# 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

<table>
<thead>
<tr>
<th><strong>Phonological notation</strong></th>
<th><strong>Definition</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>[xxx]</td>
<td>Phonetic transcription</td>
</tr>
<tr>
<td>/xxx/</td>
<td>Phonological transcription</td>
</tr>
</tbody>
</table>
| /__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:  
  
  \[
  \text{i} \\
  /x__n/ \\
  \rightarrow \\
  \text{Realized as: Indicates how a phonological form is pronounced. Often used in combination with slanting brackets and square brackets. For example, the observation that “the phoneme /j/ is pronounced as either [ʨ] or [ʥ] can be represented as:} \\
  /j/ \quad \rightarrow \quad [ʨ] \\
  \quad \quad [ʥ] \\
  \]
| xxx||xxx | Phrase boundary |
| C | Consonant |
| V | Vowel |
| ′ma | Primary stress: the most strongly stressed syllable in a phrase |
| ,ma | Secondary stress: any stressed syllable that is not the most strongly stressed in a phrase. |

<table>
<thead>
<tr>
<th><strong>Abbreviations</strong></th>
<th><strong>Definition</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>Noun phrase</td>
</tr>
<tr>
<td>VP</td>
<td>Verb phrase</td>
</tr>
</tbody>
</table>
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 “zaijiàn (117)” in the text, this signifies that I am referring especially to the word zaijiàn that is uttered by speaker A in turn 117 of Xuè 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 Bēijīng. 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 xiängsheng”.\(^2\) 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, 1985: 7-33).

\(^2\) 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 Wàiyyǔ 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 Sānlì met Zhāng Qìngsēn are born in that city, they lived and practiced their art primarily in Tiānjin. This does not mean that Xué Wàiyyǔ 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 loose 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 xiàngshēng 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)

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. Hocket’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 lungs. 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 level. 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 [hɑː]](image)

The waveform in figure 2.1 is produced from a recording of me saying [hɑː:]. 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 [ɑː] 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”.

18
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 [æː] 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♭ 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 [æː] 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 airstream 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 (shhhhh) 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** (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 uvula, the consonant is **uvular** (English has no uvular consonants, but the “r” in French “rouce” 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 aspired, 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 [ɲ]. 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 scientists. 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.
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].

**Figure 2.2:** Spectrograms of vowels [a] (left) and [i] (right)
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 [ɑ] will still be an [ɑ], 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 if 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 Norwegian 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.
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 [tʰɑŋ ɛn] (“the thought”) and [tʰɑŋ ɛn] (“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;
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 is 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. 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 fānqiè 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 鄂 红 切

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 绀 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.

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labio-dental</th>
<th>Alveolar</th>
<th>Retroflex</th>
<th>Alveolo-palatal</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>/p/ [pʰ]</td>
<td></td>
<td>/t/ [tʰ]</td>
<td></td>
<td></td>
<td>/k/ [kʰ]</td>
</tr>
<tr>
<td></td>
<td>/b/ [p]</td>
<td></td>
<td>/d/ [t]</td>
<td></td>
<td></td>
<td>/g/ [k]</td>
</tr>
<tr>
<td>Affricate</td>
<td>/c/ [tsʰ]</td>
<td>/ch/ [tsʰ]</td>
<td>/q/ [ʨʰ]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td>/l/ [l]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>/m/ [m]</td>
<td>/n/ [n]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td></td>
<td></td>
<td>/r/ [ɻ]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Mandarin initial consonants

Perhaps the most striking feature of this system of consonants is that there is no opposition of voiced an 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.
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 mophthong, 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-glode if it occurs before the nuclear vowel, and an off-glode if it occurs after. Triphthongs consist of a syllabic vowel along with both an on-glode and an off-glode. 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:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/n/</td>
<td>[n]</td>
</tr>
<tr>
<td>/ng/</td>
<td>[ŋ]</td>
</tr>
</tbody>
</table>

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-glode, with an off-glode or with both. Table 3.3 displays the finals that end with a consonant. Since no consonant can follow a triphthong or an off-glode, the rows give the height of the nuclear vowels and specify whether or not it is preceded by an on-glode. The columns give the final consonant.

When the final has an on-glode 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 [yঐn] ”eye” (Lin, 2007: p. 126). In the tables, these forms are given to the right of the normal spelling.
The Close unrounded vowel phoneme /i/ has three different realizations, depending on which initial precedes it. It is pronounced [ɿ] after /zh/, /ch/, /sh/ and /r/; [i] 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.](image)

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:
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” [tʰɐː˥˥˥ təɿ]

After T2: [3] shéi de 谁的 “whose?” [ʂɛɪ˧˥ təɿ]

After T3: [4] nǐ de 你的 “your(s)” [niː˨˩ təɿ]

After T4: [1] dà de 大的 “big one(s)” [tɐː˥˩ təɿ]

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 change 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) zhǎo huǒ “to look for fire” becomes homophonous with zháo huǒ “to catch fire”:

zhǎo huǒ 找火 “to look for fire”: [˨˩˦˨˩˦] → [˨˩˦˨˩˦]
zháo huǒ 找火 “to look for fire”: [˩˧˥˩˨˩˦] → [˩˧˥˩˨˩˦]

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”: [˨˩˦˥˩] → [˨˩˦˥˩]

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 [˧5] to a level [˥5], 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):

sānniánjī 三年级 “third-year class”: [˥˥˧˥] → [˥˥˧˥]

cōngyóubǐng 葱油饼 “onion oil cake”: [˥˥˧˥] → [˥˥˧˥]

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”.

33
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):

- ˈyàoshi 要是 “if”
- yàoˈshi 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):
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 [z]. 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). Were possible, examples from English has 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**

- [pʰ] **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”
- [tʰ] **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”
- [kʰ] **Aspirated velar**: like the “c” in English “cold”
- [k] **Unvoiced velar**: like the “c” in English “reluctant”
- [ɡ] **Voiced velar**: like the “g” in English “gap”
- [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
[ʐ] **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.

[z] **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”)

[w] **Voiced uvular;** like the “r” in French “roue”

[h] **Unvoiced glottal;** like the “h” in English “help”

**Affricates**

[ʦʰ] **Aspirated alveolar;** like the “ts” in Japanese (and English) 津波 tsunami, only more strongly aspirated

[ʦ] **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.

[ʦʰ] **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.

[ʦ] **Unvoiced retroflex;** see previous entry

[dz] **Voiced retroflex;** see previous entry

[ʨʰ] **Aspirated alveolo-palatal;** like the “Ch” in Japanese “千葉 Chiba” (place name). See alveolo-palatal fricatives

[ʨ] **Unvoiced alveolo-palatal;** See previous entry

[ʥ] **Voiced alveolo-palatal;** Like the “ji” in Japanese “実 jitsu” (“truth”)

**Nasals**
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](image-url)
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 lead to any greater analytical clarity. As an example of the latter, the symbol [ɯ] 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 [y] 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 [ʅ], 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 [ʊ] 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 [ʅ]; 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.
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”)
- [ɨ]: 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]
- [ɿ]: Is similar to [ɨ], 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**

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

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

Special vowel articulations
[ɜ] Diacritic signifies that the vowel is nasalized. Like French “sans” (“without”)
[ɔ] Diacritic signifies that the vowel is r-colored. Like American “bird”
[ɜː] Diacritic signifies that the vowel is particularly long: like English “meal” when compared to “mill”
[ˈ] 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 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 H 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/, /nián/ 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. 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 wò niàn liù nián shū

[ˌkʊŋhə xuəˈniənliʊənɪəνˈʃuː]    

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. 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.

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”:  [˦˧˥] →  [˦˧˥]

找火 zhǎo huǒ “to look for fire”:  [˨˩˦] →  [˨˩˦]

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
tone sometimes appear to have a slightly rising or falling contour. In these cases, the contour is represented using a normal tone letter.

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

**Tones of tonic syllables**

<table>
<thead>
<tr>
<th>Tone</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>[45]</td>
<td>High and slightly rising variant of tone 1; usually not as sharply rising as tone 2.</td>
</tr>
<tr>
<td>[35]</td>
<td>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.</td>
</tr>
<tr>
<td>[214]</td>
<td>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.</td>
</tr>
<tr>
<td>[21]</td>
<td>Low falling variant of tone 3; in practice it is the first half of [214].</td>
</tr>
<tr>
<td>[14]</td>
<td>Rising variant of tone 3; in this material it is usually, but not always, lower than the [35] of tone 2.</td>
</tr>
<tr>
<td>[212]</td>
<td>An intermediate variant of [214] and [21]; it rises towards the end, but not as high as [214].</td>
</tr>
</tbody>
</table>

**Neutral tone variants**

<table>
<thead>
<tr>
<th>Tone</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>[2]</td>
<td>Low, short and level, but not as low as [1]</td>
</tr>
</tbody>
</table>

[42] [˦] Falling neutral tone; not as sharply falling as the [51] of tone 4.

[24] [˨˦] 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

[ˈkʊŋ] Primary stress: the most heavily stressed syllable in a phrase, and the turning point of the intonation contour

[ˌkʊŋ] Secondary stress: Any syllable which has more stress than all adjacent syllables, but which is not the most heavily stressed syllable in a phrase

[kʊŋˌkʊŋ] 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 Wâiyû. 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. Speșnev (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 Wâiyû?
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. This went contrary to the common assumption in earlier literature, 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.”

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):

\[
\frac{\text{duration} \times \text{loudness} \times F_0}{100 \times 000} = \text{Syllable volume}.
\]

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 didao 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 vaules “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.
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

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Expression</th>
<th>Syllable</th>
<th>Duration</th>
<th>F0 mean</th>
<th>F0 peak</th>
<th>Db mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>màozi 帽子 (171)</td>
<td>[maoɻ]  [dzɿ]</td>
<td>188 42</td>
<td>327 210</td>
<td>358 214</td>
<td>79 70</td>
</tr>
<tr>
<td>A</td>
<td>màozi 帽子 (173)</td>
<td>[maoɻ]  [dzɿ]</td>
<td>170 47</td>
<td>126 213</td>
<td>285 255</td>
<td>78 68</td>
</tr>
<tr>
<td>A</td>
<td>màozi 帽子 (175)</td>
<td>[maoɻ]  [tsɿ]</td>
<td>187 46</td>
<td>205 253</td>
<td>296 256</td>
<td>83 68</td>
</tr>
<tr>
<td>B</td>
<td>màozi 帽子 (154)</td>
<td>[maoɻ]  [dzɿ]</td>
<td>280 68</td>
<td>260 149</td>
<td>357 151</td>
<td>82 68</td>
</tr>
<tr>
<td>B</td>
<td>màozi 帽子 (156)</td>
<td>[maoɻ]  [dzɿ]</td>
<td>270 54</td>
<td>247 136</td>
<td>330 136</td>
<td>78 62</td>
</tr>
<tr>
<td>B</td>
<td>màozi 帽子 (168)</td>
<td>[maoɻ]  [dzɿ]</td>
<td>203 45</td>
<td>278 230</td>
<td>356 232</td>
<td>85 80</td>
</tr>
<tr>
<td>B</td>
<td>màozi 帽子 (174)</td>
<td>[ma.ɻ]  [dzɿ]</td>
<td>129 51</td>
<td>226 191</td>
<td>232 194</td>
<td>83 77</td>
</tr>
<tr>
<td>A</td>
<td>zhuōzi 桌子 (115)</td>
<td>[tsu.oɿ]  [tsɿ]</td>
<td>160 43</td>
<td>225 193</td>
<td>234 198</td>
<td>82 79</td>
</tr>
<tr>
<td>A</td>
<td>tāmen 他们 (3)</td>
<td>[tɿn]  [mɿnɿ]</td>
<td>114 61</td>
<td>239 246</td>
<td>251 258</td>
<td>77 77</td>
</tr>
<tr>
<td>B</td>
<td>zhēnde 真的 (14)</td>
<td>[tʂɔnɿ]  [dɿ]</td>
<td>78 77</td>
<td>365 244</td>
<td>394 324</td>
<td>72 80</td>
</tr>
<tr>
<td>A</td>
<td>hánchen 寒碜 (205)</td>
<td>[tʂɛnɿ]  [tɿɿ]</td>
<td>79 44</td>
<td>179 193</td>
<td>187 194</td>
<td>78 65</td>
</tr>
<tr>
<td>A</td>
<td>shènme 什么 (35)</td>
<td>[ʂɿmɿ]  [mɿ]</td>
<td>63 46</td>
<td>181 243</td>
<td>200 256</td>
<td>74 75</td>
</tr>
<tr>
<td>B</td>
<td>shènme 什么 (140)</td>
<td>[ʂɿmɿ]  [mɿ]</td>
<td>53 21</td>
<td>265 390</td>
<td>304 391</td>
<td>82 80</td>
</tr>
<tr>
<td>A</td>
<td>xiěde 写的 (15)</td>
<td>[ɕeɿɿ]  [dɿ]</td>
<td>112 31</td>
<td>297 240</td>
<td>307 251</td>
<td>77 77</td>
</tr>
</tbody>
</table>

Table 5.1: Acoustic measurements of neutral tone expressions
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 4. 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)](image)

---

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 [ʈʂən˥˥˥] 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.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Compound</th>
<th>Syllable</th>
<th>Duration</th>
<th>F0 mean</th>
<th>F0 peak</th>
<th>Db mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>chàdiǎnr 稍点儿 (32)</td>
<td>[ʈʂʰɐː˥˩]</td>
<td>132</td>
<td>212</td>
<td>249</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[tʊə˨˩˨]</td>
<td>173</td>
<td>208</td>
<td>229</td>
<td>82</td>
</tr>
<tr>
<td>A</td>
<td>bàodǔr 爆肚 (143)</td>
<td>[pɑʊ˥˩]</td>
<td>148</td>
<td>222</td>
<td>243</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[tu˞ɚ˨˩˦]</td>
<td>203</td>
<td>240</td>
<td>283</td>
<td>81</td>
</tr>
<tr>
<td>A</td>
<td>zàijiàn 再见 (115)</td>
<td>[tsɛɛ˨˩]</td>
<td>134</td>
<td>182</td>
<td>214</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ʨiɛn˦˦]</td>
<td>153</td>
<td>206</td>
<td>244</td>
<td>82</td>
</tr>
<tr>
<td>A</td>
<td>duìhuà 对话 (109)</td>
<td>[tʰɛɻ˧˥]</td>
<td>113</td>
<td>204</td>
<td>224</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[χəɻɻ˨˩]</td>
<td>112</td>
<td>206</td>
<td>221</td>
<td>78</td>
</tr>
<tr>
<td>B</td>
<td>duìhuà 对话 (110)</td>
<td>[tʰɛɻ˧˥]</td>
<td>100</td>
<td>234</td>
<td>248</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[χəɻɻ˨˩]</td>
<td>146</td>
<td>233</td>
<td>260</td>
<td>82</td>
</tr>
</tbody>
</table>

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 [tʰɛi]. 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)](image)

### 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 Wàiyù. 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 looses 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”:  \[ ˥˩  \rightarrow  ˥˧  \]

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 threat 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, 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

---

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 for 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”:

\[ \text{[˧˥ ˥˩]} \rightarrow \text{[˧˦ ˥˥˥]} \]

扶戟 fújǐ “to hold a halberd”:

\[ \text{[˧˥ ˨˩˦]} \rightarrow \text{[˧˥ ˧˩˦]} \]

打熬 dǎ’áo “To suffer”

\[ \text{[˨˩ ˧˥]} \rightarrow \text{[˨˩ ˨˥]} \]

扶乩 fújī “planchette writing”

\[ \text{[˧˥ ˥˥˥]} \rightarrow \text{[˧˦ ˥˥˥]} \]

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 “\(\backslash\)” 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 loosing 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 Wàiyū 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 Wàiyū, 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 boundary. 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 biyè (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. Duìhuà (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.

<table>
<thead>
<tr>
<th>Stress</th>
<th>Expression</th>
<th>Syllable</th>
<th>F0 high</th>
<th>F0 low</th>
<th>intonation</th>
</tr>
</thead>
<tbody>
<tr>
<td>weak-strong</td>
<td>biyè 毕业 (7)</td>
<td>[piɿ] [ieɿ]</td>
<td>240</td>
<td>208</td>
<td>Falling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>239</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>weak-strong</td>
<td>wèn lù 问路 (87)</td>
<td>[vənɿ] [luɿ]</td>
<td>301</td>
<td>270</td>
<td>Rising</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>337</td>
<td>190</td>
<td></td>
</tr>
<tr>
<td>weak-strong</td>
<td>duìhuà 对话 (109)</td>
<td>[tɿ engagements])</td>
<td>208</td>
<td>188</td>
<td>Falling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>221</td>
<td>184</td>
<td></td>
</tr>
<tr>
<td>weak-strong</td>
<td>zaijiàn 再见 (115)</td>
<td>[tsɛɿ] [wiɿnɿ]</td>
<td>214</td>
<td>170</td>
<td>Rising</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>244</td>
<td>154</td>
<td></td>
</tr>
<tr>
<td>strong - weak</td>
<td>tài shuài 太帅 (15)</td>
<td>[tʰaeɿ] [ɕʼæɿ]</td>
<td>267</td>
<td>214</td>
<td>Falling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>163</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>strong - weak</td>
<td>tài gòu 太够 (23)</td>
<td>[tʰeɿ] [kouɿ]</td>
<td>210</td>
<td>169</td>
<td>Falling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>163</td>
<td>138</td>
<td></td>
</tr>
<tr>
<td>strong - weak</td>
<td>tài huì le 太会了 (37)</td>
<td>[tʰaeɿ] [xɿɛɿ]</td>
<td>304</td>
<td>191</td>
<td>Falling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>218</td>
<td>186</td>
<td></td>
</tr>
<tr>
<td>strong - weak</td>
<td>tài huì le 太会了 (39)</td>
<td>[tʰaeɿ] [xɿɛɿ]</td>
<td>322</td>
<td>177</td>
<td>Falling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>215</td>
<td>179</td>
<td></td>
</tr>
<tr>
<td>strong - weak</td>
<td>tài huì le 太会了 (73)</td>
<td>[tʰaeɿ] [xɿɛɿ]</td>
<td>418</td>
<td>228</td>
<td>Falling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>178</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>strong - weak</td>
<td>zuì kuài 最快 (89)</td>
<td>[tsɿɛɿ] [ksʼæɿ]</td>
<td>372</td>
<td>273</td>
<td>Falling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>336</td>
<td>238</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1: Pitch values of tone 4 disyllabic expressions

Wèn lù (89) and zaijiàn (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 zaijiàn (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)
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 → [tsæ˨˧ ʨiɛn˥˩]
```

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 Beijing xiangsheng, 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 Xue Waiyu, 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 Lao Li mai haojiu 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 → [˨˩˦ ˨˩˦ ˨˩ ˨˩˦ ]

→ [˨˩˦ ˨˩ ˨˩˨ ˨˩˦ ]

→ [˨˩˦ ˨˩ ˨˩˨ ˨˩˦ ]

→ [˨˩˦ ˨˩ ˨˩˨ ˨˩˦ ]

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) [lau˨˩]
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.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Number of occurrences</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>[214]</td>
<td>13</td>
<td>4,6 %</td>
</tr>
<tr>
<td>[21]</td>
<td>125</td>
<td>44,1 %</td>
</tr>
<tr>
<td>[14]</td>
<td>81</td>
<td>28,6 %</td>
</tr>
<tr>
<td>[212]</td>
<td>23</td>
<td>8,2 %</td>
</tr>
<tr>
<td>[NT]</td>
<td>41</td>
<td>14,5 %</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>283</strong></td>
<td><strong>100 %</strong></td>
</tr>
</tbody>
</table>

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: 我说得也挺熟啊
   Wǒ shuō de yě tíng shóu ā

“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 něiguóhuà shuō de zuì shóu ne?

“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, [lɑʊ] (1) precedes a syllable with primary stress:

1. A: 张庆森说相声是老资格
   Zhāng Qìngsēn Shuō xiàngsheng shì lǎo zīgé.

“Zhang Qingsen is a real veteran at performing xiangsheng”
The most common variant of tone 3 is [21], which makes up almost half of the total occurrences in Table 7.1. Just as stated in the tone sandhi rules, this variant occurs when followed by any tone except tone 3 as well as sometimes in phrase-final position. In the material I have analyzed, [21] does not show up before another tone 3.

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 [ʈʂəo], both adopt the rising variant. Note that this is also the longest string of tone 3 syllables that occurs in Xue Waiyu:

133. A: […]你找我说不上来[…]
    nǐ zhǎo wǒ shuō bù shàng lái
    [ni˨˩˦ˌʈʂəo˨˩˦ːʊə̹˨˩ˈʂʊə̹˥˥˥‿pʊ꜉�ˌʐʌŋˌlae˦]
    "Zhang Qingsen is a real veteran at performing xiangsheng"

Tone 3 also changes to its rising form when followed by a tone 3 syllable that is reduced to a neutral tone. Even when a tone 3 syllable is merged with another tone 3 syllable, it adopts the sandhi form. Hence the two following examples:

145. A: […]这我也会呀。[…]
    zhè wǒ yě huì ya
    [ʈʂəo˥˩ˌuo˨˩ɪɜ꜈bu꜉ˈxʊɛɪ˥˩ɪɐ꜌]
    "Even I wouldn’t know that."

141. A: […]我也想知道[…] 
    Wǒ yě bù zhīdào
    [ʈʂəu˩˧ pʊə̹˨˩ˈʐʊə̹꜆]
    "Neither do I know […]"

In the present material, 14 occurrences of [14] take place when followed by a fully pronounced tone 3. 12 take place when followed by a reduced tone 3 of one of the kinds exemplified above. This means that there are 49 instances of [14] that are not followed by any kind of tone 3 syllable, reduced or otherwise. In other words, the majority of the occurrences
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?

[ˌniː ʂʅ kɛ ꞈ˘ ʃʅ ꞈ˘ kɛ ꞈ˘ niː ʂʅ ꞈ˘ niː nyː ꞈ˘ tɛː ꞈ˘]

"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.

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

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 Wàiyù, 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 xiàngsheng shì lǎo zīgé.
   [ˌʈʂɑŋ˧˨˦˥ tʂʰiŋ˧˥ sœŋ˧˥] ʒøː˨˩ ʃɛŋ˧˥ lœŋ˨˩˧ lɑːʊ˨˩˧ lœːŋ˧˥ ʈʂɿː˦˨ kɑː˨˩ ʃuō˧˥˧]
   "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.
   [ˌʔɜɪ꜈꜊ɪɨː ꞌɪɿ ˌkʰʌ˨˩ pʊ꜈ˈkɒn˨˩ nəm˧ ꞌsøːa:˥]
   "Oh no, I wouldn’t dare say that!"

3. A: 听他们提过。
   Tīng tāmen tíguó.
   [ˌtʰɪŋ˦˥ tɐː˦˥ mɘ ꞌtʰɪː˧˥ ɡʊɑː˥˩]
   "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 [sœŋ˧]. 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 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 /ɡ/ of the syllable [ɡʱːɑː˥] is an example of a normally unvoiced consonant that assimilates the voice of the preceding vowel.
4. B: 是吗?
Shì ma.
[ˈʂʅ ˥˩intosh]
"Is that so?"

5. A: 说您的文化水平很高。
Shuō nín de wénhuà shǔipíng hěn gāo.
[ʂʊ oː̃ː˥˥ˌnin tə̃ː˧˥ʼpʰŋɊ xənɊ Ɂkæ̃ ʅ]
"They say you are a very educated person"

6. B: 高什么呀。反正年过几年书。
Gāo shénme. Fǎnzhèng niànguo jī nián shū.
[ʔ h ˈkauɊ ɁənɊ meɊ ˌfæŋɊ dəŋɊ ˨˩˧˥ŋ əəɊ təɊ ɁnɊ ɁsuɊ]
"Oh, there’s nothing educated about me. Well anyway, I went to school for a few years..."

7. A: 大学毕业。
Dàxué bièɊ.
[ˈtaː˧ ɊɃ biɊɊ ieɊ]
"University graduate!"

---

33 The syllable, [və̃ː] 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 [fæŋɊ dəŋɊ] 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 [ˈtaː] could well have been transcribed as [513].
8. B: 啊，没有没有。共合我念六年书。
Ah, no, no, nothing like that! All in all I’ve gone to school six years.”

9. A: 客气。
Kèqi.
"Now you’re being polite.”

10. B: 嘿这实在。
A zhè shízài.
"No, this is true!”

11. A: 您只客气。
Nín zhǐ kèqi.
"You’re just being polite.”

12. B: 怎么。
Zěnme.
"Why do you say that?”

---

36 Phrase spoken while laughing, hence the [h] in ‘mhe’.

37 The final vowel of syllable ‘ʈʂɘɝː’ assimilates the retroflex place of articulation of the following initial consonant. Note that ‘ʈʂɘɝː’ 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。
Kèqi. Nín juéduì bù néng niàn wǔ liù nián shū.
"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.
"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.
"Your handwriting is really good looking, I think”

16. B: 哎，就是常写嘛。
Ai, jiù shì cháng xiě ma.
"Heh, well, that just comes from writing a lot”

17. A: 应当要这样。
Yīngdāng yào zhèyang.
"That is the way it should be”

18. B: 是吗?
Shìma?
"Oh?”
19. A: 相声演员必须要有文化。
Xiàngsheng yǎnyuán bìxū yào yǒu wénhuà.
"Xiangsheng artists have to be well educated"

20. B: 就是。
Jiùshì.
"Well said!"

21. A: 像您这个文化，说相声儿，富富余余。
Xiàng nín zhège wénhuà, shuō xiàngsheng, fùfù yùyú.
"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: [i̹].

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

\/_en/ → [v]  
/_ei/ → [v]  
/_ai/ → [v]  
/_an/ → [v]  
/_o/ → [o]

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 [gyː] and [χoː] are very long syllables, as the speaker draws out the words as if he is thinking about what to say next. Note that [gyː] 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, fānzhèng còuhé.
[ʔəɪ ‖ fæŋɻ dɛnɿ tʃʰouɿ xʌɿ]42
"Well, I make do at least."

23. A: 太够。太够了。
Tài gòu. Tài gòu le.
[tʰɐ e ˥˩ koʊɿ ‖ tʰɐ e kʊɿ]
"It's more than enough! More than enough!"

24. B: 这不能那么说。
Zhè bùnéng nàme shuō.
[tʂʰɿ puɿ nɤɿ nəmɿ ꞌʂw oː˥]
"Oh no, that's taking it too far."

25. A: 可也有缺点。
Kě yě yǒu quē diǎn.
[kʰɤ ˩˦ ieː ˩˦ ‖ tɕʰyɛː ꞌtæŋɿ]
"But you still have one shortcoming"

26. B: 啊我?
Ā wǒ?
[ʔəʊ ɿ ʍoːɿ]
"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 realised as [nəm]. 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 diǎn quēdiǎn.

"A tiny shortcoming"

28. B: 哪一点儿您指出来。
Nǎ yìdiǎnr nín zhǐ chū lái.

"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.

"You, who have performed xiangsheng for all these years,"

30. B: 是是。
Shì, shì.

"yes, yes."

31. A: 不会说外国话。
Bù huì shuō wàiguóhuà.

"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 [ɚ] 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 [ɚ], while no stressed syllable is this heavily reduced.

46 The tones of syllables [na] and [náh] 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.
32. B: 呵...这点，差点儿，差点儿。啊这是个缺点哪。

A zhèdiǎn, chà diǎnr, chà diǎnr. a zhè shìge quēdiǎn á?

"Ah yes, at that point I... fall a little short. Is that really such a shortcoming?"

33. A: 俄,说相声...说相声哪儿不会说外国话的。

Á shuō xiàngshengr... shuō xiàngshengr, nǎr bú huì shuō wàiguóhuà de?

"A xiangsheng-performer... How can a xiangsheng-performer not know any foreign?!!"

34. B: 哦?

Ó?

"Oh?"

35. A: 哪儿不会说外国话的，这叫什么说相声哪?

Nǎr bú huì shuō wàiguóhuà de? Zhè jiaò shénme shuō xiàngshengr na?

"How can he not know any foreign? What kind of xiangsheng would that be?"

36. B: 那这么说，您会说外国话?

Nà zhème shuō nín huì shuō wàiguóhuà?

"So, when you put it like that, does it mean that you can speak foreign?"

37. A: 嗯? 太会啦。

N? tài huì la.

"Oh, do I!"
38. B: 你瞧。
Nín qiáo?
[ nin\˧˥iptables:]1 "Really?"

39. A: 太会啦。
Tài huì la.
[ tʰae\ ˩˩˨ xʊɛɪ\ ˥˩la.] "Do I ever!"

40. B: 那么您会哪国话?
Nàme nín huì něiguóhuà?
[ʔə�\ LowerCase\ nin\ ˧˥ xʊɛɪ\ ˥˩ \næɛ\ 籴\ ˧˥ \χɔɐː\ ˥˩]47 "Then, which language do you know?"

41. A: 哪国话都会呀。
Něiguóhuà dōu huì ya.
[ɛi\ ˥˩˨ k\ ˧˥ xʊɛɪ\ ˥˩ \tuo\ xʊɛɪ\ ˥˩]48 "I know all languages!"

42. B: 哇?
 Ō?
[əa] "Oh?!"

47 The three stressed syllables [nin\ ˧˥iptables\ xʊɛɪ\ ˥˩næɛ\ 籴\ ˧˥χɔɐː\ ˥˩] are separated by pauses. If this were not the case, I would have transcribed them [nin\ iptables\ xʊɛɪ\ ˥˩næɛ\ 籴\ ˧˥χɔɐː\ ˥˩]. The pause between the two syllables with secondary stress [xʊɛɪ\ iptables] and [næɛ\ 籴\ ˧˥] is shorter than that between [nin\ iptables\] and [xʊɛɪ\ iptables].

48 Interestingly enough, the two speakers pronounce the word něi 哪 differently; speaker A says [nei] and speaker B says [næi]. This stays consistent throughout their entire dialogue.
43. A: 全会呀。
Quán huì ya.
[ˈʨʰʏæ̃˧˥xʊeɪəl]⁴⁹
“I know them all!”

44. B: 哪国话都会呀?
Něiguóhuà dōu huì ya?
[ˈnæɛɭkʊəɭ tsù ˧˥ ꜆xùə]⁵⁰
“You know all languages?”

45. A: 各国的话都会。我到过外国。
Gèguó de huà dōu huì. Wo dauguo waiguo.
[ˈkɯɜː˥˩ kʊəɭdʌɭxʊː˦˨ uɪəɭ ائهم]⁵¹
“I know every single language. I've been to other countries, you know.”

46. B: 哦您哪?
Ō nín na?
[Sœɭɪnəɭ]⁵²
“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 [tɑʊɭ] could well be treated as two separate syllables [tɑɭ uɭ]. 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 [kʊəɭɭɑɭɭ] are pronounced so close together that they may well be treated as a single syllable [kʊəɭ]. I have treated it as two syllables because it has two relatively long vowel segments [æ] 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.
"I went abroad as a schoolgirl"

48. B: 哦！
Ō!
"Oh!"

49. A: 我在外国…
Wǒ zài wàiguó…
"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?
"Hold on a second!… You’re a woman?!

51. A: 啊？
Ā?
"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 Thrithongs when in tones 3 or 4. Similar patterns of tonally correlated variation are found in this material as well. The syllable [l'u:] 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?

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

53. A: 就是男。。。说错了。
Jiùshì nán shuōcuò le.

"I'm a man of cou... Oh, slip of the tongue!"

54. B: 这说错了。
A, zhè shuōcuò le?

"That was a slip of the tongue?"

55. A: 出洋留学的男学生。
Chūyáng liúxué de nán xuéshēng.

"I went abroad as a... schoolboy!"

56. B: 哈。
Hā.

"heh..."

57. A: 男学生！
Nán xuéshēng.

"A schoolboy!"
58. B: 啊，就是个留学生？
A: jiùshì ge liú xuéshēng?

"Oh, so like an exchange student then?"

59. A: 对啦，对啦。
Dui la, dui la.

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

60. B: 那么您会哪国话?
Nàme nín huì něguóhuà?

"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，意国话，反正眼面儿前儿的这[就]话，全行。
Chàbùduō ba, fǎnzhèng. Yīngguóhuà, Fǎguóhuà, Déguóhuà, Sūliánhuà, Yìguóhuà.
Fǎnzhèng yǎnmiàn qiánr de zhè (X) huà, quán xíng

"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 nèiguóhuà shuō de zuì shóu ne?

"Oh, that many… Then which language do you know best?"

63. A: 最熟，中国话。

Zuì shóu, Zhōngguóhuà.

"I'm most fluent in Chinese."

64. B: 哎，啊?

Ai, a?

"What?!"

65. A: 中国话。

Zhōngguóhuà.

"Chinese."

66. B: 啊，中国话！我说得也挺熟阿。说那外国话。


"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?
"You said fluent. Wouldn’t I be most fluent in Chinese?"

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

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

70. B: 都差不多？
Dōu chàbùduō?
"All the same?"
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。


B: 那么这个英国话，您会吗？

Nàme zhège yīngguóhuà, nín huì ma?

“Then what about English? Do you know that?”

A: 哈，太会了，太会了。

Hā, tài huì le. Tài huì le.

“Hah, do I ever!”

B: 真的？

Zhēnde?

“Really?”
75. A: 阿
A.
[ʔʌː]  "hmm"

76. B: 那(…)那我跟您学点儿行吗?
Nà… nà wǒ gēn nín xiáo diǎnr xíng ma?
[nʃ həʔ n̥loyː kənɻ̩ nɪn˥˩ ʔɛdəɻ tʃuiɻ ɕuŋ˧˥]  "Then... can I study with you?"

77. A: 行哇。
Xíng wa.
[ʃiəŋ˧˥ uɐː˥˧]  "Sure!"

78. B: 啊?
Á?
[ɦəː]  "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…
[ɻeɻəʊ˨˩ ɲɻɻeɪn˥˧ nɾɻɻ eɻəʊ̩ nɾɻɻ sɻəʊ˨˩ tɕuŋ˧˥]  "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?
[naɛ˨˩ tʃəə̝˧]  "Any day?"
81. A: 啊
Ā.
[ɦa:]  "hmm"

82. B: 甭哪天了，就今天在这儿学学。
[ˌpɒŋイン₃ɪə̹˧˨˦ˈɕɪɑːʊŋ]  "Never mind ‘any day’. I want to learn some here! today!"

83. A: 在这儿学不行。
Zài zhèr xiáo bù xíng.
[tʂəɭˈɕɪʌŋ]  "That’s not possible."

84. B: 怎么?
Zěnme?
[tʂamɿˈmaːɣ]  "Why?"

85. A: 这儿学不了。
Zài zhèr xiáo bù liǎo.
[lˈtʂəɭˈɕɪʌbɿˈlɑːɿ]  "You can’t study here."

86. B: 那怎么?
Nà zěnme?
[nɔɿɿˈtsɔmɿˈmaːɿ]  "Why is that?"
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。


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 huǐr bǎ yīngguóhuà quándōu xiáohuì le.
89. A: 你最快了，你得学半年。
Nǐ zuì kuài le，nǐ děi xiáo bàn nián.

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

90. B: 哎，不是，我们不学那么多啊，我就学一两句呀。
Ā，bù shì。Wǒ men bù xiáo nà me duō a。Wǒ jiù xiáo yī liǎng jù ya。

"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 gànmá ne？Xiáo yì liǎng jù yǒu shénme yòng a，yě yòng bù shàng a。

"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: 呃——，是那样儿啊，我跟您学，让您听着我的喉音怎么样，说这喉音得能学外国语，我将来就跟您多学。
Ā——，Shì nèiyànr a，wǒ gēn nín xiáo，ràng nín tīngzhe wǒ de hóuyīn zěnmeyàng。Shuō zhè hóuyīn de néng xiáo wàiguóhuà，wǒ jiāng lái jiù gēn nín duō xiáo。

"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。
Dui.
[t¹əɪ]
"Right!"

94. B: 嗯，这喉音不好，我就不学。
Ng, zhè hóuyīn bù hǎo, wǒmen jiù bù xiáo.
[mɚ tʂəŋ ɿ iɿ ɿ iɿ bɤ ɿ ɿ iɿ ɿ iɿ]
"If my guttural sounds are no good, we won’t need to study any more."

95. A: 对，对队队，这叫对。
[t⁰əɪ] ɿ t⁰əɪ ɿ t⁰əɪ ɿ t⁰əɪ]
"Right, right. That’s right."

96. B: 这怎么样呢?
Zhè zěnme yàng ne?
[tʂʰ eɿ tʂəmɿ meɿ aŋɿ nəɿ]
"How is that?"

97. A: 这个外国话，主要就这音。
Zhège wàiguóhuà, zhǔyào jiù zhè yīn
[tsɿ eɿ vəɿ oɿ tʂəɿ tʂəɿ tʂəɿ jɿ n]
"The really important thing when studying foreign languages is the pronunciation."

98. B: 啊啊。
A… a
[ʔa ʔa]
"Yeah..."
99. A: 你字儿对了因不对不行。
Nǐ zìr duì le yīn bú duì bù xíng

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

100. B: 哦。
Ō…

"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 xiáo wàigúoyǔ, xiáo hǎo jǐ nián, gēn wàiguóreń shuō, wàiguórén bù dǒng. Zhè zěnme huí shì a? yīn bú duì.

"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。
Ō…

"Ah…"
103. A: 音不对。
Yīn bú dui.
['jin⁶ bu⁴ tue.']
"The pronunciation is wrong!"

104. B: 那您要教我呢?
Nín yào jiāo wǒ ne?
['nin⁴ jia⁴ wou⁴ ne']
"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 děi dui, Nǐ xué nēiguóhuà?
['wo⁶ jia⁴ a, jia⁴ nǐn⁴... jin⁶ de⁴ du⁴, ni⁵ xué⁴ ne⁴ i⁴ gwo⁴ hwa⁴']
"I'm going to teach you… What language do you want to learn?"

106. B: 英国话。
Yīngguóhuà.
['jιŋgwo⁴ hwa⁴']
"English"

107. A: 教你英格兰的正音。
Jiāo nǐ yīnggélán de zhèngyīn.
['jia⁴ jia⁴ ying⁴ gwo⁴ l⁴ de⁴ zhen⁴ gi⁴ en⁴']
"Then I’ll teach you correct Oxford pronunciation"
108. B: 你瞧？那你受累吧。

Á? Nín qiáo? Nà nǐ shòulèi ba.

"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?

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

110. B: 啊…对话是什么意思?

A… duìhuà shì shénme yìsi?

"Ah… What’s a dialogue?"

111. A: 对话，问答话了，打听道问路啊、买东西、问价钱哪。问答话了。

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

"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:

\[
\begin{align*}
/ou/ & \rightarrow [o:] \\
/ai/ & \rightarrow [æ:] \\
/iao/ & \rightarrow [ɐː]
\end{align*}
\]

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

113. A: 对话。
Duìhuà.
"Dialogue..."

114. B: 单句是什么?
Dān jùr shì shénme?
"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 jùr jiùshì wèn shēngzì le, jiùshì míngcí le. Zhuōzi zěnme shuō a, nín qǐng zuò zěnme shuō ya, zàijiàn zěnme shuō ya.
"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’..."
117. A: 就是这个。
    Jiùshì zhège.
    
    "Things like that."

118. B: 啊，一句一句的了。
    A, yì jù yì jù de le.
    
    "Ah, so it's one sentence at a time?"

119. A: 对，这是对，对了。
    Duì. Zhè shì duì… Duì le.
    
    "That's right, that's right."

120. B: 那我学单句儿了。
    Nà wǒ xiáo dānjùr le.
    
    "Then I’ll learn some simple phrases."

121. A: 学单句儿也行。
    Xiáo dānjùr yě xíng.
    
    "You can study simple phrases if you want."

122. B: 那这行。
    Ā zhè xíng.
    
    "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.  
[ʂə́ ɿ ʂə́ ɿ ɿ ]°"It’s perfectly all right to study simple phrases.”

124. B: 那您受累吧。  
Nà nín shòu lèi ba.  
[na nǐ ʂə́ ɿ ]°"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.  
[ʔə́ ʃ ː ʔ ə́ ʃ ː ʔ ɿ ]°"Ok, how about this…”

127. A: 别胡来，别胡按啊。  
Bié hú lái, bié hú'àn a.  
[n pɔ́ ɿ xuː ʃ ə́ ʃ ː xuː ʃ ə́ ʃ ː ]°"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?
[ˌpʰeː˥˩ xuː˧ ai]
"Mess it up?!"

129. A: 哦，你打算学，打算揭我。
Ō, nǐ dǎsuàn xiào, nǐ dǎsuàn jiē wǒ?
[ʔəɪ˩˦ˌnǐ tɐː˨˩˨˩ˌɕɪɐʊ˩˧˥ˌtɐː˨˩˨˩ˌəɪ꜊ˌɕɪɑo˧˥ˌoãː˨]61
"That’s right. Are you planning to study or are you planning to cause trouble for me?"

130. B: 你瞧?
Nǐ qiáo?
[ni˨˩˧ ˌʨʰɐː]61
"Excuse me?!"

131. A: 啊?
A?
[w:ə]
"Huh?"

132. B: 打算跟你学啊。
Dǎsuàn gēn nǐ xiáo a!
[ˌtəː˥˩ sɐn˥ ni˧˥ ɕ’oa˧ ai]61
"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.
A: 哎，你要打算学，你痛痛快儿快儿，好好儿学。你打算歪心眼儿，我别扭，当这么些人你给我来个难看，你找我说不上来，你诚心把我问住，趁早别的学。嗯，你要学，你痛痛快儿快儿好好儿学，你按规矩问，你就打算问，按规矩来，别胡来你行。

B: 按照规矩学啊?

A: 恩，按规矩问。

"If you’re going to study, then you study hard and for real. If you’re really just looking for trouble you’ll make me look bad in front of all these people. If you are going to ask something that you know I can’t teach you, then it’s better if you don’t ask at all. So if you are going to study, you have to be sincere and study hard. You have to ask me according to the rules. If you are going to ask, you do it by the rules. It’s all right as long as you don’t mess up"

"Study according to the rules?!"

"Yes, ask according to the rules."

133. A: 哎，你要打算学，你痛痛快儿快儿，好好儿学。你打算歪心眼儿，我别扭，当这么些人你给我来个难看，你找我说不上来，你诚心把我问住，趁早别的学。嗯，你要学，你痛痛快儿快儿好好儿学，你按规矩问，你就打算问，按规矩来，别胡来你行。


"If you’re going to study, then you study hard and for real. If you’re really just looking for trouble you’ll make me look bad in front of all these people. If you are going to ask something that you know I can’t teach you, then it’s better if you don’t ask at all. So if you are going to study, you have to be sincere and study hard. You have to ask me according to the rules. If you are going to ask, you do it by the rules. It’s all right as long as you don’t mess up"
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 fǎr xiáo. Shénme jiào ànzhào guījù wèn na? Zhè yīngguóshū méi niàn guo, méi xiào guo yíngguóhuà, tóu yī jù wǒmen xiān wèn wén shénme zhème bù zhīdào.

"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.

"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?

"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?

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

---

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

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

64 Nin the syllable [ɔm], 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…

"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


"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?

"Do the foreigners have words that don’t exist?"
143. A: 那当然呢。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, zěnme shuō. Nà zěnme shuō a?

"Of course they have! If you ask me how foreigners say… ‘Baodu’. How can I answer that? What would I say then?"

144. B: 哦，对了。

Ō, duì le.

"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…

"Do you really think foreigners have 'baodu'? Or 'douzhi'? Even I wouldn’t know that."

146. B: 反正，普通有的东西。

Fǎnzhèng, pǔtōng yǒu de dōngxi.

"Well, normal stuff then."
147. A: 你得有的。
Nǐ děi yǒu de.
[ni˨˩tɛi˨˩˨tɕə˧˦˦][72]
"It has to be words that exist."

148. B: 哦，那... 那成。
Ō, nà... nà chéng.
[ʔɔːǁnɐ˨˦ǁnɐ꜌ٍʈʂʰɜ̃ː]
"Ok... That's all right."

149. A: 嗯只要有这/着就成。
Ēn, nǐ zhǐ... zhǐ yǒu zhe jiù chéng.
[ɛ˧ ni˨˩tsɿǁtsɿ˦tʂɿ꜌ɪɔʊ˨˩ʈʂɿ˨˦ɗɪʃə˥˩ˌʈʂʰʌŋ˧˥][72]
"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 Yīngguórén guǎn źhe ġe màozi jiào shénme?
[ʔəːǁʔʌ꜌‿oɕɝ˥˥˥ˈɖʐeɪ˥˩ǁʥʏ꜌‿bɿǁʈʂe꜌ ]
"Let's study this word then: How do foreigners say 'hat'?"

151. A: 嗯?
É?
[hɔ][72]
"What?"

---

[72] The tone of [ni˨˩] is difficult to hear clearly, and the sound quality allows no reliable instrumental analysis. It may also be [I].
152. B: 管这个帽子叫什么？
Guǎn zhège màozi jiào shénme?
[kʷʐɛŋ dʐɛk̪ jɛawʔ_tsal ʨiaʊʔ]73
"How do they say 'hat'?"

153. A: 先学别的，先学别的，先学别的。
Xīan xiáo bié de, xiān xiáo bié de… xiān xiáo bié de.
[ɕɪ̃˥˥˥ɕɪɒŋ˥˥˥ˈbiɛː˨˧tɔɿ || ɕɪ̃˥˥˥ɕɪɒŋ˥˥˥ˈbiɛː˨˧tɔɿ || (0.98) || ɕɪ̃˥˥˥ɕɪɒŋ˥˥˥ˈbiɛd˧] "Let's do something else first. Let's do something else."

154. B: 啊?
Á?
[ʁɐː] "Huh?"

155. A: 先学别的。
Xīān xiáo bié de.
[ɕɪ̃˥˥˥ɕɪɒŋ˥˥˥ˈbiɛː˨˧tɔɿ] "Do something else first."

156. B: 怎么?
Zěnme?
[tsɿməɿ] "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，你这玩意。
Gàosu nǐ xiān xiào bié de. Yào bù zěnme zhège rén kěhèn ne, nǐ zhè wányír.

"I'm telling you to do something else! That's really why I hate this guy..."

158. B: 哦?
Ó?

"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 jiāo rén wàiguóhuà de, húlái. Nǐ zhè shì hú lái, nǐ zhè shì húlái le. Húlái.

"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。
Zhège màozi zhè...

"So a hat..."
161. A: (XX) 胡来！胡来！
(XX) húláí. Húláí.
[(XX)75 ʃʊɻ˨˩˧˨ ʃʊɻ˨˩˧˨ (3.58) || ʃʊɻ˨˩˧˨]
"You're just causing trouble!"

162. B: 噢哈！
Ha.
[ʔɔː χɔ]
"Hah!"

163. A: 胡来。
Hú lái.
[ʃʊɻ˨˩˧˨]
"Causing trouble!"

164. B: 合着帽子这句您不会是把。
Hézhe màozi zhè jù nín bú huì, shiba.
[ʃɻ˨˦ɻɔ ʃəɻ˦˨ɻ tɕiɻ dɻɻ˨˩˧˨ tɕeɻ tɕɛɻ ßimɻ puɻ˦˨˦ ʃɐɻɻɻ ʃəɻɻɻ bəɻ]
"This word 'hat'... You don't really know it, do you?"

165. A: 嗯？你怎么知我不会呀？
Nǐ zěnme zhī wǒ bú huì ya?
[ʃɻʅəɻ mêɻʅ ʂɻʅ ʂɻʅ ɻəɻ ʂɻʅ puɻ˦ɻ ɕəɻɻɻ eɻɻəɻ bəɻ]
"How do you know that?"

75 Speaker A interrupts at this point, and I have not been able to make out the first syllables of this turn, nor the last syllable of the previous turn. My personal interpretation is that speaker A just produces some nonsense sounds for the sake of interruption alone (some equivalent to “tut-tut”, or “apapapa”).

76 This word, which I have presumed to be zeme 怎么, is pronounced with both a different initial and a different tone from what I expect; [ʂəmɻ]. Note that the tone of the preceding syllable ['niɻ] apparently undergoes tone sandhi, as if [ʂəmɻ] were a tone 3.
166. B: 你瞧这意思，我跟你学这帽子你（XXX）“胡来”。学这帽子怎么叫胡来呢？这帽子这句您是会是不会？

Nǐ qiáo zhè yìsi, wǒ gēn nǐ xiào zhè màozi nǐ (XXX) húlái. Xiào zhè màozi zěnme jiào húlái ne? Zhè màozi zhè jù nín shì huì shì bù huì?

167. A: 你干嘛非得学帽子呢？

Nǐ gànmá fēiděi xiáo màozi ne?

168. B: 啊，我想起这帽子来了。

Ā, wǒ xiǎngqǐ zhè màozi lái le.

169. A: 帽子没法儿说啊。

Màozi méifǎr shuō a.

---

77 This passage appears to be pure nonsense. Phonetically it is something like [ʃə əŋ ɪ ʔ ə ʂ u ɔː ˥˥˥ ɑ˦˨].
78 This final /n/ seems to assimilate the place of articulation of a following /shi/, but in the actual speech /shi/ has been omitted, and only the final [n] remains.
79 [ʃʰ ə ʂ u ɔː] is another example of two syllables that could just as well be analysed as a single syllable.
170. B: 嗯，啊？这帽子没法儿说啊？
Äi, ā? Zhè màozi méi fār shuō a?
"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。
Maòzi duō le. Cǎomào, pímàozi, zhānmàozi, shuō něige, něige duì ya?
"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，那就甭分了，就这普通，这就帽子。
O… Ā, nà jiù bèng fēn le. Jiù zhè pǔtōng, jiù zhè màozi.
"Ah… Well, don’t separate them. Just an ordinary hat."

173. A: 就是帽子啊？
Jiù zhè màozi a?
"Just a hat."

174. B: 嗯。
Ai.
"Yes."
175. A: 哪国话？
Néiguóhuà?
哪国话？
[ˈnɛɪɡʊˌhuːə] "Which language?"

176. B: 英国话
Yīngguóhuà.
英国话
[ˈjɪŋɡʊˌˈχuə] "English"

177. A: 英国话？
Yīngguóhuà?
英国话？
[ˈjɪŋɡˌnoʊˌˈxʊə] "English?"

178. B: 嗯。
n.
嗯。
[m] "Yeah."

179. A: 帽子。外国人说话用喉音。
Màozi a? Wàiguórén shuō huà yòng hóuyīn.
帽子。外国人说话用喉音。
[ˈmɑoˌdzu̯]; [ˈvæəˌɡʊˌtʃən]; [ˈʃoʊˌtoʊˌrɑːŋ] "Hat... Foreigners use throat sounds when they talk."

180. B: 啊，啊。
A, a.
啊，啊。
[ˈʔɔʔ] "Yeah, yeah."
181. A: 咱们中国话儿用丹田音。
We Chinese use sounds from our abdomen.

182. B: 是。
Right.

183. A: 它是用喉音。
They use throat sounds.

184. B: 啊?
Ok.

185. A: 帽子?
Hat?

186. B: 嗯。
Yeah.
187. A: 啊，我先拜托各位啊。
āi, wǒ xiān bāituō gèwèi ā.

“Then, I have a request for everybody.”

188. B: 干嘛?
Ganma?

“What?”

189. A: 我可不知道哪位会这个外国语啊，我也不知道哪位同志念过英文，啊，
是如果我说得对，或者是不对，请您哪，千万别管
Wǒ kě bù zhīdào něiwèi huì zhège wàiguóyǔ ā, wǒ yě bù zhīdào něiwèi tóngzhì niàn
guo Yīngwén ā. Shì rúguǒ wǒ shuō de duì, huòzhě shì bú duì, qǐng nín ā, qiānwàn bié
guǎn.

“I don’t know who among you have studied foreign. Neither do I know if any of you
have learnt English. So, whether I say this correctly or not… Please… don’t worry
about it.”

190. B: 啊？别管？合着对不对呢，别管？
Ái, bié guǎn? Hézhe duì bú duì ne, bié guǎn?

“Huh? Don’t worry about it? Don’t worry if it’s right or wrong?!”

80 Speaker A turns to the audience.
191. A: 帽子啊？
Màozi ā?
[ˈmaʊ̯ dzoɻ aː]  
“Hat?”

192. B: 哦——
Ō…
[ʔɔː]
“Hmmm…”

193. A: 外国话用喉音。
Wàiguóhuà yòng hóuyīn.
[ˌvaːɪˈɑːlˌtʂɿ]  
“Foreigners use throat sounds.”

194. B: 那您只说吧。
Nà nín zhǐ shuō ba.
[ʔaɿnɪn]  
“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.

198. B: 哦，这个帽子叫。。。帽帽?
Ō, zhège màozi jiào māomāo?

---

81 Something is said here, but because of the applause and laughter it is difficult to hear exactly what.
82 This part is inaudible.
199. A: 有门儿，行。  
Yǒu ménr, xíng.  
[ʊə ˧˥ ˥˥˧]  
“That’s right! You’re getting the hang of it!”

200. B:  啊。  
Ā…  
[ɑ̃ː]  
“Uh…”

201. A: 先记得啊。  
Xiān jìzhe a.  
[ɛ́n ˧˥ ˥˧]  
"First you have to remember that!"

202. B:  啊，啊。  
A a.  
[ʔeɿ]  
"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 xiá a, chū xué, chū xué zhà liàn, xiān xiáo zhè hǎoxiáo de. Xiān bié dào wài tou shuōzhe a.  
[ɛ́n ˧˥ ˥˧]  
"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 [ʔə] illustrates why the neutral tone is not analysed simply as an unstressed syllable. [ʔə] is a neutral tone and is also stressed.
204. B：嗯？那怎么？
Ai？Nà zěnme？
[ʔzǐ| | (1.81) | n̩ɭ| ʦəm˨˩˧ mɔ́l]
"What? Why is that?"

205. A：嗯！先别到外头说。
Ài，xiān bié dào wàitóu shuō.
[ʔzǐ| | eɬem˦˨ b̥e˥˧ taoɭ| ʰa² tɔol ʃoː˥]
"That’s right. Don’t let anybody hear you say it!"

206. B：为什么？
Wèishénme?
[ʔveːɭ ʃəm˨˩ meː]
"Why?"

207. A：因为外国话你会得少啊，你见人你先别露阿，你见人一露外国话，人啪
问你一句，你没有，多寒碜，不会得少吗？个人就在肚子记着
Yīn wèi wàiguóhuà nǐ huì de shǎo a，nǐ jiàn rén nǐ xiān bié lòu a。Nǐ jiàn rén yí lòu
wàiguóhuà，rén pā，wèn nǐ yí jù，nǐ méiyōu。Duō hánchen。Bú hui de shǎo ma？Géren
jiù zài dùzi jizhe.
[jiːŋ˨˩˧ vɔɪ˧ ,can˥˧ ɡo˧ ,n̩ɭ| ʃəm˨˩˧ n̩ɭ ]
"Because you still know very little foreign, so you shouldn’t just blurt it out when you
meet someone. If you meet someone and just blurt out some foreign, they’ll just reply
with a question or something, and then you’re finished! How embarrassing is that?
You still know very little, am I right? Don’t forget it!"

84 This expression géren 个人 “you, yourself” is a North Chinese colloquialism, and should not be confused with
gérén 个人 “individual”. The sentence “géren jiù zài dùzi jizhe” means something like “make sure you keep it
with you (and don’t forget it)”.

123
208. B:  嗯。。。哎
O… Ai.
[ʔɔː˧˨ˈʔɤi]
“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。
[ˈpʰeː˨˦ɒʰo˥˥ || tə˧˥ ʰnɛɭ˧ ʰpʰeː˧ɭ ʰʷoː˧˥ ʔə˧˥ || ʰtʰɛː˧˥ ʰvan˥ pʰeː˧ɭ təo˧˥ ʰvaɭˌpʰoː˧˥ ʔoː˧˥ || ʰtəɪɭˌtəɭ˩ ʰtəɪɭˌtəɪɭ ʰbʰɭ˧ʔ oː˧˥]
“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…哦一一，嗯?我们穿的这个袜子叫什么?
Ô, Ái wǒmen chuān de zhège wàzi jiào shénme?
[ʔoː˧˨ || ʔəɪɭ “oʊmɭ ʰtsʰɛn˥ ʰnɛɭ˧ ʰdzekɭ kəɭ ʰtaɭˌdzɭɭ ʰdzɭɭ ʒəmɭ]
“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。
Wāwā. Zhèwányìr qíshí méi ma, zhè wányìr méi ma.
[ʰæː˨˦ɛɭɭ ʰməː˨˦ ʰmeɭɭ meːɭ || ʰtʰɭɭɥɭ ʰɭɭɭ ʰmeɭɭ meɭɭ]
‘Wawa’. There’s nothing to it really! You can see now that there’s nothing to it.”

212. B:  啊。
A…
[ʔaːɭ]
“Oh…”
213. A: 一层窗户纸儿。
Yì céng chuānghùzhǐr.
"Once you get the basics it all falls into place."^86

214. B: 哦，这个袜子脚袜袜?
O, zhège wàzi jiào wàwa?
"So 'a sock' is called 'wawa'?"

215. A: 有门儿。
Yǒu ménr.
"You're getting the hang of it!"

216. B: 啊——。那我们吃的这个包子叫什么?
a… Nà wǒmen chī de zhège bāozi jiào shénme?
"Then what about 'a stuffed bun', the kind we eat?"

217. A: 包包儿。
Bāo baor.
"'Baobao'."

^85 This syllable is not audible enough for a completely accurate analysis. It is therefore possible that the initial glottal plosive [ʔ] does not occur.

^86 This line literally really means "One sheet of window-paper". It is the first part of the Chinese saying “yìcéng chuānghùzhǐr, yitōng jiù pò 一层窗户纸，一捅就破” . "Like a thin veil, a single touch will penetrate it".
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 bāobao? Nà wǒmen chī de nèige jiǎozi ne? Jiǎo jiāojiāo, shìma?

“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.

“Hey, that’s right! You do know some foreign!”

220. B: 我呀！
Wǒ ya.

“Who, me?”

221. A: 不对了吗？
Bú duì le ma?

“What you just said was right, wasn’t it?”

87 The syllable [ɻɐ] is one of the few cases in this data set where r-coloring of a vowel seems to work progressively.
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；哇，我吃那包包”这叫外国话，这叫？

Zhè jiào wài guó huà, zhè jiào？Zhè xiàng wǒmen jiā de xiǎohái gāng xiáo huà. Nèige xiǎohái zhǐ nà dōngxi jiào bù míngbai, guǎn zhè máozi zhǐ jiào ‘wa, wǒ yào nà màomàor’, ‘wa, wǒ chuān nà wàwar’, ‘wa, wǒ chī nà bāobao’. Zhè jiào wàiquóhuà, zhè jiào？

“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?”

A: 漫漫儿来呀！

Mànmànér lái ya.

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

B: 啊。

A.

“Yeah...”

A: 先记着。

Xiān jìzhe.

“First make sure you remember it!”
226. B: 是啊。
Shì a.
[ʂʅː˥˩˧˩꜌]
“Right...”

227. A: 啊，千万别到外边儿说着啊。
Ai, qiānwàn bié dào wàibiānr shuōzhe a.
“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“我吃那饺饺”。我怎么了，我!
Wa, wǒ shì bù néng shuō zhe. Wǒ jīn nián sìshí duō sui le, jiàn rénjia wǒ shuō wǒ chǐ nà jiǎojiào. Wǒ zěnme le, wǒ!
[ʊɔ˨˦ʊo˨˩’ʂʅː˥˩pə tʂʰʅː˥˥ Scrolls: (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


------ 1997, 语言问题 yǔyán wèntí [Problems of Language], Bēijīng, 商务印书馆 Shāngwù Yinhshūguǎn.


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.


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.


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 / Kàtōng 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.


“张庆森 Zhāng Qìngsēn”, 百度百科 bǎidùbǎikē, 2 June 2010, accessed 20 November 2010,


133