] and [214] is illustrated by the examples in figures 7.1 and 7.2.

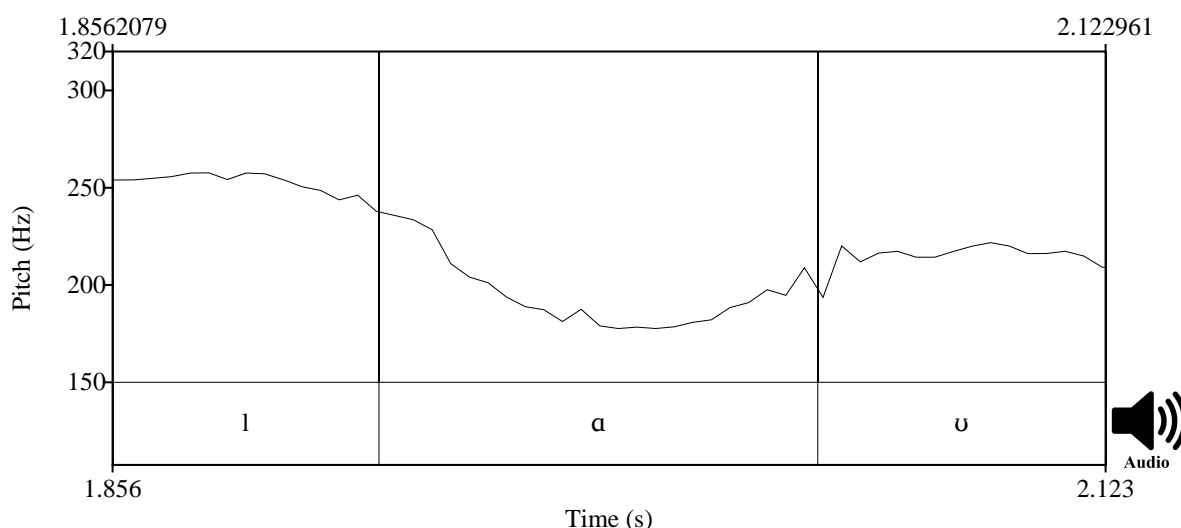


Figure 7.1: Pitch contour of lǎo 老 (1) [lǎo˥˩]

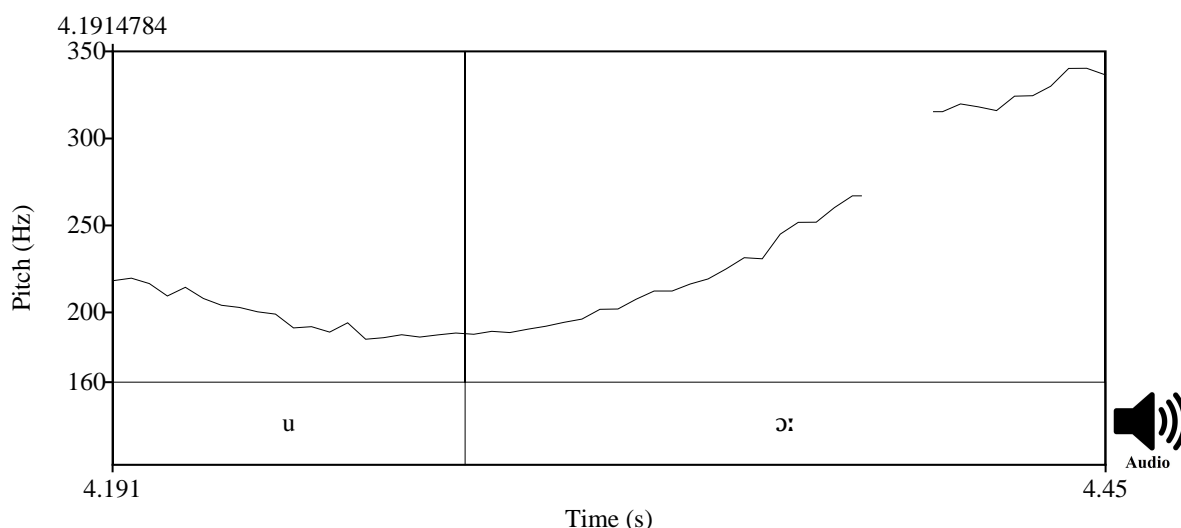


Figure 7.2: Pitch contour of wǒ 我 (66) [ʊ˥˥˥]

Some of the syllables that are nominally tone 3 have lost their tonal contours altogether. These have been treated in the transcription as neutral tones, and are marked in this discussion as [NT]. While [NT] is simply a neutral tone, and therefore not strictly a variant of tone 3, it has some repercussions on the discussions of the other tone 3 variants. Specifically, it has been reported that tone 3 changes to [14] when followed by a neutral tone, if that neutral tone is an underlying tone 3 (Shen 1989: 55). Taking the four variations into account, along with the tone 3 syllables reduced to neutral tones, the total occurrences of tone 3 are displayed in table 7.1. The number of occurrences of each variant is listed along with the percentage of the total occurrences of tone 3.

Variant	Number of occurrences	Percent
[214]	13	4,6 %
[21]	125	44,1 %
[14]	81	28,6 %
[212]	23	8,2 %
[NT]	41	14,5 %
Total	283	100 %

Table 7.1: Variations of tone 3

As expected, the ideal form [214] makes up a relatively small number of the total occurrences of tone 3; only 4,6 %. This variant shows up at the end of phrases, which is just as predicted

by the tone sandhi rules. However, it also occurs in non-final position, even when it is not subject to tone sandhi blocking. Remember Shen's (1989: 60-61) claim that stressed syllables keep their ideal form. This material does not seem to conform to that expectation. That is; all syllables that keep the [214] shape are stressed, but not all stressed syllables keep the [214] shape. Apparently, the syllable must have particularly heavy stress in order to produce this effect. The typical example in this material of a tone 3 keeping its ideal shape is when the tone 3 syllable gets primary stress. For example:

66. B: 我说得也挺熟阿



Wǒ shuō de yě tíng shǒu ā

Audio [luːʃ sʷɔːl deːlɛ tʰɪŋˌsoːlˈɑː]

“Even I can speak [Chinese] pretty fluently”

Some syllables that are subject to special circumstances also seem to sometimes keep the [214] shape. One typical example is [nai] (64), which is a merger of two distinct syllables. [nai] (64) has two relatively long vowel segments, one of which, [i], rarely shows up as an off-glide. Because of this, and because [nai] may be construed both as “něi 哪” and as “nǎ yí 哪一”, this situation is an ambiguous one. [nai] might just as well have been analyzed as two syllables.

62. B: 那么您哪一国话说得最熟呢。



Nàme nín nǚguóhuà shuō de zuì shǒu ne?

[[nəm'ɪmə'ɫɪnɪn'ɪnəɪ k'oʊɿχ'oʊɿɿʊoʊɿtɜːt s'ɛ'ɿʊɿnəʊɿ]]

“Then which language do you know best?”

The “waste bin” variant [212] often shows up in similar circumstances to [214]. It is common in phrase final position, and also in syllables with phrasal stress. In addition, [212] shows up in a position where the [214] variant does not; immediately preceding a syllable with primary stress. In *Xué Wàiyǔ*, a tone 3 which directly precedes a syllable with primary stress is often realized as [212], but never as [214]. Here, [laʊ̯] (1) precedes a syllable with primary stress:

1. A: 张庆森说相声是老资格



Zhāng Qìngsēn Shuō xiàngshēng shì lǎo zīgé.

[tʃsan¹ tɕʰin¹ ʋ¹ sən¹ ʔ s^woː¹ ɕ¹an¹ ʋ¹ səiː¹ ʃuː¹ laʊ¹ ʔ tʃiː¹ kaː¹]


"Zhang Qingsen is a real veteran at performing xiangsheng"

The rising variant of tone 3 [14] is assumed by Chao (1968: 27) to occur only when a tone 3 syllable is followed by another tone 3 syllable. In the string of tone 3 syllables in turn 135, the first two, [ni] and [tʂao], both adopt the rising variant. Note that this is also the longest string of tone 3 syllables that occurs in Xué Wàiyǔ:

of the rising variant of tone 3 (66,6 %) seem to be entirely unrelated to tone sandhi as it is traditionally described. Examples like the following are typical:

52. B: 您是个男的您是个女的?

Nǐ shì ge nán de nǐ shì ge nǚ de?



Audio

[ni:ʌ ʃɿ˥˥kɛ˥˥ 'nan˥˥tɛ˥˥ ni˥˥ ʃɿ˥˥kɛ˥˥ nɥ˥˥˥tɛ˥˥˥]

"Are you a man or are you a woman?"

This appearance of a rising tone 3 not followed by another tone 3 is similar to that observed by Susan Shen (1989: 52-57). She attributed this unexpected change of tone 3 into a rising tone to the effects of intonation. If the tone appears in the first part of a phrase, where the intonation is usually rising, global effects may override the local tone sandhi effects and tone 3 may change into a rising tone. To investigate whether this might be the case in the present material as well, I have grouped the occurrences of [14] according to whether they appear in a phrase-initial position or not. Table 3.5 lists the number of occurrences in the various positions, as well as the percentage of the total occurrences of [14].

The data in table 3.5 show that [14] does occur in non-initial position, even when it does not precede another tone 3. However, it does so relatively rarely; only 9 out of 49 times. This seems to suggest that the same mechanism is at work here as that described by Shen (1989: 52-57), and that intonation influences the output of tone sandhi. A tone 3 syllable preceding another tone would be expected to change into [21], but in these cases it changes into [14] instead. It is also worth noting that a tone 3 followed by another tone and occurring in phrase-initial position does not always change into a rising tone. The normal [21] output of half tone 3 sandhi also sometimes occurs in phrase-initial positions. Of all phrase-initial occurrences of tone 3 (discounting those that keep their full [214] shape) 21 are low-falling [21] variants. This is roughly half as many as the occurrences of rising variants in the same environment.

	Initial position	Non-initial position	Total
Preceding any other tone	45 (55,5 %)	9 (11,1 %)	54 (66,6 %)
Preceding reduced tone 3	9 (11,1 %)	3 (3,7 %)	12 (14,8 %)
Preceding tone 3	9 (11,1 %)	6 (7,5 %)	15 (18,6 %)
Total	63 (77,7 %)	18 (22,3 %)	81 (100 %)

Table 3.5: Variations in the rising form of tone 3

Interestingly, the tone 3 syllables that are followed by another tone 3 also occur predominantly in phrase-initial position. All told, this means that almost 80 % of the total occurrences of [14], whether as an output of tone sandhi or not, occur phrase-initially. On the basis of this, it could be argued that tone sandhi as a whole is more likely to occur in an environment of rising intonation. However, the data set is hardly large enough that such arguments would carry any real value, especially considering that there are no instances of tone 3 sandhi blocking in environments with falling intonation. Most strings of tone 3 syllables in this material coincide with a rising trend in the intonation, and to speculate what would happen if they did not would be purely hypothetical.

In my opinion, the most important insight that can be gained from these data concerning the rising variant of tone 3 is that it occurs independently of the traditionally described tone 3 sandhi. Moreover, these non-sandhi occurrences of [14] are far more frequent. Not only does this mean that Chao's (1968: 27) tone sandhi rules are insufficient to account for the behavior of tone 3 in this material; it also means that intonation seems to be a more important factor than tone sandhi when it comes to determining the shape of this tone. As a final note, it might be worth considering that the rising variant of tone 1 (which I have transcribed as [45], and which is further remarked on in footnote 30) also seems to occur as a result of rising intonation. In fact, the majority of its occurrences take place in the same phrase-initial environment as [14]. This could also be an indication that intonation may in many cases influence the shape of lexical tones.

7.3 Conclusions

In this chapter I have shown that the sandhi rules proposed by Chao (1968), and widely accepted in general phonological literature (Chen, 2000; Duanmu, 2000; Shih, 1997), fail to account for the full behavior of tone 3 in Xué Wàiyǔ. It does, however, account for most of it.

The low-falling low-rising variant of tone 3 [212] has been added in addition to the three variants described by Y. R. Chao (1968). This should not be seen as an attempt to create an additional phonological category, but rather as a reflection of the intermediary stages, or “gray zones”, that are always found in phonetic analyses. The occurrences of the falling variant of tone 3 [21] are entirely in line with the predictions of the half tone 3 sandhi rule. The rule stipulates that the falling variant occurs before any other tone than a tone 3, and that appears to be the case in the present material. The falling variant sometimes occurs in phrase final position, which is also in accordance with Chao's (1968) description. The low-falling high-

rising variant [214] is usually presumed to appear only in phrase-final position, but in this material it seems to appear in intra-phrasal positions as well. However, this happens only in special cases; either in syllables with very heavy stress, or in syllables that are formed from merging of two adjacent syllables. Susan Shen (1989: 60-61) suggests that any stressed tone 3 syllable will retain its ideal tone [214], but this does not appear to be the case here.

The case of the rising variant [14] is less in line with phonological predictions. Tone 3 becomes a rising tone when followed by another tone 3, but in this material it also changes to a rising contour in other environments. This change is not predicted by Chao's (1968: 27) tone sandhi rules. The non-sandhi occurrence of [14] is most likely to take place in the beginning of an intonation phrase, where the global intonation contour is rising. This corresponds well to Shen's (1989: 52-57) proposal that intonation may in some cases override the effects of tone sandhi.

In *Xué Wàiyǔ*, the majority of the occurrences of the [14] variant of tone 3 seem to be a result of intonation and not of tone 3 sandhi. Tone 3 sandhi, which Duanmu (2000) suggests is “perhaps the best known phonological process in [Standard Chinese]” (237), therefore seems to be of only secondary importance here. This might point to a deviation of this material from the generic Mandarin Chinese discussed by most phonologists.

8 Phonetic transcription

1. A: 张庆森说相声是老资格。

Zhāng Qìngsēn Shuō xiàngshēng shì lǎo zīgé.

[tʂaŋ¹²⁹ tɕʰiŋ³⁰ ʃsən³⁰ ʃʷo: ʈ tɕʰaŋ³⁰ ʃsəŋ. ʃʃ. ʃaʃ ʈ tsɿ: ʈ ka: ʃ]³¹

"Zhang Qingsen is a real veteran at performing xiangsheng"

2. B: 啊呀可不敢那么说啊。

Āiyā, kě bù gǎn nàme shuō a.

[ʔʂɿ³² ʈ tɕʰa: ʈ kʰʌ³² ʈ pɕʰ³² ʈ kan³² ʈ nəm³² ʈ ʃʃ³² a: ʈ]

"Oh no, I wouldn't dare say that!"

3. A: 听他们提过。

Tīng tāmen tíguo.

[tʰɪŋ³² ʈ tɕʰ: ʈ mən³² ʈ tʰi: ʈ g³² a: ʃ]³²

"I've heard them say it"

²⁹ This is an example of a tone 1 getting a rising contour [45] instead of the flat contour [55] it has in citation form. In this material, the rising variant of tone 1 often, though not always, seems to occur in the beginning of phrase, where the intonation contour is generally rising.

³⁰ The final consonant in the syllable /sēn/ assimilates the following consonant's place of articulation, and is realized here as [sən³⁰]. Final nasal consonants frequently assimilate a retroflex place of articulation in this material.

³¹ In citation form, the word /lǎo zīgé/ has a tone 2 on the last syllable, but in this instance it gets a kind of falling tone. It appears to be a characteristic of informal spoken Chinese to give the final syllable in a phrase a neutral tone, and as I pointed out in section 1.1.4, neutral tones have been known to have a falling contour [42] at the end of phrases. There are many similar cases in this piece, and I have transcribed them the same way. Another thing to note is the vowel of the final syllable. The typical realization of the syllable /ge/ is [kɤ], but in this case the vowel is a fully open [a]. This and similar phenomena occur quite frequently in phrase-final positions.

³² In section 1.2.1 I discussed briefly the lack of phonemic opposition between voiced and unvoiced consonants in Mandarin phonology. However, that voiced plosives and affricates are not phonemic does not mean that such consonants do not occur phonetically (Duanmu, 2000: p. 27; Lin, 2007: p. 158-160). The initial /g/ of the syllable [g³²a: ʃ] is an example of a normally unvoiced consonant that assimilates the voice of the preceding vowel.

"Is that so?"

Shuō nín de wénhuà shǔipíng hěn gāo.

[ʃ̺oː1, nin1, tə1, vāː1³³, ʃ̺eː1, ʃ̺eː1, p̺hɪŋ1, xən1, kəɔ̯1]

"They say you are a very educated person"

Gāo shénme. Fǎnzhèng niànguó jǐ nián shū.

[ʔh 'kauʔ ,səmʔ mɛʔ || fænʔ dɛnʔ³⁴ 'niʔnʔ ʔgʊʔ ,tɛiʔ n'ɛnʔ ,suʔ]

"Oh, there's nothing educated about me. Well anyway, I went to school for a few years..."

Dàxué bìyè.

[³⁵ta:V ɕ^yə1 biV₁ie:V]

"University graduate!"

³³ The syllable, [vẽ:ɫ] drops its final consonant /n/, something which seems to happen quite frequently in the present material. The final /ng/ is also sometimes dropped, but not as often as /n/. When the final consonant is dropped, the vowel always remains nasalized.

³⁴ The half tone 3 and the following rising tone in [fæn↓dæn4] together form a tonal contour very similar to the citation form of tone 3 [↗].

³⁵ Some scholars (Chao, 2005a; Lin, 2007: pp. 227-232) have claimed that tones under the influence of intonation may sometimes get an extra contour added to their original shape, while others (Shen, 1989: pp. 75-76) maintain that this is not the case. This stressed tone 4 actually gains a rising contour at the end, and could appear to be an example of the former. Because it uses a fixed number of tone letters, and because I have decided not to add extra contours to the tone letters, my transcription does not show this detail. In a more impressionistic notation, the tone of the syllable [ʰta:] could well have been transcribed as [513].

8. B: 啊, 没有没有。 共合我念六年书。
 A méiyǒu méiyǒu. Gònghé wǒ niàn liù nián shū.
 [ʔɜ: || 'mheʔ ʔɔʔ mɛ:ʔ a.ʔ ||³⁶ (0.943) || ,kuŋʋ xΛ:ʔ ʔəʔ n'ɛnʔ 'lɪəʔ n'ɛŋʔ su: ʔ]
"Ah, no, no, nothing like that! All in all I've gone to school six years."
9. A: 客气。
 Kèqi.
 [kʰɤ:ʋ tɕʰɪ.ʔ]
"Now you're being polite."
10. B: 嘿这实在。
 A zhè shízai.
 [ʔɜt ʔʂɜ:ʋ³⁷ ,ʂɪ:ʔ tsæʋ]
"No, this is true!"
11. A: 您只客气。
 Nín zhǐ kèqi.
 [nɪnʔ tʂɪʔ kʰɤ:ʋ tɕʰɪ.ʔ]
"You're just being polite."
12. B: 怎么。
 Zěnmē.
 [tsʌmʔ maʋ]
"Why do you say that?"

³⁶ Phrase spoken while laughing, hence the [h] in ['mheʔ].

³⁷ The final vowel of syllable [ʔʂɜ:ʋ] assimilates the retroflex place of articulation of the following initial consonant. Note that [ʔʂɜ:ʋ] itself has a retroflex initial, but that this is not likely to be the source of the r-coloring because the latter does not set in until about midway into the vowel segment. In this case, the r-coloring seems to work only regressively, but there are other cases where it appears to work progressively as well.

13. A: 客气。您绝对不能念五六年书。
Kèqì. Nín juéduì bù néng niàn wǔ liù nián shū.
[^hkʰɿ̌ʋ̌ tɕʰ || nin¹ tɕʰe:ʋ̌ de¹ bu¹ nən̩¹ n'ẽ̌ʋ̌ ɿu:ʋ̌ l'əʊʋ̌ n'ẽ̌¹ sǔ ʔ]
"You're being polite... There's no way you could have gone to school just six years"
14. B: 真的六年。
Zhēnde liù nián.
[^htʂəň̩ ʔ dæ¹ ɿɿəʋ̌ n'æ̌.¹]
"It's really six years!"
15. A: 我看你写的字儿，太帅了。
Wǒ kàn, nǐ xiě de zìr, tài shuài.
[^hʋ̌ ə.ʋ̌ kʰan¹ ɿni:¹ ʔ ɕʰe:ʋ̌ n̩¹ tɕʰə̌ʋ̌ || ^htʰaɛ̌ʋ̌ ʂ̌ʋ̌ æ̌ʋ̌]
"Your handwriting is really good looking, I think"
16. B: 哎，就是常写嘛。
Ai, jiù shì cháng xiě ma.
[ʔɜa: || tɕʰə̌ʋ̌ ʋ̌ ʔ tʂʰəň̩ ʔ ɕʰe:ʋ̌ mɐ:ʋ̌]
"Heh, well, that just comes from writing a lot"
17. A: 应当要这样。
Yīngdāng yào zhèyang.
[ɿŋ̌ ʔ tʰəň̩ ʔ ʰəʊʋ̌ tʂʰə̌ʋ̌ ʔ ẽ̌:ʋ̌]
"That is the way it should be"
18. B: 是吗？
Shìma?
[^hʂ̌ʋ̌ mɐ:¹.¹]
"Oh?"

19. A: 相声演员必须要有文化。

Xiàngsheng yǎnyuán bìxū yào yǒu wénhuà.

[ʼɛ̃˥˥aŋ˥˥xəŋ˥˥.l̥ ˥˥i̯æŋ˥˥yɜn˥˥ || py˥˥³⁸ ˥˥ey˥˥ ˥˥aʊ˥˥ou˥˥.l̥ ˥˥vəŋ˥˥³⁹ ˥˥χ˥˥e˥˥.l̥]

"Xiangsheng artists have to be well educated"

20. B: 就是。

Jiùshi.

[ʼdʒo˥˥.l̥ʂ˥˥.l̥]

"Well said!"

21. A: 像您这个文化，说相声儿，富富余余。

Xiàng nín zhège wénhuà, shuō xiàngsheng, fùfu yùyú.

[ɛ̃˥˥ɜn˥˥ ˥˥nin˥˥ [ʂeg˥˥gɜ˥˥.l̥ ˥˥vẽ˥˥ ˥˥χ˥˥e˥˥.l̥ ||⁴⁰ (0.873)⁴¹ || ɛ̃˥˥o˥˥ ˥˥ ˥˥ɛ̃˥˥aŋ˥˥xəŋ˥˥.l̥ || hm ||

˥˥fu˥˥.l̥ ˥˥vu˥˥ ˥˥jy˥˥ ˥˥jy˥˥]

"With an education like yours... when you perform Xiangsheng, it's really... hm... it's more than enough"

³⁸ This syllable seems to display some kind of vowel harmony. The vowel is more rounded than [i], but slightly less so than the [y] of the next syllable. The way it is represented in the transcription is not entirely accurate, since my notation does not distinguish intermediate levels of rounding. An alternative way of transcribing this syllable might be with a "more rounded" diacritic: [i̯].

³⁹ Initial /w/ is realized in a number of ways in this text. Although there are exceptions, the general tendency seems to be:

/_en/	→	[v]
/_ei/	→	[v]
/_ai/	→	[v]
/_an/	→	[v]
/_o/	→	[ʊ]

San Duanmu (2000) proposes that "[v] can be used before any vowel except [o]" (p. 25), and the tendency in the present material seems to correspond well to that assumption.

⁴⁰ [gɜ˥˥.l̥] and [χ˥˥e˥˥.l̥] are very long syllables, as the speaker draws out the words as if he is thinking about what to say next. Note that [gɜ˥˥.l̥] has a neutral tone, which is generally presumed to have a duration that is "relatively short" (Chao, 1968: p. 35).

⁴¹ Parentheses with decimal numbers have been used to represent the duration (in seconds) of pauses that are longer than half a second.

22. B: 呵,反正凑合。
He, fǎnzhèng còuhé.
[ʔəɪ || fænʌdʒənʅ 'tsʰouʌ xʌʅ]⁴²
"Well, I make do at least."
23. A: 太够. 太够了。
Tài gòu. Tài gòu le.
[tʰæʌ koʊ || tʰæʌ kouʌ]
"It's more than enough! More than enough!"
24. B: 这不能那么说。
Zhè bùnéng nàme shuō.
[tʂʌʅ 'puʌnŋʅ nɜmʌ⁴³ ʂʷo:ʅ]
"Oh no, that's taking it too far"
25. A: 可也有缺点。
Kě yě yǒu quē diǎn.
[kʰɿʌ i̯e:ʌ i̯oʊʅ tʰɛʰɿɛ:ʅ tiænʅ]
"But you still have one shortcoming"
26. B: 啊我?
Ā wǒ?
[ʔəʊʅ i̯uʌ:ʅ]
"Who, me?"

⁴² The last two syllables in this phrase and the first two of the next are unclear because speaker A interrupts.

⁴³ The tendency of certain disyllabic expressions to merge into a single syllable is well documented in general Mandarin phonology (for example in Chao, 1968: pp. 54-55). Here it happens with the expression nàme 那么, which is realized as [nɜm]. This apparently only happens where the last syllable of the expression is in the neutral tone.

27. A: 有一点儿缺点。
Yǒu yì diǎnr quēdiǎn.
[¹oʊʌ ¹i:ʌ ¹tʰɿɿ ¹tɕ^he: ʌ ¹tʰænʌ]
"A tiny shortcoming"
28. B: 哪一点儿您指出来。
Nǎ yìdiǎnr nín zhǐ chū lái.
[¹na:ʌ ¹ɿɿ ¹tʰa:ʌ ¹nɪnʌ ¹tʃʌʌ ¹tʃ^hʊʌ ¹læʌ]⁴⁴
"Please point it out for me then."
29. A: 您哪，说相声儿这么些年哪。
Nín na, shuō xiàngshengr zhème xiē nián a.
[¹nɪnʌ ¹naʌ || ¹ʃ^wo:ʌ ¹ɕʰaʌ ¹ʃ^hʌʌ ¹tsʌmʌ ¹ɕʰe: ʌ ¹niɛnʌ ¹nɜhʌ]⁴⁶
"You, who have performed xiangsheng for all these years,"
30. B: 是是。
Shì, shì.
[¹ʃʌʌ ¹ʃʌʌ]
"yes, yes."
31. A: 不会说外国话。
Bú huì shuō wàiguóhuà.
[¹puʌ x^uɿʌ ¹ʃ^wo:ʌ ¹ʋaʌ ¹g^uo:ʌ ¹χoʌʌ]
"don't know any foreign languages!"

⁴⁴ The diphthong /ai/ in /lai/ is here reduced to a monophthong [æ]. This is a quite frequent phenomenon in this data set, and occurs with both speakers. Curiously enough, it seems to occur almost exclusively with the particular morpheme lái 来 "to come".

⁴⁵ Duanmu (2000: 80-82) argues that stressed and unstressed (i.e. neutral tone) syllables have inherently different syllable structures. The reduction of syllables like /sheng/ to [ʃ] may be seen as evidence in support of this argument. In the present material it is quite common for unstressed /sheng/ to be realized as [ʃ], while no stressed syllable is this heavily reduced.

⁴⁶ The tones of syllables [naʌ] and [nɜhʌ] are, nominally, neutral tones. In line with the system I introduced in section 4.3, they have therefore been transcribed [42]. It should be noted however, that both the tones are actually just as sharply falling as, and apparently indistinguishable from, a tone 4. A more impressionistic notation would have been better suited to picking up this kind of phonetic detail.

38. B: 你瞧。
Nín qiáo?
[nin¹ ʔe^hɛ:¹]
“Really?”
39. A: 太会啦。
Tài huì la.
[t^hae^h x^oɛɪ^hla.]
“Do I ever!”
40. B: 那么您会哪国话?
Nàme nín huì něigióhuà?
[ʔə^h nəm^h ʔnin¹ x^uɛɪ^h næɛ^h k^wɔ¹ x^oɛ:^h] ⁴⁷
“Then, which language do you know?”
41. A: 哪国话都会呀。
Něiguóhuà dōu huì ya.
[nɛɪ^h k^wɔ¹ x^oɛ^h tɕ x^oɛ:^h p.] ⁴⁸
“I know all languages!”
42. B: 哟?
Ō?
[^oa]
“Oh?!”

⁴⁷ The three stressed syllables [ʔnin¹ x^uɛɪ^h næɛ^h] are separated by pauses. If this were not the case, I would have transcribed them [ʔnin¹ x^uɛɪ^h næɛ^h]. The pause between the two syllables with secondary stress [x^uɛɪ^h] and [næɛ^h] is shorter than that between [ʔnin¹] and [x^uɛɪ^h].

⁴⁸ Interestingly enough, the two speakers pronounce the word něi 哪 differently; speaker A says [nɛɪ] and speaker B says [næɛ]. This stays consistent throughout their entire dialogue.

43. A: 全会呀。

Quán huì ya.

[^ltɕ^hæ[˥] x[˥]eɪ[˥]u[˥]a.]⁴⁹

“I know them all!”

44. B: 哪国话都会呀?

Nǎiguóhuà dōu huì ya?

[^lnæɛ[˩] k[˥]o:[˩] ʃ[˥]o[˥]ɐ[˥] to.ɿ x[˥]æ:[˥]] ⁵⁰

“You know all languages?”

45. A: 各国的话都会。我到过外国。

Gèguó de huà dōu huì. Wǒ dàguo waiguo.

[^lk[˥]ʊ:[˥] k[˥]ɔ:[˥] d[˥]ʌ[˥] ʃ[˥]o[˥]ɐ[˥] tu:[˥] x[˥]ɛ:[˥] || ^uə:[˥] ʃ[˥]tau[˥] ⁵¹ uaɛ[˥] ^lk[˥]ɔ:[˥] ʃ[˥]ɑ:[˥]] ⁵²

“I know every single language. I’ve been to other countries, you know.”

46. B: 哦您哪?

Ō nín na?

[ʃɔ[˥] ʃ[˥]nɪn[˥] na:[˥]]

“You have?”

⁴⁹ This turn is nearly inaudible because of applause in the background.

⁵⁰ Background applause makes this turn difficult to analyse instrumentally, although it can be heard much more clearly than the previous one.

⁵¹ The syllable [tau[˥]] could well be treated as two separate syllables [ta[˥] u.[˥]]. That it is pronounced in this way could suggest that it is actually a merging of the two syllables dàoguo 到过 “to have been”. This interpretation also makes sense from a grammatical point of view.

⁵² The syllables [^lk[˥]ɔ:[˥] ʃ[˥]ɑ:[˥]] are pronounced so close together that they may well be treated as a single syllable [^lk[˥]ɔɑ[˥]]. I have treated it as two syllables because it has two relatively long vowel segments [ɔɑ] and two distinct tone contours (one falling and one rising). It has been claimed, however, that an additional contour may be “added” to tones under certain circumstances, and this would make the distinction less clear cut. (Chao, 2006b; Lin, 2007: pp. 227-232; see also footnote 35).

47. A: 我是出洋留学的女学生。
 Wǒ shì chū yáng liúxué de nǚ xuéshēng.
 [ɿ^uɔ̌ʌ̌ŋľ | ťʂ^hu: ɿ̌ʌ̌ŋ | ľ^u:ɿ̌ ɛ̌^yɛ̌ʌ̌dɛ̌ | 'ny:ɿ̌ ɛ̌^yɛ̌:ɿ̌ŋʌ̌ŋ |]⁵³
"I went abroad as a schoolgirl"
48. B: 哦!
 Ō!
 [ʔɔ:]
"Oh!"
49. A: 我在外国...
 Wǒ zài wàiguó...
 [ɿ^uoʌ̌ tsaě ʌ̌væ̌ɛ̌ k^uo:ɿ̌]⁵⁴
"when I was abroad..."
50. B: 您等点儿说吧。。。啊，你是个女的？
 Nín děng diǎnr shuō ba... ā nǐ shì ge nǚde?
 [niň | 'dɔ̌ŋ | ď^oaɿ̌ | ʂ^{wo}o ɿ̌pɛ̌ | || (1.03) || ʔəň | ɿ̌ni:ɿ̌ | ʂ̌ľ | kɛ̌ | 'ny:ɿ̌ ta:ɿ̌]
"Hold on a second!... You're a woman?!"
51. A: 啊?
 Ā?
 [ɑ:ɿ̌]
"Huh?"

⁵³ Howie (1973: p. 69) and Halliday (1992: p. 307) have reported that the finals /iu/ and /ui/ are realized differently when coupled with different tones. Very generally, both these finals have a tendency to be pronounced as diphthongs when in tones 1 or 2, and as triphthongs when in tones 3 or 4. Similar patterns of tonally correlated variation are found in this material as well. The syllable [ľ^u:ɿ̌] is a typical example of what happens when /iu/ is spoken with tone 2.

⁵⁴ Because of applause and laughter in the background, the phonetic transcription of turns 48, 49 and 50 are only approximate. Some inaudible passages have been omitted from the text.

52. B: 您是个男的您是个女的?
 Nǐ shì ge nán de nǐ shì ge nǚ de?
 [ni:ʌ ʂɿ̌kɛ̌ 'nanʌtɛ̌ | niʌ ʂɿ̌kɛ̌ | ny:ʌtɛ̌ |]
"Are you a man or are you a woman?"
53. A: 就是男。。。说错了。
 Jiùshì nán shuōcuò le.
 [ʔɔ:ʌ 'tɛ̌u:ʌ ʂɿ̌ | nanʌ || 'ʂ̺oʌ tsʰoʌ |]
"I'm a man of cou... Oh, slip of the tongue!"
54. B: 这说错了。
 A, zhè shuōcuò le?
 [xə ʈʂɜ:ʌ ʂ̺o:ʌ ʈʂʰaʌ |]
"That was a slip of the tongue?"
55. A: 出洋留学的男学生。
 Chūyáng liúxué de nán xuéshēng.
 [ʈʂʰuʌ ʈʂɛ̌ŋ ʈʂɿ̌ʈ:ʌ 'ɛ̌yɛ̌:ʌ dɛ̌n | || 'nanʌ ɛ̌yɛ̌:ʌ ʂ̺ɛ̌ŋ |]
"I went abroad as a... schoolboy!"
56. B: 哈。
 Hā.
 [χa:]
"heh..."
57. A: 男学生!
 Nán xuéshēng.
 [nanʌ 'ɛ̌yɛ̌:ʌ ʂ̺ɛ̌ŋ |]
"A schoolboy!"

58. B: 啊，就是个留学生？

A, jiùshi ge liú xuéshēng?

[ʔɒ | || tɕ'ə:V.ʂɿ.l.gə.l'li:V ɕ'əV.ʂɿŋ]

"Oh, so like an exchange student then?"

59. A: 对啦，对啦。

Duì la, duì la.

[t'vəiV.lə || t'wəV.lə]

"That's right! That's right!"

60. B: 那么您会哪国话？

Nàme nín huì nēguóhuà?

[nəm'li.mə'li.ninV x'ɛV 'næiV.k'oV.χ'v:V]

"Then what languages do you speak?"

61. A: 差不多吧，反正。英国话，法国话，德国话，苏联话，意大利话，反正眼面儿前儿的这[就]话，全行。

Chàbùduō ba, fǎnzhèng. Yīngguóhuà, Fǎguóhuà, Déguóhuà, Sūliánhuà, Yìguóhuà.

Fǎnzhèng yǎnmiànr qiánr de zhè (X) huà, quán xíng

[tʂ'hɑ:V.bʊ.l'duo:V.bʌV.vənɿV.ʈʂ' | || (1.41) || 'jinV.goV.χ'vɑV.fəV.ʍV.χ'vəV || 'tʂgV.χ'vəV

ʂɿV.pʂ'V.χ'vɑV || 'suV.l'eʂV.χ'vəV.i:V goV.χ'vəV || (0.41) || fəV.ŋV. i.æmV. tɕ'h'əV.ŋV

tʂɿV.tɕ'əV.χ'vəV || tɕ'h'ɛnV.ɕiŋV]

"Oh, pretty much all of them. English, French, German, Japanese, Russian, Italian...

Pretty much any language you can think of, I know."⁵⁵

⁵⁵ Speaker A uses the popular names for these languages; literally "England-talk", "france-talk", "Japan-talk", "Soviet-Union-talk", etc.

62. B: 会那么多... 那么您哪一国话说得最熟呢。
 Huì nàme duō... Nàme nín nǐguóhuà shuō de zuì shóu ne?
 [x⁵⁵əi²¹ 'nəm²¹ mǎ²¹ t⁵⁵ɔ: ɿ || (2.19) || nəm²¹ mǎ²¹ nin⁴ 'nai⁴ k⁵⁵ɔ ɿ ʃ⁵⁵o⁵⁵ ɿ t⁵⁵ɜ:
 ts⁵⁵ɛ⁵⁵ ɿ ʃ⁵⁵o⁵⁵ ɿ nǎ⁵⁵]
- "Oh, that many... Then which language do you know best?"*
63. A: 最熟, 中国话。
 Zuì shóu, Zhōngguóhuà.
 ['ts⁵⁵ɛi⁵⁵ ʃ⁵⁵u⁴ ɿ t⁵⁵ʃuŋ ɿ k⁵⁵wo²¹ ɿ ʃ⁵⁵o⁵⁵ ɿ]
- "I'm most fluent in Chinese."*
64. B: 哎, 啊?
 Ai, a?
 [ʔ³i²¹ 'ʔɛ:]
- "What?!"*
65. A: 中国话。
 Zhōngguóhuà.
 ['t⁵⁵ʃuŋ ɿ gə²¹ ɿ ʃ⁵⁵o⁵⁵ ɿ]
- "Chinese."*
66. B: 啊, 中国话! 我说得也挺熟阿。说那外国话。
 A, Zhōngguóhuà. Wǒ shuō de yě tíng shóu ā. Shuō nà wàiguóhuà.
 [ʔ³ah hə? 't⁵⁵ʃuŋ ɿ kuə ɿ ʃ⁵⁵ɔa || (..) 'u⁵⁵ɔ: ɿ ʃ⁵⁵wo²¹ ɿ de²¹ ɿ ɛ²¹ t⁵⁵hɪŋ²¹ ɿ ʃ⁵⁵o: ɿ ʃ⁵⁵o⁵⁵ ɿ nǎ⁵⁵ ɿ va⁵⁵ ɿ
 g⁵⁵wo²¹ ɿ ʃ⁵⁵o⁵⁵ ɿ]
- "Ah, Chinese... I can speak that pretty good myself. I'm talking about foreign languages!"*

67. A 熟嘛？最熟可不中国话吗？
 Shóu ma? Zuì shóu kě bù Zhōngguóhuà ma?
 [ʰsou̯mǎ || ʰts̺e:̯ ʃo̯k̺b̺ | t̺ʂɤ̯ ʈgo̯t̺ʃ̺e̯m̺.]
"You said fluent. Wouldn't I be most fluent in Chinese?"

68. B: 外国话哪？
 Wàiguóhuà ne?
 [ʰuæ̯k̺wo̯t̺ʃ̺e̯n̺.]
"And what about foreign languages?"

69. A: 外国话都差不多。
 Wàiguóhuà dōu chàbùduō.
 [ʰa̯i̯k̺o̯t̺ʃ̺e̯t̺ʂ̺ | t̺ʂ̺e̯b̺duo̯]
"Foreign languages are all the same to me."

70. B: 都差不多？
 Dōu chàbùduō?
 [tu:̯ ʰt̺ʂ̺e̯b̺duo̯t̺ʂ̺o̯:]
"All the same?"

71. A: 〈哦〉你要问我会多少句，那我说不上来，我也没统计过，我也没算算，法国话我会多少，德国话我会多少，日本话我会多少，我也没算过账，反正我知道，会点儿。

Ài, nǐ yào wèn wǒ huì duōshǎo jù, nà wǒ shuō bù shàng lái. Wǒ yě méiyǒu tǒngjì guo, wǒ yě méi suànsuan. Fǎguóhuà wǒ huì duōshǎo, Déguóhuà wǒ huì duōshǎo, Rìběnhuà wǒ huì duōshǎo. Wǒ yě méi suàn guo zhàng. Fǎnzhèng wǒ zhīdào, huì diǎnr.

[ʔɛ¹ ni.ɿaʊ.⁰ẽv.⁰oɿ x̥⁵ɛ¹v.⁵tʰɔ̌v.⁵ə.⁵l.⁵dzy:v || ,nɐ:v.⁵ɔɿ 'ɕʷo:v.⁵bʊ.⁵zɛŋv.⁵læɛ¹ ||
,və:ɿmə¹v.⁵tʰɔŋv.⁵dʒɿk.⁰oɿ ,⁵əɿmə¹v.⁵s⁰ɛn.⁵sʷɜ̌v || (.)fɐ.⁵oɿx̥²a.⁵v.⁵x̥⁵ɛ¹v.⁵tʰo.⁵zɛɔɿ ||
'tɕɿ.⁵gu.⁵x̥²a.⁵v.⁵x̥⁵ɿ.⁵tʰo.⁵zɐ:v.⁵ || ,ɿl.⁵b̌ɜ̌.⁵'ɛʷa.⁵v.⁵x̥⁵ɿ.⁵tʰo.⁵ || '⁵əɿ.⁵m⁰ɿɿ
s⁰ɛŋ.⁵gə.⁵ɿdʒɛŋv || ,fɛŋ.⁵dʊm.⁵ɿ.⁵[ɕl.⁵tə.⁵ || 'x̥⁰ɛɿ.⁵tʰɛɿ]

"Right! You have to ask how many sentences I know, and that I can't say. It's not like I've kept score! 'How much French do I know?', 'how much German do I know?', 'how much Japanese do I...' I haven't kept a record, I just know some.

72. B: 那么这个英国话，您会吗？

Nàme zhège yīngguóhuà, nín huì ma?

[nmɿm⁰ɿ.⁵[ɕɛ¹v.⁵kɛ:ɿ.⁵'ɿŋ.⁵k.⁵ɔɿ ,x̥⁵ɐv.⁵nɿ.⁵x⁰ə¹.⁵b.⁵]

"Then what about English? Do you know that?"

73. A: 哈，太会了，太会了。

Hā, tài huì le. Tài huì le.

[xəx̥ 'tʰaev.⁵x̥⁵ɛ:v.⁵lɔ.⁵ || 'tʰɐɿv.⁵x̥⁵ɛɿv]

"Hah, do I ever!"

74. B: 真的？

Zhēnde?

[ʔ[ɕɛn.⁵ta:v.⁵]

"Really?"

75. A: 阿
A.
[ʔa:]
”hmm”
76. B: 那(...)那我跟您学点儿行吗?
Nà... nà wǒ gēn nín xiáo diǎnr xíng ma?
[nəh həʔ || nɛ̃ʋ^uɔ̃ kɛn¹niŋ¹ ʔ¹ɛ̃ⁱɑ̃³ʋ¹ʔ¹ɜ̃¹ɿ¹ | ɛiŋ¹ʋ¹mɛ̃¹]
”Then... can I study with you?”
77. A: 行哇。
Xíng wa.
[ʔɛi²ŋ¹ʋ^uɛ̃¹ɿ¹]
”Sure!”
78. B: 啊?
Á?
[ɦɛ̃¹ɿ¹]
”yeah?”
79. A: 学啊。哪天你学你找我去啊。我教。
Xiáo ā. Něitiān nǐ xiáo nǐ zhǎo wǒ qù a. Wǒ jiāo...
[ʔ¹ɛ̃¹ɑo¹ʋ^oɛ̃¹ || ʔ¹nɛ̃¹ʋ¹t^{hi}ɛn¹ nɪ¹ʋ¹ɛ̃¹ɑo¹ nɪ¹ʋ¹[ʂɑo¹ʋ^oɜ̃¹ʋ¹tɛ̃^yɛ̃¹ || ʋ^oɜ̃¹ʋ¹tɛ̃¹ɔ̃¹ɿ¹]
”Study as much as you want! Any day you want to learn some, you just come and find me. I’ll teach...”
80. B: 哪天?
Něitiān?
[naɛ̃¹ ʔ¹t^hɛ̃^oɑ̃¹ɿ¹]
”Any day?”

81. A: 啊
 Ā.
 [ɦɑː]
 “*hmm*”
82. B: 甬哪天了，就今天在这儿学学。
 Béng nǚitiān le. Jiù jīntiān zài zhèr xiáoxiáo.
 [pɒŋ¹˥˥ nɑɛ˨˩ tʰiɜn˨˩ ɿle˨˩ || tɕʰə˨˩ ɿtɕin˨˩ tʰiɛn˨˩ ɿ tsɑɛ˨˩ ɿtɕʰə˨˩ ɿtɕʰə˨˩ ɿtɕʰə˨˩]
 “*Never mind ‘any day’. I want to learn some here! today!*”
83. A: 在这儿学不行。
 Zài zhèr xiáo bù xíng.
 [tsɑɛ˨˩ ɿtɕʰə˨˩ ɿtɕʰə˨˩ ɿtɕʰə˨˩ ɿtɕʰə˨˩]
 “*That’s not possible.*”
84. B: 怎么？
 Zěnmě?
 [tsɛm˨˩ ɿmɑː˨˩]
 “*Why?*”
85. A: 这儿学不了。
 Zài zhèr xiáo bù liǎo.
 [tɕʰə˨˩ ɿtɕʰə˨˩ ɿtɕʰə˨˩ ɿtɕʰə˨˩]
 “*You can’t study here.*”
86. B: 那怎么？
 Nà zěnmě?
 [nɜ˨˩ ɿtsɛm˨˩ ɿmɑː˨˩]
 “*Why is that?*”

87. A: 我没那么大能耐。我当时把你叫啦，学不了。打听道儿成了，问路打听道儿成了。劳您驾北子湾往哪边走啊，那儿要坐电车，一句话就告诉你了，这你学外国话，一句话就会啦，告诉你了，就行了；不成。

Wǒ méi nàme dà néngnài. Wǒ dāngshí bǎ nǐ jiào huì la. Xiǎo bù liǎo. Dǎtīng dào
chéng le. Wèn lù dǎtīng dào chéng le. Láo nín jià běizi wān wǎng něibiān zǒu a? Nǎ
yào zuò diànchē. Yí jù huà jiù gào su nǐ le. Zhè, nǐ xiǎo wàiguó huà zhè, yí jù huà jiù
huì la, gào su nǐ la, jiù xíng le, bù chéng.

[illegible]

“I don’t have that kind of skill. I can’t teach you just like that... No way! I can teach you to ask for directions... To find your way around and ask for directions. ‘Excuse me, how do I get to Běizi Wān?’ ‘Oh, you just take the train from over there’. Like that, in a single sentence... As if you could I could teach you foreign just like that, in a single sentence! No way!”

88. B: 啊，您哪儿能当时一会儿把英国话全都学会了。

Ā, nín nǎr dāng shí yì huǐ bǎ zhè yīngguóhuà quándōu xiáohuì le

[ʔəh f nɪnʌˈnaɪlɤʔ, tənɹɯŋl iː\χwəʔ pæːʌtʂə, ɪŋɹkʷɔʔχʰe\ ˈteˑɦnɪdə
 ɕʰao\χʰeˑ\ la]

"Then how could you have learnt English in such a short time?!"

⁵⁶ [tʃeʔ] ends on a glottal plosive with no audible release. This segment is somewhat reminiscent of some phonemic syllable finals in other Chinese languages, for example the Wu topolects (Chao, 1967). In this case, it occurs when the speaker interrupts himself.

89. A: 你最快了, 你得学半年。

Nǐ zuì kuài le, nǐ děi xiáo bànnián.

[ni˥˥ tsʰɛi˥ kʰə̌aɛ˥˩ ɲɛ˥˩ ni˥˩ dɿ˥˩ ʃɛ˥˩ ʌo˥˩ ˥pan˥˩ nʰɛ˥˩]

"At the very least you have to study for half a year!"

90. B: 啊, 不是, 我们不学那么多啊, 我就学一两句 呀。

Ā, bú shì. Wǒmen bù xiáo nàme duō a. Wǒ jiù xiáo yì liǎng jù ya.

[ʔə˥˩ ˥pɔ˥˩ tʰɪ˥˩ || ˥om˥˩ ˥pu˥˩ ʃɛ˥˩ ʌ˥˩ nəm˥˩ ˥tʰo˥˩ || ˥ə̌ʔ˥˩ dʰə̌ ʃɛ˥˩ ʌo˥˩ ˥i˥˩ ˥li˥˩ ʌŋ˥˩ ˥dzy˥˩ ˥e˥˩]

"What? No! We don't need to study that much! I just want to learn a sentence or two."

91. A: 那是干吗呢, 学一两句有什么用啊, 也用不上啊。

Nà shì gānmá ne? Xiáo yì liǎng jù yǒu shénme yòng a? Yě yòng bù shàng a.

[˥na˥˩ ˥ʃi˥˩ ʃi˥˩ kəm˥˩ ˥mɛ˥˩ ˥ne˥˩ || ʃɛ˥˩ ʌ˥˩ ˥i˥˩ ˥li˥˩ ʌŋ˥˩ ˥tɛy˥˩ ˥o˥˩ ʃəm˥˩ ˥i˥˩ ʌŋ˥˩ ˥bɛ˥˩
˥ʃi˥˩ ʌŋ˥˩ ˥ɲɛ˥˩]

*"Why on earth do you want that? What good will it do you to learn a sentence or two?
It will be completely useless!"*

92. B: 啊——, 是那样儿啊, 我跟你学, 让您听着我的喉音怎么样, 说这喉音
得能学外国话, 我将来就跟您多学。

Ā... Shì nèiyàng a, wǒ gēn nín xiáo, ràng nín tīngzhe wǒ de hóuyīn zěnmeyàng. Shuō
zhè hóuyīn de néng xiáo wàiguóhuà, wǒ jiānglái jiù gēn nín duō xiáo.

[ʔə˥˩ || ʃi˥˩ ˥naə˥˩ ˥e˥˩ || ˥o˥˩ kən˥˩ ˥nin˥˩ ˥ʃɛ˥˩ ʌ˥˩ ˥ɲɛ˥˩ ˥nin˥˩ ˥tʰiŋ˥˩ ˥dʰə̌ ˥ə̌ ˥tə̌ ˥χʊ˥˩ ˥i˥˩ n˥˩
˥tsəm˥˩ ˥i˥˩ ʌŋ˥˩ || ʃɛ˥˩ ˥tʰɛ̌ ˥dʰɛ̌ ˥χʊ˥˩ ˥i˥˩ n˥˩ ˥de˥˩ ˥nɛŋ˥˩ ˥ʃɛ˥˩ ʌo˥˩ ˥vɛi˥˩ ˥ko˥˩ ˥χʰɛ˥˩ || ˥ə̌ ˥dʰɛ̌ ˥i˥˩ ˥li˥˩ ˥dʰə̌
kən˥˩ ˥nin˥˩ ˥tʰuo˥˩ ˥ʃɛ˥˩ ʌ˥˩]

*"All right, how about this... I'll study with you and let you listen to how I pronounce
the guttural sounds. If my guttural sounds are good enough that I can go on learning
foreign languages, I'll study some more with you in the future."*

93. A: 对。
Duì.
[t^ʰəi²¹]
"Right!"
94. B: 嗯，这喉音不好，我就不学。
Ng, zhè hóuyīn bù hǎo, wǒmen jiù bù xiáo.
[n¹ t͡ɕʰɿ³¹ | 'χ⁰u¹ j¹ɿ²¹n¹ | bu²¹ χa: | 'ɔm²¹ t͡ɕʰə¹ | bu²¹ ɕ'əɔ¹]
"If my guttural sounds are no good, we won't need to study any more."
95. A: 对，对队队，这叫对。
Duì. Duì duì duì. Zhè jiào duì.
[t^ʰəi²¹ || | t^wə²¹ t^wə²¹ t^ʰəi²¹ || | 't͡ɕei²¹ t͡ɕ'əɔ¹ t^wəi²¹]
"Right, right. That's right."
96. B: 这怎么样呢？
Zhè zěnmē yàng ne?
['t͡ɕ'ei²¹ t͡səm²¹ mɛ¹ ɕ'əŋ²¹ na: |]
"How is that?"
97. A: 这个外国话，主要就这音。
Zhège wàiguóhuà, zhǔyào jiù zhè yīn
[t͡ɕe¹ ɕə¹ | 'vae²¹ ɕ⁰o¹ χ⁰ə²¹ | t͡ɕu²¹ ɕə¹ t͡ɕə²¹ t͡ɕə¹ | 'ji²¹n¹]
"The really important thing when studying foreign languages is the pronunciation."
98. B: 啊啊。
A... a
[ʔ²¹ ʔ²¹]
"Yeah... "

99. A: 你字儿对了因不对不行。

Nǐ zì duì le yīn bú duì bù xíng

[ni/ 'tsɜ̌ɹ/ t̪ʰɛ/ lɛ/ jim/ bu/ t̪ʰe/ bu/ ɕiẽ/]

"It's no use saying the right words if you have the wrong pronunciation."

100. B: 哦。

 $\bar{O}_{...}$

[?ɔ:]

"Oh..."

101. A: 有的人学外国话, 学好几年, 跟外国人一说, 外国人不懂, 那怎么回事儿啊? 音不对。

Yǒu de rén xiǎo wàiguóyǔ, xiǎo hǎo jǐ nián, gēn wàiguórén shuō, wàiguórén bù dòng.

Zhè zěnmé huí shì a? yīn bú duì.

[ʔɪˈlʰoʊʤ tɛˌj̥t̚ɕ̟ || ɛˈdoʋ pax̟gwo˧ ʔyːʤ || ˈɛˈdoʋ xeoʋvɪˌniɛn || kəm]

'ba: _gə: _jə:n _ə: _s̥o: _ɿ̃ \| _uɜ: _o̥: _jə:m _bɔ: _don \| _tə: _tsam _mo: _ʔə: _ɿ̃ _ə: _jin

pu1 ,tʊe: \w]⁵⁷

"Some people study foreign, they study it for years even, and when they speak to a foreigner, the foreigner doesn't understand anything. Why is that? The pronunciation is wrong."

102. B: 哦。

 $\bar{O}_{...}$
$$[\text{O:}]$$

"Ah..."

⁵⁷ There are two things worth noting in this last phrase. One is that the final nasal consonant of [ʔjinʔ] does not assimilate the place of articulation of the following consonant. This is probably a result of the syllable being particularly emphasized, and makes it different from other finals like it in this same phrase, such as [kəŋʔ] and [ɬəŋʔ]. Note that the same happens in turn 105. The other peculiarity is the tone of the last syllable [ɬi⁵ɛːw], which is spoken in a very theatrical manner. It could, in fact, also have been analyzed as two distinct syllables, each with a Tone 4. Finally, the vowel of the word [doŋʔ] differs from that of all other /-ong/ finals, which are pronounced [uŋ].

103. A: 音不对。
Yīn bú duì.
[ˈjin¹ bu¹ t̚sɛːʋ]
"The pronunciation is wrong!"
104. B: 那您要教我呢?
Nín yào jiāo wǒ ne?
[ˈnin¹ ˈɛəʋ t̚sɛːʋ ˈɔːʋ naːʋ]
"So what are you going to teach me?"
105. A: 我要教啊，教你，音得对。你学哪话?
Wǒ yào jiāo a, jiāo nǐn... Jīn dēi duì, Nǐ xiáo něiguóhuà?
[ˈɔːʋ ˈɛəʋ t̚sɛːʋ ˈɔːʋ t̚sɛːʋ nɪn¹ || ˈjin¹ t̚sɛːʋ t̚sɛːʋ || niːʋ ˈɛːʋ neːʋ ɲgˈɛʋ]
"I'm going to teach you... What language do you want to learn?"
106. B: 英国话。
Yīngguóhuà.
[ˈjiːŋ ɲgɔːʋ ɲˈɑːʋ]
"English"
107. A: 教你英格兰的正音。
Jiāo nǐ yīnggélán de zhèngyīn.
[t̚sɛːʋ niːʋ ˈjiːŋ ɲgɛːʋ lænˈdɛːʋ t̚sɛːŋ ɲjinˈɲ]
"Then I'll teach you correct Oxford pronunciation"

108. B: 啊? 你瞧? 那你受累吧。

Á? Nín qiáo? Nà nǐ shòulèi ba.

[ʔo: || n.ɿ 'tɕʰe:ɿ || nɛ:ɿ 'niɿ ʃo:ɿ⁵⁸ ɿæ.ɿ ɿpɛ.ɿ]

"Wow! Well, go ahead then!"

109. A: 你学单句儿, 学...学...学是对话?

Nǐ ciáo dānjùr, xiáo... xiáo... xiáo shì duìhuà?

[niɿ 'ɕ'ɑ̃:ɿ || tɕn ɿtɕ'ɕ'ɿ || ɕ'ɑoɿ 'ɕ'ɕ'ɿ || ɕ'ɕ'ɿ ʃɿ ɿt'ɕ'ɿ ɿχ'ɕ'ɿ]

"Do you want to learn simple phrases or dialogues?"

110. B: 啊...对话是什么意思?

A... duìhuà shì shénme yìsi?

[ʔo:ɿ t'wɛɿ ɿχ'ɕ'ɿ ɿzəm ɿmin]

"Ah... What's a dialogue?"

111. A: 对话, 问答题了, 打听道问路啊、买东西、问价钱哪。问答题了。

Duìhuà, wèndāhuà le. Dǎtīng dào wèn lù a, mǎi dōngxi, wèn jiàqián na. Wèndāhuà le.

[t'wɛɿ ɿχ'ɕ'ɿ ɿwən de 'χ'ɑɿ || ɿta ɿt'ɿŋ ɿtao ɿwən ɿlu: ɿvɛ.ɿ || maeɿ 'tuŋ ɿɕ'ɿ || ɿwən]

'tɕ'ɕ'ɿ ɿtɕ'ɕ'ɿ:ɿ ɿnɿ || ɿwən ɿta: ɿχ'ɕ'ɿ]

"Dialogues are questions and answers. Asking for directions and finding your way around, buying things, asking about prices; questions and answers."

⁵⁸ Here, the diphthong of /shou/ is reduced to a monophthong, something which happens quite frequently with this syllable final. It seems similar to what can be observed to happen with final /ai/ in the following syllable, and final /iao/ in the preceding phrase:

/ou/	→	[o:]
/ai/	→	[æ:]
/iao/	→	[ɿ:]

Note that, in this material, off-glides are often dropped in rapid speech (as shown here), but on-glides nearly always kept.

112. B: 哦(...) 那不行了, 太麻烦。

O... Nà bù xíng le, tài máfan.

[ʔɔ:ʔ nʌʔ pɕʔ ʔɕiŋʔ ʔɕɛʔ || tʰɛ:ʔ ʔmaʔ fɜnʔ]

"Oh, let's not do that. It's too much hassle."

113. A: 对话。

Duìhuà.

[tʰweʔ ʔχʔɐ:ʔ]

"Dialogue..."

114. B: 单句是什么?

Dān jù shì shénme?

[ʔtanʔ ʔdʒʔɐ:ʔ ʔʃʔ.ʔ ʔzʌmʔ ʔmʌʔ]

"What are simple phrases?"

115. A: 单句就是问生字了, 就是名词了: 桌子怎么说啊, 您请坐怎么说呀, 再见怎么说呀。

Dān jù jiùshì wèn shēngzì le, jiùshì míngcí le. Zhuōzi zěnmē shuō a, nín qǐng zuò zěnmē shuō ya, zàijiàn zěnmē shuō ya.

[ʔtɜnʔ ʔdʒʔɐ:ʔ tɕʔo:ʔ ʔʃʔ.ʔ ʔɕɐnʔ ʔʃɕɐŋʔ ʔtɕʔɜʔ ʔɕɛʔ || tɕʔɕʔ.ʔ ʔmɪŋʔ tɕʔɜʔ ʔɕɛʔ || ʔʃʔuoʔ ʔtɕʔɜʔ

ʔtɕʔɜmʔ ʔmɜ:ʔ ʔʃʔɕʔ:ʔ ʔɐ:ʔ || nʔ tɕʔɜʔɜŋʔ ʔtɕʔɕʔ:ʔ ʔtɕʔɜmʔ ʔmɜ:ʔ ʔʃʔɕʔ:ʔ ʔɐ:ʔ || tɕʔɜʔ ʔtɕʔɜʔ ʔtɕʔɜmʔ

ʔʃʔɕʔ:ʔ ʔɐ:ʔ]

"Simple phrases are basically just words; like nouns. How do you say 'table'. How do you say 'please have a seat', how do you say 'goodbye'..."

116. B: 啊。。。。

O...

[ʔɔ:ʔ]

"Oh..."

117. A: 就是这个。
 Jiùshì zhège.
 [tɕəʔ¹ 'tɕəʔ. gəʔ¹]
"Things like that."
118. B: 啊，一句一句的了。
 A, yí jù yí jù de le.
 [ʔəʔ¹ 'i:ʔ¹ tɕy¹:ʔ¹ dʒy.¹ təʔ¹.ləʔ¹]
"Ah, so it's one sentence at a time?"
119. A: 对，这是对，对了。
 Duì. Zhè shì duì... Duì le.
 ['tʷe:ʔ¹ tsəʔ¹ | tʷeɪʔ¹ || 'tʷe:ʔ¹.ləʔ¹]⁵⁹
"That's right, that's right."
120. B: 那我学单句儿了。
 Nà wǒ xiáo dānjùr le.
 [nəʔ¹ ʔ¹ əʔ¹ ɕ'ɑʔ¹ 'tan¹ ʔ¹ tɕyʔ¹.ləʔ¹]
"Then I'll learn some simple phrases."
121. A: 学单句儿也行。
 Xiáo dānjùr yě xíng.
 [ɕəʔ¹ 'tæn¹ ʔ¹ tɕy¹ ʔ¹.ləʔ¹ ɕi¹ ɣ¹]
"You can study simple phrases if you want."
122. B: 那这行。
 Ā zhè xíng.
 [ʔəʔ¹.dʒɛʔ¹ 'ɕ'ʒ¹]
"Then that's what we'll do."

⁵⁹ Note that final /ei/ drops the off-glide in some syllables, just like finals /ou/, /iao/ and /ai/ (see footnote 58).

123. A: 学单句儿也成。
 Xiáo dānjùr yě chéng.
 [ɣəːl 'tɕɿn ɿ ʈe˥əːl ʈɕʰɿ̃ʌ˥]
"It's perfectly all right to study simple phrases."

124. B: 那您受累吧。
 Nà nín shòu lèi ba.
 [naːl 'nĩ ɿ ɕəːl ʌeːːl ɿ ʌ˥˥]
"Please, go ahead then!"

125. A: 嗯。
 Ng...
 [ŋ]
"Eh..."

126. B: 啊，我（...）我学这句吧。
 A, wǒ... Wǒ xiáo zhè jù ba.
 [ʔəːl || ʊː˥˥ || ʊ˥˥ ɕʰɿ̃ʌ˥ ɿ ʈɕʰɿ̃ʌ˥ ʈɕʰɿ̃ʌ˥ ʌeːːl]
"Ok, how about this..."

127. A: 别胡来，别胡按啊。
 Bié hú lái, bié hú'àn a.
 [n || p'əː˥ 'xuː˥ ʌeː˥ || p'ə˥ 'xuː˥ ɿ̃ʌ˥ ʔ˥˥]⁶⁰
"Don't run wild now. Don't mess it up."

⁶⁰ Neutral tone syllables with the zero-initial (like the /a/ in this phrase), are assumed by Duanmu (2000: 83) to assimilate the final of the preceding syllable, or to be pronounced with an initial [j] or [ɦ]. In this material, some syllables of this kind also receive an initial glottal plosive [ʔ], which seems to give them extra emphasis and also creates a kind of juncture between them and the preceding syllable.

128. B: 别胡按?

Bié hú'àn?

[p^hé:ʔ 'xu:ʔãʔ]

"Mess it up?!"

129. A: 哦，你打算学，打算揭我。

Ō, nǐ dǎsuàn xiáo, nǐ dǎsuàn jiē wǒ?

[ʔəʔ || niʔ | tɕ:ʔɿʔ | ɕ'ɛʊnʔ | tɕ:ʔənʔ | 'tɕɿ:ʔɕ'ɔ:ʔ]⁶¹

"That's right. Are you planning to study or are you planning to cause trouble for me?"

130. B: 你瞧?

Nǐ qiáo?

[niʔ 'tɕ^hɛ:ʔ]

"Excuse me?!"

131. A: 啊?

A?

[ɛ:ʔ]

"Huh?"

132. B: 打算跟你学啊。

Dǎsuàn gēn nǐ xiáo a!

[tɕ:ʔ sɿʔ gənʔ niʔ 'ɕ'aoʔɕ'ɑ:ʔ]

"I'm planning to learn from you!"

⁶¹ Notice all the variants of the word 打算 dǎsuàn "to plan" that appear from turn 129 to 133. What they all have in common is the heavy reduction of the final syllable suàn.

136. B: 我没法儿学。什么叫按照规矩问哪？这英国书没念过，没学过英国话，头一句我们先问什么这么不知道。

Wǒ méi fǎn xiào. Shénme jiào ànzhào guījù wèn na? Zhè yīngguóshū méi niàn guo, méi xiào guo yīngguóhuà, tóu yī jù wǒmen xiān wèn shénme zhème bù zhīdào.

[^uɔɭ 'me:ɿ faʒɿ ɿɿa:ɿ ||⁶² (1.71) || ʃəmɿ muɿl dʒ'ɐʒ ɿ anɿ dʒɔɿ 'kur:ɿ tɐy^uɿ ɿɐnɿ naɿ ||

(0.68) $\parallel \text{tʃe}^1 \text{ } ^1\text{j}^{\text{ə}} \eta \text{ } ^1\text{k}^{\text{w}}\text{o} \text{ } ^1\text{s}^{\text{u}} \text{ } ^1\text{m}^{\text{e}} \text{ } ^1\text{i} \text{ } ^1\text{n}^{\text{i}} \text{ } ^1\text{x}^{\text{ə}} \parallel \text{ } ^1\text{m}^{\text{e}} \text{ } ^1\text{i} \text{ } ^1\text{e}^1 \text{ } ^1\text{a}^1 \text{ } ^1\text{g}^1 \text{ } ^1\text{i} \eta \text{ } ^1\text{g}^1 \text{ } ^1\text{x}^{\text{ə}} \parallel$

⁶³ ^htoʊʌ yɿ dʒ uɿ əmɿ ɪɿ ɛi̯ ɿ wæ̯ ɿ sɛmɿ məɿ ɿ tɕɛmɿ pɔɿ ɿ dɑ̯ ɿ

"I can't do this... Study according to what rules?! I have never learnt to read or speak English. I don't even know what phrase I'm going to ask about first!"

137. A: 你按照规矩问就成啊。

Nǐ ànzhào guījù wèn jiù chéng la.

[nɪl, ʁen\, tʃə.ɿ 'kʰi: ʔ tɕʰɔ.ɿ, ʁen\ dʒɔ.ɿ, tʃʰən\ ɿlɔ:ɿ]⁶⁴

"It's no problem as long as you follow the rules."

138. B: 什么叫规矩，这头一句先问什么？

Shénme jiào guījù zhè tóu yī jù xiān wèn shénme?

[l̥səmʈ̚mɪl tɕʰəʋ ʔgʷɪː ʈ̚tɕɪːl dʒəːl̥həʋʈ̚ɪːl tɕɪːʋ ɛɪʑ ʈ̚vænʋsəmʈ̚maːʋ]

"What rules are you talking about?! What's the first phrase?!"

139. A: 不是头一句。外国话头一句是什么干嘛?

Bú shì tóu yī jù. Wàiguóhuà tóuyījù shì shénme gànma?

[pu:1 sə:l tʰou1 j i:1 tɕe:1 || vaeg1 χʷe:1 tʰoʊ1 y i:1 tɕe:1 sɿ:1 sɿm1 mɔ:1 kam1 mɛ:1]⁶⁵

"There is no 'first phrase'. Why on earth would you want to learn how to say 'first phrase' in foreign?"

⁶² [ɛɪɑ:ɹ] behaves here like a tone 3. Apparently it even induces tone sandhi in the preceding tone 3.

⁶³ Notice how the syllable /guĩ/ in these last three turns has a different segmental realization from /duĩ/. This is probably due to the tone difference.

⁶⁴ Nin the syllable [ʔənʌ], the zero-initial is realized as an uvular fricative.

⁶⁵ The last two syllables of turn 139 are interrupted by speaker B, and are barely audible. The phonetic transcription of these is therefore only half reliable.

140. B: 那您说按照规。。。。

Nà nín shuō àn guī...

[nɜːl ˌnɪnˈʃʊːɔːŋ ʔanˈkʰɪːŋ]⁶⁶

"But you talk about rules..."

141. A: 不是头一局，你按照规矩问，你别胡来。外国得有的这个东西，有这个名词儿，它有这个名词的，你问我，我可能说。根本连外国人都没这么句话，你问我，那我也，我怎么说，我也说不上来啊。

Bú shì tóu yī jù. Nǐ ànzhào guījù wèn, nǐ bié hú lái. Wàiguó děi yǒu de zhège dōngxi, yǒu zhège míngcí. Tā yǒu zhège míngcí de, nǐ wèn wǒ, wǒ kěnéng shuō. Gēnběn lián wàiguórén dōu méi zhème jù huà, nǐ wèn wǒ... nà wǒ yě... wǒ yě zěnmē shuō? Wǒ yě shuō bù shàng lái ya.

[(XX)]⁶⁷ ˈtʰoʊ˥˥ ɿ˥˥ ˌtɛy˥˥ || nɿ˥˥ ˌkʰɛn˥˥ ʃʊ˥˥ ˈkʰɪːŋ ˌtɛy˥˥ ˌpən˥˥ || nɿ˥˥ pɿɛ˥˥ ˈχu˥˥ ˌlaɛ˥˥ || dz

ˌvaɪ˥˥ ˌkʰo˥˥ ˌdeɪ˥˥ ˈoʊ˥˥ ˌtɛ˥˥ ʃʊ˥˥ ˌtuŋ˥˥ ˌtɕei˥˥ ˈoʊ˥˥ ʃʊ˥˥ ˌgu˥˥ ˌmɿ˥˥ ˌtsʰɜ˥˥ || tʰa˥˥ ˈoʊ˥˥

ʃʊ˥˥ ˌgʊ˥˥ ˌmɿ˥˥ ˌtsʰɪ˥˥ ˌtɛ˥˥ || nɿ˥˥ ˈtʰɛ˥˥ ˌɔ˥˥ || ˈo˥˥ ˈkʰə˥˥ ˌnəŋ˥˥ ˌʃʊ˥˥ || ˌkəm˥˥ ˌbɛ˥˥ ˌlʰɛ˥˥

ˌva˥˥ ˌgʊ˥˥ ˌtɕɛn˥˥ ˌdo˥˥ ˌmeɪ˥˥ ˌzəm˥˥ ˌdzu˥˥ ˌχʰɛ˥˥ || nɿ˥˥ ˈtʰɛ˥˥ ˌtɕa˥˥ || (1.29) || na˥˥ ˈu˥˥ ˌɜ˥˥ || ˈtʰɛ˥˥

ˌtsəm˥˥ ˌʃʊ˥˥ || ˈu˥˥ ˌtɕa˥˥ ˌʃʊ˥˥ ˌbʊ˥˥ ˌtɕa˥˥ ˌlæ˥˥ ɛ˥˥]

"There is no 'first phrase'. You just follow the rules and don't mess things up. You have to ask about things and word that exist in other countries. If you ask about a word that they have over there, then I can teach you. If you ask me about something even the foreigners don't have a word for, then how can I know what it is?!"

142. B: 着外国有没有的话？

Zhè wàiguó yǒu méiyǒu de huà?

[tʃʌ˥˥ ˈvaɛ˥˥ ˌgʊ˥˥ ˈo˥˥ ˌmɛɪ˥˥ ˌo˥˥ ˌdʌ˥˥ ˌxʷɛ˥˥]

"Do the foreigners have words that don't exist?"

⁶⁶ Turn 140 is barely audible due to background noise. The phonetic transcription is only half reliable here.

⁶⁷ The (XX) stands for a section which is so difficult to hear that I have not managed to give a phonetic transcription.

143. A: 那当然哪。你问我，外国人，吃爆肚，怎么说，那怎么说啊？
 Nà dāngrán la. Nǐ wèn wǒ, wàiguórén, chī bào dǔ, zěnmē shuō. Nà zěnmē shuō a?
 [nɜ́ɔ ʔaŋ ʔɛ́n lɛ: ʔ⁶⁸ || nɪ ʔv̥ʂ ʔo ʔ || (0.52) || ʔv̥ɛ ʔgɔ ʔɛ́n ʔ (0.73) ʔɕʰl̥p̥ ʔpɑ ʔv̥ ʔtuə ʔ ||
 (0.52) || tsum. ʔmə ʔɕʰɔ: ʔ || (1.87) || ʔnɜ́ ʔtsɐm ʔɕʰɔ: ʔ]
"Of course they have! If you ask me how foreigners say... 'Baodu'. How can I answer that? What would I say then?"⁶⁹
144. B: 哦，对了。
 Ō, duì le.
 [ʔɔ: ʔv̥ɛ: ʔlɜ.]
"Oh, right..."
145. A: 外国人那儿有爆肚啊。豆汁儿，这我也不会呀。豆。。。
 Wàiguórén nǎr yǒu bàodǔ a. Dòuzhī, zhè wǒ yě bú huì ya. Dòu...
 [ʔv̥ɛ ʔgɔ ʔɛ́n ʔn̩ ʔɑ ʔpɑ ʔv̥ ʔtɔ ʔv̥ ʔ⁷⁰ || (0.87) || tou ʔdʒɛ ʔ || (0.90) || ʔɕo ʔ ʔo ʔ ʔɜ ʔbu ʔ
 ʔx̥ ʔɛɪ ʔv̥ ʔ || tou ʔ]
"Do you really think foreigners have 'baodu'? Or 'douzhi'? Even I wouldn't know that."⁷¹
146. B: 反正，普通有的东西。
 Fǎnzhèng, pǔtōng yǒu de dōngxi.
 [fɑŋ ʔtɕʰɜŋ ʔpʰu ʔtʰɔŋ ʔ ʔv̥ɔ ʔtɪ ʔtuŋ ʔɕɪ ʔ]
"Well, normal stuff then."

⁶⁸ This syllable [lɛ:] is very drawn out and long, despite neutral general presumption that the duration of neutral tone syllables is "relatively short" (Chao, 1968: p. 35).

⁶⁹ Bàodǔ 爆肚 is a Chinese dish of fried tripe.

⁷⁰ The half tone 3 and the following neutral tone in [ʔv̥ ʔv̥] together forms a tonal contour very similar to the citation form of tone 3 [ɤ].

⁷¹ Dòuzhī 豆汁: A drink made from fermented soy beans.

147. A: 你得有的。

Nǐ děi yǒu de.

[ni˩ tɕei˩ ʰɔ̌˩ tɕʰe˩] ⁷²

"It has to be words that exist."

148. B: 哦，那... 那成。

Ō, nà... nà chéng.

[ʔɔ: || nɐ˩ nɐ˩ ʰtɕʰɛ̌:]

"Ok... That's all right."

149. A: 嗯只要有这/着就成。

Ēn, nǐ zhǐ... zhǐ yǒu zhe jiù chéng.

[ɛ˩ ʰni˩ tsɿ˩ || tsɿ˩ ʰɔ̌˩ tɕʰə˩ dʒə˩ ʰtɕʰɛ̌]

"As long as they have it, it's ok."

150. B: 啞...，啊，我学这句儿吧，这英国人管这个帽子叫什么？

Ā... Ā, wǒ xiáo zhè jù ba. Zhè Yīngguó rén guǎn zhè ge màozi jiào shénme?

[ʔə: || ʔʌ˩ ɔ˩ ɛ̌ ʔ dʒɛ˩ dʒɛ˩ bʌ˩ || tɕʰe˩ jɿŋ˩ kʰə˩ tɕʰən˩ kʰən˩ tɕʰeg˩ ə˩ ʰmao˩ tsɿ˩ tɕʰa˩ ʰ
ɕʌm˩ mɛ̌]

"Let's study this word then: How do foreigners say 'hat'?"

151. A: 嗯？

É?

[hɛ̌]

"What?"

⁷² The tone of [ni˩] is difficult to hear clearly, and the sound quality allows no reliable instrumental analysis. It may also be [ɿ].

152. B: 管这个帽子叫什么?

Guǎn zhège màozi jiào shénme?

[_lkʷə̌ŋʅ dʒɛkʅkaʅ 'maoʅtsɿ.ʅ tɕiɑʊʅ]⁷³

"How do they say 'hat'?"

153. A: 先学别的, 先学别的, 先学别的。

Xiān xiáo bié de, xiān xiáo bié de... xiān xiáo bié de.

[_lɕiẽŋ ɕʰɑ̌ŋ 'biɛ:ʅtəʅ || _lɕiẽŋ ɕʰɑ̌ŋ 'biɛ:ʅtəʅ || (0.98) || ɕĩŋ ɕʰɑ̌ŋ 'biɛdʅ]

"Let's do something else first. Let's do something else."

154. B: 啊?

Á?

[ɤv:]

"Huh?"

155. A: 先学别的。

Xiān xiáo bié de.

[ɕĩŋ ɕʰɑ̌ŋ 'piɛ:ʅtəʅ]

"Do something else first."

156. B: 怎么?

Zěnmé?

[ʰtsʌmʅmɛ:]

"Why?"

⁷³ There may be a "shénme 什么" at the end of this turn, but it is difficult to hear as he is being interrupted by speaker A.

157. A: 告诉你先学别的。要不怎么这种人可很呢，你这玩儿意。
 Gàosu nǐ xiān xiáo bié de. Yàobu zěnmē zhègè rén kěhèn nē, nǐ zhè wányìr.
 [ḡaʊ̯ʌ̌n̩i̯, ɛ̌'ě: ɿ ɛ̌'ʌ̌ʊ̯ ɿ 'pɿɛ: ʌ̌də ʌ̌ || (0.667) || ɿjaʊ̯bʌ̌ tsəmʌ̌mə̌ ʌ̌tʂɛi̯ḡʊ̯ʌ̌ʌ̌ʂʌ̌ 'kʰɿ: ʌ̌
 xənʌ̌nɛ || ʌ̌niʌ̌ḡə̌ 'ʊ̯ɿʌ̌]
"I'm telling you to do something else! That's really why I hate this guy..."
158. B: 哦?
 Ó?
 [ʔo:]
"What?"
159. A: 你好心好意地叫他外国话，他胡来，你这是胡来，你这是胡来了。胡来!
 Nǐ hǎoxīn hǎoyì de jiào rén wàiguóhuà de, húlái. Nǐ zhè shì hú lái, nǐ zhè shì hú lái le.
 Húlái.
 [ʌ̌ni: ʌ̌χa: ʌ̌i̯ ʌ̌χǎʌ̌ ɿi̯: ʌ̌ðe. ʌ̌tɕiǎʌ̌ ʌ̌vaɛ̌ ɿ ʌ̌χʷɛ: ʌ̌ðə ʌ̌xuʌ̌ ʌ̌æ: ʌ̌ || ʌ̌niʌ̌ tʂə̌ ʌ̌xuʌ̌ ʌ̌'æ: ʌ̌ || ʌ̌niʌ̌
 tʂə̌ ʌ̌xuʌ̌ ʌ̌'æ: ʌ̌lə̌ || (1.47) || huʌ̌ ʌ̌'æ: ʌ̌]⁷⁴
"You try and teach him foreign with the best of intentions... You're messing it up. All you do is cause trouble!"
160. B: 这帽子这。
 Zhègè màozi zhè...
 [ə̌ ʌ̌tʂə̌ ʌ̌'mǎʌ̌ tsɿ. ʌ̌dʒɛ̌ (X)]
"So a hat..."

⁷⁴ It appears speaker B is talking in speaker A's mouth at the beginning of this turn, but I have not been able to make out what he says.

170. B: 嗯, 啊? 这帽子没法儿说啊?
 Āi, á? Zhè màozi méi fǎr shuō a?
 [ʔəiʔ ʔeːʔ || (0.99) || ʔəːdzɿː ʔmaːdzɿː meːʔfəːʔ ʔɕɔːʔ]
"What? There is no way to say 'hat'?"
171. A: 帽子多了, 草帽、皮帽子、毡帽子, 说哪个, 哪个对呀。
 Màozi duō le. Cǎomào, pímaozi, zhānmaozi, shuō něige, něige duì ya?
 [ʔmaːdzɿː ʔtʰoːʔleː || ʔtsʰaːdzɿː ʔmaːdzɿː pʰiːʔ ʔmaːdzɿː || ʔɕemʔ maːdzɿː || ɕʷoʔ
 ʔnɛiʔ kɐː || nɛiʔ ʔtʰɛːʔɕɔːʔ]
"There are all kinds of hats. Straw hats, leather hats, felt hats... Which one do you want me to say? Which one is the right one?"
172. B: 哦 (...) 啊, 那就甬分了, 就这普通, 这就帽子。
 O... Ā, nà jiù béng fēn le. Jiù zhè pǔtōng, jiù zhè màozi.
 [ʔɔː || ʔeː naː ʔeɪʔ ʔɔŋʔfɛːʔɕɔː || ʔɛːdzɿː ʔpuːʔtʰoŋʔ ʔtɛːʔtɛːʔ ʔmaːdzɿː]
"Ah... Well, don't separate them. Just an ordinary hat."
173. A: 就是帽子啊?
 Jiù zhè màozi a?
 [ʔtɛːʔəː ʔmaːdzɿː]
"Just a hat?"
174. B: 嗯。
 Ai.
 [ʔeɪʔ]
"Yes."

175. A: 哪国话?

Něiguóhuà?

[¹nɛi ɿ gɔʊ²¹ ʅ⁵⁵ ɛː ʋ]

"Which language?"

176. B: 英国话

Yīngguóhuà.

[¹jɪŋ⁵⁵ ɿ ŋɔ²¹ ʅ⁵⁵ ɤː ʋ]

"English"

177. A: 英国话?

Yīngguóhuà?

[¹jɪŋ ɿ ŋɔ ʅ⁵⁵ ɤː ʋ]

"English?"

178. B: 嗯。

n.

[m]

"Yeah."

179. A: 帽子。外国人说话用喉音。

Màozǐ a? Wàiguórén shuō huà yòng hóuyīn.

[¹mao⁵⁵ ɿ dzɿ.ɿ || ¹ʋaɛ⁵⁵ ɿ gɔʊ²¹ ʅ⁵⁵ ɤː ʋ ɿ ʋɔŋ⁵⁵ ʅ⁵⁵ ɤː ʋ ɿ ʃou²¹ ɿ jin ɿ]

"Hat... Foreigners use throat sounds when they talk."

180. B: 啊，啊。

A, a.

[¹ʔɔ⁵⁵ ɿ ʔɔ.ɿ]

"Yeah, yeah."

181. A: 咱们中国话儿用丹田音。
Zénmenr Zhōngguóhuà shì yòng dāntián yīn.
[ʈsəmʅmǎʅ ʈʂŋ ʈʂŋʊʅx^oɐ:ʅǎ.ʅʅŋ ʈtan ʈ tʰɛnʅʅ'in ʈ]
"We Chinese use sounds from our abdomen."
182. B: 是。
Shì.
[ʂʂʅ]
"Right."
183. A: 它是用喉音。
Tā shì yòng hóuyīn.
[tʰɜ:ʅzǎʅ ʈŋʅʅʅ^oʊʅʅin ʈ]
"They use throat sounds."
184. B: 啊?
Á?
[ʔɐʅ]
"Ok."
185. A: 帽子?
Màozi?
[ʅmaoʅtsɿʅ]
"Hat?"
186. B: 嗯。
Āi.
[ʔɛɪʅ]
"Yeah."

191. A: 帽子啊?

Màozǐ ā?

[¹ma⁵⁵dzɿ⁴ a:¹]

“*Hat?*”

192. B: 噢——

Ō...

[ʔɔ:¹]

“*Hmmm...*”

193. A: 外国话用喉音。

Wàiguóhuà yòng hóuyīn.

[¹ʋae⁵⁵ɣɔ:¹χ⁰a:⁵⁵ i⁵⁵ŋ⁵⁵ ¹χou⁴ɿn¹]

“*Foreigners use throat sounds.*”

194. B: 那您只说吧。

Nà nín zhǐ shuō ba.

[¹na¹ ¹nɿn¹ tɕʰɿ¹ ¹ʂ^wɔ¹ bə:¹]

“*Oh, just say it already.*”

195. A: 啊。。。帽帽。外国话这音，就这音，(XXX)。漫漫儿来，这得漫漫儿来，刚学，初学乍练，先学这好学的。

196. B: (XXX) 这个帽子叫什么?
(XXX) zhège màozi jiào shénme?

- 197: A: 帽帽。
Māomāo.
[¹mao¹ moʔ¹]
“*Maomao*”

⁸¹ Something is said here, but because of the applause and laughter it is difficult to hear exactly what.

199. A: 有门儿，行。

Yǒu ménr, xíng.

[¹o⁵⁵ ʈ⁵⁵ mǎn˥˥ || ɛ˥˥ŋ˥˥]

“That’s right! You’re getting the hang of it!”

200. B: 啊。

Ā...

[ã:]

“Uh...”

201. A: 先记得啊。

Xiān jìzhe a.

[ɛ˥˥n˥˥ ˥˥tɕi˥˥ ʈ⁵⁵ʈ⁵⁵ɛ˥˥ | ʔɐ˥˥]⁸³

“First you have to remember that!”

202. B: 啊，啊。

A a.

[˥˥ʔɛ˥˥ ʈ⁵⁵ɦɛ˥˥]

“All right, all right.”

203. A: 先记着，刚学啊，初学，初学乍练，先学这好学的。先别到外头说着啊。

Xiān jìzhe. Gāng xiáo a, chū xiáo, chū xué zhà liàn, xiān xiáo zhè hǎoxiáo de. Xiān bié dào wàitóu shuōzhe a.

[ɛ˥˥ʈ⁵⁵n˥˥ ˥˥tɕi˥˥ ʈ⁵⁵ʈ⁵⁵ɛ˥˥ || kɑŋ˥˥ ˥˥ɛ˥˥ɑ˥˥ ʈ⁵⁵ɛ˥˥ || ˥˥ʈ⁵⁵ʈ⁵⁵u˥˥ ˥˥ɛ˥˥ɑ˥˥ || ˥˥ʈ⁵⁵ʈ⁵⁵u˥˥ ˥˥ɛ˥˥ʈ⁵⁵˥˥ ˥˥ʈ⁵⁵ʈ⁵⁵ɛ˥˥ ˥˥ɦɛ˥˥˥˥ ||

˥˥ɛ˥˥ʈ⁵⁵ɛ˥˥ʈ⁵⁵ʈ⁵⁵˥˥ ˥˥kɑ˥˥ ˥˥ɛ˥˥ʈ⁵⁵ʈ⁵⁵ɛ˥˥ ˥˥dɔ˥˥ || (5.30) || ˥˥ɛ˥˥n˥˥ ˥˥pɿɛ˥˥ ˥˥tɑ˥˥ ˥˥vɑɛ˥˥ ˥˥tɕi˥˥ ˥˥ʂ˥˥o˥˥ ˥˥ʈ⁵⁵ʈ⁵⁵ɛ˥˥ ˥˥ʔɐ˥˥]

“First you have to remember that. You just started, you’re just a beginner. First you have to study these simple things. ... Whatever you do, don’t let anybody hear you say it!”

⁸³ The syllable [ʔɐ] illustrates why the neutral tone is not analysed simply as an unstressed syllable. [ʔɐ] is a neutral tone and is also stressed.

208. B: 嗯。。。哎
O... Ai.
[ʔɔːɿ ʔɜɿɿ]
“Oh...”
209. A: 别说，到哪儿也别说啊，千万别到外边儿说，记着了，千万别说。
Ài. Bié shuō, dào nǎr bié shuō a. Qiānwàn bié dào wàibianr shuō. Jìde le, qiānwàn bié shuō.
[ʔeːɿ ʔɿʷo ʔ ʔaːɿ ʔnæɿ ʔeːɿ ʔɿʷɔː ʔ ʔeːɿ ʔ ʔeːɿ ʔ van ʔeːɿ ʔtaoːɿ ʔvæɿ ʔpəːɿ ʔɿʷoː ʔ ʔ ʔeɿ ʔtʰɿ ʔlɿ ʔ ʔeːɿ ʔvɿ ʔbʰəːɿ ʔɿʷoːɿ]
“That’s right! Don’t say anything. Wherever you go, whatever you do, just don’t say anything! Remember that! Whatever you do, don’t start speaking!”
210. B: 哦... 哦——，嗯？我们穿的这个袜子叫什么？
Ò, Ái wǒmen chuān de zhège wàzi jiào shénme?
[ʔɔːɿ ʔ ʔæɿɿ ʔ ʔom ʔ ʔɿʰvən ʔnɿ ʔ ʔzɛk ʔkɜːɿ ʔvaːɿ ʔdzɿɿ ʔdzʰəɿ ʔɿmɿɿ]
“Oh... But hey, what do you call ‘a sock’? You know, the kind that you put on your feet.”
211. A: 袜袜。这玩儿一其实么嘛，这玩意儿没嘛。
Wāwà. Zhèwányìr qíshí méi ma, zhè wányìr méi ma.
[ʔeːɿ ʔ ʔeːɿ ʔ ʔɿʰvən ʔnɿ ʔ ʔzɛk ʔkɜːɿ ʔvaːɿ ʔdzɿɿ ʔdzʰəɿ ʔɿmɿɿ]
“‘Wawa’. There’s nothing to it really! You can see now that there’s nothing to it.”
212. B: 啊。
A...
[ʔaːɿ]
“Oh...”

218. B: 哦，那个包子叫包包儿？那我们要吃那个饺子呢？叫饺饺，是吗？
 O, nàge bāozi jiào bāobaor? Nà wǒmen chī de nège jiǎozi ne? Jiào jiǎojiao, shìma?
 [ʔəːʔ ˌnɛg˥˩ ˌgə˥˩ ˈpɑʊ˥˩ ˌdzɛ˥˩ ˌtɕʰɑʊ˥˩ || ˈpɑʊ˥˩ pɑ˥˩ || (2.34) || nɬm˥˩ ˌmə˥˩ ˌʔɕʰ ˌtɕʰɿ˥˩ ˌnɛg˥˩ gɛ˥˩
 ˈtɕʰiɑʊ˥˩ ˌtsɛ˥˩ ˌnæ˥˩ || tɕʰɑʊ˥˩ ˈtɕʰɑʊ˥˩ tɕʰɛ˥˩ || (1.39) || ˈʂɿ˥˩ mɛ˥˩ː˩]
“So a ‘steamed bun’ is called a... ‘baobao’... Then what about a dumpling? I suppose it’s called a ‘jiaojiao’... right?”
219. A: 嗯，行啊，这外国话你有点儿啊
 Ai, xíng wa, zhè wàiguóhuà nǐ yǒu diǎnr a.
 [ʔəːʔ ˌɕiŋ˥˩ ˌʔə˥˩ ˌvɑɛg˥˩ ˌχʰɛ˥˩ ˌnɿ˥˩ ˈoʊ˥˩ tɕʰɿ˥˩ ˌtɕʰɿ˥˩⁸⁷]
“Hey, that’s right! You do know some foreign!”
220. B: 我呀！
 Wǒ ya.
 [ʔəːʔ ˈuɑ˥˩ ˌʔa˥˩]
“Who, me?”
221. A: 不对了吗？
 Bú duì le ma?
 [bʊ˥˩ ˈtʰɛi˥˩ mɛ˥˩ː˩]
“What you just said was right, wasn’t it?”

⁸⁷ The syllable [ɕʰ] is one of the few cases in this data set where r-coloring of a vowel seems to work progressively.

222. B: 这叫外国话？这像我们家那小孩儿刚学话。那个小孩儿指那东西叫不明白，管那帽子就叫“哇，我要那帽帽儿；哇，我穿那袜袜；挖，我吃那包包”这叫外国话，这叫？

Zhè jiào wài guó huà, zhè jiào? Zhè xiàng wǒmen jiā de xiǎoháir gāng xiào huà. Nèige xiǎoháir zhǐ nà dōngxī jiào bù míngbai, guǎn zhè màozi zhǐ jiào ‘wa, wǒ yào nà mào maor’, ‘wa, wǒ chuān nà wàwar’, ‘wa, wǒ chī nà bāobaor’. Zhè jiào wàiguóhuà, zhè jiào?

[illegible]

“Is this really foreign? It’s like when my kid just learnt to speak. He pointed at things but he didn’t know what to call them. So he called pointed at the hat and said ‘wah, give maomao’, ‘wah, put on wawa’, ‘wah, eat baobao’! Is this really foreign?”

223. A: 漫漫儿来呀!

Mànmānr láí ya.

[mæm\,mʒʔ 'lae\,ea\]

“You have to take it slow, remember?!”

224. B: 啊。

A.

[?e:ʌ]

“Yeah...”

225. A: 先记着。

Xiān jìzhe.

[e¹en ɿ 'tɕi: _ tʃʌ.ɿ]

“First make sure you remember it!”

226. B: 是啊。

Shì a.

[¹ʂɿ: ʋə: .]

“*Right...*”

227. A: 啊，千万别到外边儿说着啊。

Ai, qiānwàn bié dào wàibiānr shuōzhe a.

[ʔæɪ ʔ || ¹tɕ^hæŋ ʈvã: ʋ p¹e¹ʈtã^ʋ ʋæ¹ʈpɕ: | ʂ^ʋo ʈtɕa: | ʔe: ʋ]

“*Hey, whatever you do, don’t tell anybody!*”

228. B: 嗯，我是不能说去啊，我今年四十多岁了，见人家，我说“我吃那饺子”。我怎么了，我！

Wa, wǒ shì bù néng shuō zhe. Wǒ jīn nián sìshí duō suì le, jiàn rénjiā wǒ shuō wǒ chī nà jiǎojiaor. Wǒ zěnmē le, wǒ!

[^ʋe: || ^ʋɔ¹ ^ʋo¹ ¹ʂɿ: ʋ pə: ʈnəŋ | ʂ^ʋɔ: ʈdʒa: || ^ʋo¹ ʈtɕin ʈnⁱẽ ʈ¹sɪ ʋɕ: | ʈwə: ʈ s^wɛ ʋ dɛ

ʈdʒ¹ẽ ʋ ʈdʒ¹æ: ʋ || ^ʋə ʈdʒ^wo ʈ^ʋɕ: | ʈʂ^hɿ: ʈnæ: | ¹tɕiɔu¹ dʒiɔ: ʋ (XXX)]

“*You can bet I won’t tell anybody! I’m more than forty years old, and when I meet people I’m supposed to say ‘wah, want eat jiaojiao’?!’*”

9 Concluding remarks

In this thesis, I set out to answer the question:

- How can I best represent the speech sounds of a Mandarin xiàngsheng through a phonetic transcription?

Through my work, it has become obvious that this question is very difficult, of not impossible, to answer. I have devised a notation that distinguishes far more phonetic detail than Hànyǔ Pīnyīn, but also considerably less than an impressionistic notation would. Whether this is the best way to transcribe the speech sounds of this material, I cannot say. I can only hope that it reveals some phonetic mechanisms of Mandarin speech that may of interest to students of the language.

The other question I posed at the beginning of this study, was:

- What can this transcription reveal about the way Mandarin speech sounds behave when connected to form words and phrases?

This question has been addressed partly through chapters 5, 6 and 7, which each discussed one particular problem related to the theory of Mandarin phonetics and phonology. In chapter 5, I demonstrated that there is a clear correspondence between certain acoustic qualities and the degree of syllable stress. This relates to the above question because it says something about how the rhythm and flow of the utterances in this data set is realized phonetically. Chapter 6 concerned the problem of tone 4 sandhi, which I found to be non-existent, or at least phonetically undetectable, in the present material. In chapter 7 I performed an investigation of another tone sandhi phenomenon; tone 3 sandhi. Although tone 3 sandhi was shown to occur in the material I analyzed, it did not account for the entire behavior of tone 3. Perhaps most interestingly, tone 3 sandhi, which is one of the most well known and extensively studied phonological phenomena in Mandarin Chinese, was shown to be a relatively marginal phenomenon in this material.

What other information the phonetic transcription gives about the behavior of Mandarin speech sounds is up to the reader to discover. I can only hope that my work is accurate and consistent enough that the information it reveals might prove useful.

References

- Abercrombie, D. 1967, *Elements of General Phonetics*, Edinburgh, Edinburgh University Press.
- Boersma, P. & Weenink D. 2010, *Praat version 5.1.38*, Amsterdam, University of Amsterdam, accessed 8 August 2010, <<http://www.fon.hum.uva.nl/praat/>>.
- 曹剑芬 Cáo Jiànfēn 2003, “语言的节奏 Yǔyán de Jiézòu [The Rhythm of Language]”, in *Report on Phonetics 2003*, 中国社会科学院语言研究所 Zhōngguó Shèhuìkēxué Yuàn Yǔyán Yánjiùsuǒ [Chinese Academy of Social Sciences, Institute of Linguistics], pp. 24-29, accessed 3 October 2010, <http://ling.cass.cn/yuyin/report/report_2003.htm>.
- Carr, P. 2008, *A Glossary of Phonology*, Edinburgh, Edinburgh University Press.
- Chao, Y. R. 2006a, “A System of ‘Tone Letters’”, in Wu Zongqi & Zhao Xinna (ed.) *Linguistic Essays by Yuenren Chao* pp. 98-102, Beijing, 商务印书馆 Shāngwù Yīnshūguǎn.
- 2006b, “Tone and Intonation in Chinese”, in Wu Zongqi & Zhao Xinna (ed.) *Linguistic Essays by Yuenren Chao* pp. 198-220, Beijing, 商务印书馆 Shāngwù Yīnshūguǎn.
- 1997, *语言问题 yǔyán wèntí [Problems of Language]*, Běijīng, 商务印书馆 Shāngwù Yīnshūguǎn.
- 1968, *A Grammar of Spoken Chinese*, Berkley, University of California Press.
- 1967, “Contrastive Aspects of the Wu Dialects”, in *Language*, Vol. 43, No. 1, pp. 92-101.
- 1948, *Mandarin Primer*, Cambridge, Harvard University Press.
- Chen, M. Y. 2000, *Tone Sandhi: Patterns Across Chinese Dialects*, Cambridge, Cambridge University Press.
- 辞海: 缩印本(音序) Cíhǎi Suōyìnběn Yīnxù 1999, Shànghǎi, 上海辞书印书馆 Shànghǎi Císhū Yīnshūguǎn.
- Duanmu, S. 2000, *The Phonology of Standard Chinese*, Oxford, Oxford University Press.
- [对口相声]学外语_马三立,张庆森[10分] 2010, online audio file, accessed 21 february 2010, <<http://www.tingcd.net/viewthread.php?tid=4197>>
- Fletcher, J. 2010, “The Prosody of Speech: Timing and Rythm”, in W. J. Hardcastle (ed.), J. Laver (ed.) & F. E. Gibbon (ed.), *The Handbook of Phonetic Sciences. Second Edition*, pp. 523-602, Chichester, Wiley-Blackwell.

- Fry, S 1955, "Duration and intensity as physical correlates of linguistic stress", in *the Journal of the Acoustical Society of America*, vol. 27, no. 4, pp. 765-768,
- Halliday, M. A. K. 2005, "A Systemic Interpretation of Peking Syllable Finals", in Webster, J. (ed.), *Studies in Chinese Language: Volume 8 in the Collected Works of M. A. K. Halliday*, London, Pinter Publishers.
- 1994, "Spoken and Written Modes of Meaning", in Graddol, D (ed.) & Boyd-Barrett, O. (ed.), *Media Texts: Authors and Readers*, pp. 51-73, Clevedon, The Open University.
- Halliday M. A. K & Matthiessen, Christian M. I. M. 2004, *An Introduction to Functional Grammar: Third Edition*, London, Hodder Education.
- 汉语大词典: 普及本 *Hànyǔ Dà Cídiǎn: Pǔjí Běn* 2000, Shànghǎi, 汉语大词典出版社
Hànyǔ Dà Cídiǎn Chūbǎnshè.
- 汉语大字典: 八卷本 *Hànyǔ Dà Zìdiǎn: Bājuǎnběn* 2001, Wǔhàn, 湖北辞书出版社 Húběi Císhū Chūbǎnshè/四川辞书出版社 Sìchuān císhū chūbǎnshè.
- Harrington, J. 2010, "Acoustic Phonetics", in W. J. Hardcastle (ed.), J. Laver (ed.) & F. E. Gibbon (ed.), *The Handbook of Phonetic Sciences. Second Edition*, pp. 81-129, Chichester, Wiley-Blackwell.
- Hardcastle W. J. (ed.), Laver J. (ed.) & Gibbon F. E. (ed.), *The Handbook of Phonetic Sciences. Second Edition*, Chichester, Wiley-Blackwell.
- Hayes, B 1995, *Metrical Stress Theory: Principles and Case Studies*, Chicago, the University of Chicago Press.
- Hockett, C. F. 1947, "Peiping Phonology", in *Journal of the American Oriental Society*, Vol. 67, No. 4; pp. 253-267.
- Howie, J. M. 1973, *Acoustical Studies of Mandarin Vowels and Tones*, Cambridge, Cambridge University Press.
- The International Phonetic Alphabet (IPA): revised to 2005*, 2005, International Phonetic Association, Department of Theoretical and Applied Linguistics, Aristotle University of Thessaloniki.
- Kaikkonen, M. 1990, *Laughable Propaganda: Modern Xiangsheng as Didactic Entertainment*, Doctoral Dissertation, Stockholm University, Institute of Oriental Languages.
- Karlgren, B. 1918, "A Mandarin Phonetic Reader in the Pekinese Dialect. With an Introductory Essay on the Pronunciation", in *Archives d'étude orientales*, Vol. 13, Stockholm, P. A. Norstedt & sönnner.

- Kratochvil, P. 1998, "Intonation in Beijing Chinese", in Daniel Hirst (ed.) & Albert Di Cristo (ed.), *Intonation Systems: A Survey of Twenty Languages*, pp. 417-432, Cambridge, Cambridge University Press.
- 1984, "Phonetic tone sandhi in Beijing dialect stage speech", in *Cahiers de linguistique – Asie orientale*, Vol. 13, No. 2, pp. 135-174.
- 1974, "Stress Shift Mechanism and Its Role in Peking Dialect", in *Modern Asian Studies*, Vol. 8, No. 4, pp. 433-458.
- 1968, *The Chinese Language Today: Features of an Emerging Standard*, London, Hutchinson University Library.
- 1967, "On the Phonology of Peking Stress", in *Transactions of the Philological Society*, Vol. 66, No. 1, pp. 154-178.
- Ladefoged, P. 2001, *Vowels and Consonants: An Introduction to the Sounds of Languages*, Oxford, Blackwell Publishers.
- Ladefoged, P. & Maddieson, I. 1996, *The Sounds of the Worlds Languages*, Oxford, Blackwell Publishers.
- Laver, J. 1994, *Principles of Phonetics*, Cambridge, Cambridge University Press.
- Lin, Y.-H. 2007, *The Sounds of Chinese*, Cambridge, Cambridge University Press.
- 罗常培 Luó Chángpéi 1957, *普通语音学纲要 Pǔtōng Yǔyīnxué Gāngyào [An Outline of General Phonetics]*, Beijing, 科学出版社 Kēxué Chūbǎnshè.
- Mǎ Sānlǐ 2003, *我的一辈子 Wǒ de Yī Bèizi [My Life]*, 中华相声 Zhōnghuá Xiàngsheng, accessed 14 October 2010, <<http://www.xiangsheng.org/asp/html/2003/02/20030227130500-1.htm>>
- Mair, V. H. 1991, "What is a Chinese 'Dialect/Topolect'? Reflections on Some Key Sino-English Linguistic Terms", in *Sino-Platonic Papers*, No. 27, University of Pennsylvania, Department of Oriental Languages and Civilizations, accessed 8 November 2010, <http://www.sino-platonic.org/complete/spp029_chinese_dialect.pdf>
- Rygaloff, A. 1955, "La Phonologie du Pekinois", in *T'oung Pau*, Vol 43, No 3, pp. 183-264.
- Shen, S. X. 1989, *The Prosody of Mandarin Chinese*, Berkley, University of California Press.
- Shih, C. 1997, "Mandarin third tone sandhi and prosodic structure", in Jialing Wang (ed.) & Norval Smith (ed.) *Studies in Chinese Phonology*, The Hague, Mouton de Gruyter.
- 1988, "Tone and Intonation in Mandarin", in *Working papers of the Cornell Phonetics Laboratory*, No 3, pp. 83-109.

- Spešnev, H. A. 1959, “现代汉语音节重音的声学本质 Xiàndài Hànyǔ Yīnjié Zhòngyīn de Shēngxué Běnzhi [The Phonetic Basis of Syllable Stress in Modern Chinese]” in 中国语文 Zhōngguó Yǔwén, Vol. 8, No. 1, pp. 28-34.
- Stetson, H. R. 1951, *Motor Phonetics: A Study of Speech Movements in Action*, Amsterdam, North-Holland publishing company.
- Sweet, Henry 1877, *Handbook of Phonetics*, Oxford, Clarendon Press.
- 现代汉语词典 Xiàndài Hànyǔ Cídiǎn, 5th edition 2007, Běijīng, 商务印书馆 Shāngwù Yīnshūguǎn.
- 薛宝琨 Xuē Bǎokūn 1985, 中国的相声 Zhōngguó de Xiàngsheng [Chinese Xiàngsheng], Běijīng, 人民出版社 Rénmín Chūbǎnshè.
- 学外语 / 马三立 张庆森 / 卡通相声 Xué Wàiyǔ / Mǎ Sānlì Zhāng Qìngsēn / Kǎtōng Xiàngsheng 2009, CCTV, online video, accessed 2 February 2010, <<http://www.youtube.com/watch?v=i16OV5OFv0M>>
- 王力 Wáng Lì 1982, 汉语音韵学 Hànyǔ Yīnyùnxué [Chinese Phonology], Beijing, 中华书局 Zhōnghuá Shūjú.
- 王茂林 Wáng Màolín & 林茂灿 Lín Màocàn 2003, “普通话自然话语中的下倾 Pǔtōnghuà zìrán huàyǔ zhōng de xiàqīng [Declination in Standard Chinese Natural Discourse]”, in *Report on Phonetics 2003*, 中国社会科学院语言研究所 Zhongguo Shehuikexue Yuan Yuyan Yanjiusuo [Chinese Academy of Social Sciences, Institute of Linguistics], pp. 24-29, accessed 3 October 2010 <http://ling.cass.cn/yuyin/report/report_2003.htm>.
- Yin Hui 2003, “The blocking of tone sandhi in Mandarin Chinese” In *Proceedings of the 2003 Annual Conference of the Canadian Linguistic Association*, pp. 296–307. Québec, Université du Québec à Montréal.
- “张庆森 Zhāng Qìngsēn”, 百度百科 bǎidùbǎikē, 2 June 2010, accessed 20 November 2010, <<http://baike.baidu.com/view/1170919.htm>>.
- Zhōu Bīn (ed.) 2004, 新华大字典: 彩色版 Xīnhuá Dà Zìdiǎn: Cǎisèbǎn, Běijīng, 商务印书馆 Shāngwù Yīnshūguǎn.