The Pronunciation of Mandarin Chinese

According to the "canIPA Natural Phonetics & Tonetics Method"

SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: L-LIN/01

Tesi di dottorato di Marco Cerini, matricola 955751

COORDINATORE DEL DOTTORATO
Prof. Alessandra Giorgi

TUTORE DEL DOTTORANDO
Prof. Luciano Canepari
The Pronunciation of Mandarin Chinese
According to the IPA Natural Phonetics & Tonetics Method

0. Introduction
1 o.1. The official romanization – Hànyù pīnyīn fāng'ān
2 o.2. The finals
2 o.3. Phonotactics

1. Vowels and vocoids
5 1.1. Vowels
8 1.2. Phonemic conciseness vs realism: the interphonemic choice
11 1.3. On the phonemic status of -ı
16 1.4. Diphthongs
17 1.5. Vowel reductions

2. The r grameme and the so-called érhùà phenomenon
21 2.1. Origin and use of érhùà
22 2.2.4. Articulatory characteristics of érhùà
25 2.5. Taxophonic variation in [ŋ] and [Vŋ]
27 2.6. Taxophonic variation in [Vŋ] and [Vŋ]
28 2.7. Various examples of érhùà, according to our ‘suggested’ pronunciation scheme
30 2.8. Frequency and acceptance of érhùà in cultivated speech % non-neutral pronunciations
31 2.8.3. An ‘international’ pronunciation scheme of érhùà (for ‘practical’ teaching purposes only)

3. Consonants and contoids
34 3.1. Main features of Mandarin consonantal system (with a table of neutral taxophones)
35 3.2. Nasals
36 3.3. Stops and stop-strictives
39 3.4. Constrictives
42 3.5. Approximants
44 3.6. Laterals
45 3.7. Casual speech
4. **Chinese tonemes, tones, taxotones & stress**

4.1. Tonality as a phonologically distinctive feature

4.1.2-3 An introduction to the four marked tonemes and the so-called ‘zero’ (unmarked) toneme

4.2. Taxotones and stress

4.3. Primary (or ‘full’) taxotones of the four marked tonemes

4.4. Rythmic groups and taxotones

4.5. Further taxotones, including /$\langle$$/ [}$, $\langle$, /$\langle$$/ [}$, and /$\langle$$/$ [}$, $\langle$

4.6. Taxotones in three-syllable sequences, including /$\langle[/}$ ~ [}$

4.7. The so called ‘zero’ toneme, or qíngshēng

4.8. Tonal behaviour in /$\langle$$/ sequences

4.9. Tonal behaviour of bù ‘not’, méi ‘not’, yī ‘one’ (plus qī ‘seven’ and bā ‘eight’)

5. **An introduction to Chinese intonation**

5.1. Tonality and intonation

5.2. The oxford tonogram

5.3. Studying and learning intonation

5.4. The intonation of English

5.5. Intonemes and preintonemes

5.6-9. English intonational patterns

5.10. The intonation of Mandarin Chinese

5.11. The oxford symbols inventory to annotate Chinese intonation

5.12. Mandarin Chinese preintonemes

5.13. Mandarin Chinese intonemes

5.15-18. Various kinds of questions and their intonational patterns

5.19. The continuative intoneme

5.20. Examples illustrating Mandarin Chinese intonation behaviour

5.21. Emphasis

5.22. Parentheses and quotations

5.23. Natural perception vs acoustic measurements

6. **Phono-tonetic analysis of extended texts**

6.1. Phono-tonetic transcription of Aesop’s The North Wind and the Sun

6.2. Phono-tonetic transcription of Lìuyōu yúdī

7. **Phonetic transcriptions in current handbooks**

7.1. A proposal for a simpler, teaching-oriented transcription to be used in Chinese textbooks

7.2. Phono-tonetic transcription of Lìuyōu yúdī according to the ‘simplified’ scheme

7.3. Scarce attention to phonetics, and pronunciation in general, by most teachers of Chinese
7.3.1. An ‘enlightened’ use of offIPA in the old Linguaphone Institute textbook Chinese
7.3.3. The officially-sanctioned description of Mandarin Chinese as presented in the JIPA
7.3.5-6. Keeping offIPA ‘true to itself’

8. Peking
8.1. The Peking accent vs the Peking dialect
8.2. Classical Pekingese and contemporary Pekingese
8.3. Readily available examples of genuine Pekingese pronunciation
8.4. Fundamental traits

9. ‘Praarteries’
9.1. Usefulness of computer-aided analysis applied to phonetics and tonetics
9.2. Detecting consonantal traits
9.3. Detecting vocalic traits
9.4. Analysing tonality
9.5. A few sentences transcribed in canIPA, with their relevant Praat diagrams

10. Conclusions

Bibliography
**o. Introduction**

We provide the modern neutral pronunciation of modern Mandarin Chinese, or *pǔtōnghuà* /ˌpʰu̯-tʰoŋ-ɥu̯a/ [ˌpʰu̯-tʰoŋ-ɥu̯a], ie the ‘common language’, based on Peking dialect. We obviously use the official romanization system *pǐnyīn* /pʰin-jin/ [ˈpʰiŋ-ˈjın], accompanied by a careful phonetic transcription, to show and acquire a good pronunciation. ¹

At the beginning, a well meditated phonemic transcription will be provided too, in accordance with precise descriptive and teaching strategies. To improve readability, the phonemic transcription will be progressively omitted in the following chapters, wherever *pǐnyīn* suffices to highlight any relevant phonemic facts.

*The official romanization – Hányǔ pǐnyīn fāng‘àn.*

0.1. *Pǐnyīn* stands at midway between the Chinese characters and our phonetic transcriptions. This romanization scheme was developed, and officially adopted, in 1958, by the People’s Republic of China, and now is the worldwide *de facto* standard. Its full Chinese name is *Hányǔ pǐnyīn fāng‘àn*, where *pǐnyīn* roughly means ‘to put sounds together’. ² We shall present it here, even though, to be entirely loyal to the natural approach, any graphemic representation should come last, *after* the spoken language has been duly described and learnt.

The twenty-three *initials* are listed below, and arranged not alphabetically, but by logical phonic groups. Each is accompanied by its phonemic transcription and only one phonetic variant (the one used in fully-stressed syllables):

<table>
<thead>
<tr>
<th>Initial</th>
<th>Phonemic</th>
<th>Pinyin</th>
</tr>
</thead>
<tbody>
<tr>
<td>b /p/</td>
<td>[p]</td>
<td>p/ph/ [pʰ]</td>
</tr>
<tr>
<td>d /t/</td>
<td>[t]</td>
<td>t/th/ [th]</td>
</tr>
<tr>
<td>z /ts/</td>
<td>[ts]</td>
<td>s/s/ [s]</td>
</tr>
<tr>
<td>j /tʃ/</td>
<td>[tʂ]</td>
<td>x/x/ [ɕ]</td>
</tr>
<tr>
<td>zh /tʂ/</td>
<td>[tʂʰ]</td>
<td>sh/s/ [ʂ]</td>
</tr>
<tr>
<td>g /k/</td>
<td>[k]</td>
<td>h/h/ [ɦ]</td>
</tr>
</tbody>
</table>

¹. This information is obtained from the textbook *A Modern Mandarin Chinese Pronunciation* (Byron W. Chen and Susan F. G. Hui, 2008).
². According to *Chao (1955)*, p. 31, *pǐnyīn* roughly means ‘to put sounds together’.
0.2. The finals are 39 (with three occurrences of i and three different values), and we present them according to their pīnyīn graphic syllabic nucleus. This time, they are given in alphabetical order, even where their romanization has not been the happiest one (as we show in { }). The three finals which do not occur without an initial are marked with an asterisk (*):

A
a /a/ [a], ia, "ya /ja/ [ja], ua, "wa /wa/ [wa];
ai /æ/ [ʌ], uai, "wai /wæ/ [wa]; an /an/ [ʌn], ian, "yan /jæn/ [jʌn], uan, "wan /wan/ [waŋ], uian, "yuan (j/q/x + uan) /jæn/ [jʌn];
ang /an/ [ʌn], iang, "yang /jan/ [jʌn], uang, "wâng /wan/ [wæŋ];
ao /o/ [o], iao, "yào /jao/ [jæo].

E
e /e/ [ɛ, ɛɭ], ie, "ye /je/ [je], iê, "yue (j/q/x + ue) /jœ/ [jœ];
ei* /ei/ [ɛi], en /ɛn/ [ɛn], er /ɛr/ [ɛɭ];
eng* /ɛŋ/ [ɛn], ueng, "wâng /wæŋ, wən/ [wæŋ, wən].

I
i /i/ *{[^i, ji]}; (s/z/c) /ui/ [ui], (sh/zh/ch) /ui/ [ui] {we adopt ì, instead};
in, "yin /jîn/ [în, jîn]; ing, "ying /jîn/ [în, jîn];
iu, "wei /weî/ [wêî] {we adopt ìei would have been better than ui}.

O
o /o/ [ɔ, ɔɭ], ou /ou/ [ou] (ɔu); uo, "wo /wɔ/ [wo, wɔɭ] (wɔ, wɔɭ);
ong* /ɔŋ/ [ɔn], (/un) [un], iong, "yong /jɔŋ/ [jɔn]

U
u, "wu /u/, "wù /wù/ [u, wu]; û, "yu (j/q/x + u) /y/, "yû/ [y, ý];
un, "wen /wen/ [wən]; iun, "yun (j/q/x + un) /jən/ [jən];
iu, "you /jou/ [jou] (jou); we employ û whenever it stands for /y, ɻ/; besides, uen and iou would have been better than un and iu}.

Phonotactics

0.3. Chinese phonotactics obeys to very strict limitations: in fact, just 400-odd distinct syllables exist in the modern language, tonemes not considered. Not all acceptable permutations are used in fact, like shong, though dictionaries list various
words with zhong, chong and rong. Other syllables do not occur with certain tones, for historical reasons: for instance, there is no ër (er in the first tone). 4

Mandarin phonology has been analyzed according to very different criteria, but we maintain that any Chinese syllable consists of an initial element and a final one, traditionally called shēngmù /ɕm/, mu [ʃm] ‘mother of the voice’, and yùnmù /yn/, mu [yn] ‘rhyme of the voice’. 5

At the phonemic level, all initials are consonantal by their nature, even though, phonetically, a contoid is not necessarily present. In the case of close vowels /i, y, u/, then, we actually have /ji, yu, wu/ [j, y, w] 6 Before other vowels, our choice is /œ/: /œi, œy, œu/ (/œu/ does not occur) — which can be safely realized as such, [œ], or –less advisably— as [γ, fi, a, η] (: velar semi-approximant, lenis laryngeal approximant, uvular tap, and provelar semi-nasal). Nevertheless, when there is no risk of hiatus with a preceding vowel, /œ/ spontaneously reduces to ‘zero’, [ø].

In its full form, a final is a sequence of three parts: a glide /j, q, w/; a vocalic nucleus /i, y, u, e, α, σ, a/; and a coda, which can be vocalic /i, e, u, σ/, or consonantal /n, n, ŋ/. For example, combining /œ, w, a, σ/, we obtain /œa/ sha, /œσ/ shao, /œwa/ shua, /œα/ shang, and /œwan/ shuang. Also, glides can act as initials: /wa/ wa, /wan/ wang.

Many textbooks and dictionaries include a list of permitted syllables; one of the most exhaustive and reliable ones, in our view, is that provided by The Contemporary Chinese Dictionary, which is also the reference book with respect to the orthography of all Chinese words quoted in this work, including those featuring a ‘zero’ toneme %, the so-called érhuaè phenomenon. 7

Notes:

1 We assume that our readers have a working knowledge of the Canepari International Phonetic Alphabet — ‘caniPA’, and of the Natural Method, as thoroughly described in canepari, l. (1983) Phonetic Notation · La notazione fonetica, and in canepari, l. (2007) Natural Phonetics & Tonetics – articulatory, auditory, functional.

The caniPA inventory is a groundbreaking development of the ‘official’ IPA (cursorily: ‘offIPA’), not just a mere extension, and it is purposely conceived to provide a very high accuracy rate, both at the phonetic and the tonetic level. Particularly in transcribing tonal languages, the unique caniPA tonemic and tonetic notation finally makes it possible for the transcriber to annotate the exact tonality of single syllable and of entire utterances, in a readable and unambiguous form.

In this, we follow CANEPARI, L. (2007) *A Handbook of Pronunciation*, § 11.0.6 et seq, with minimal amendments.

However, the Peking dialect also features a few words with peculiar phonology, like *dèn*, which not only are absent in the official inventory, but also not *permitted*, even in theory. See also ZHOU, Y-M. (2002) *Xiandai Beijinghua yanjiu*, particularly §§ 3 and 6.

Contrary to Western tradition, where dictionaries commonly list words according to their initial letter, classical Chinese dictionaries generally considered single characters –instead of words– arranged according to their ‘radical’ components and the number of calligraphic traits. Pronunciation, at first, was indicated by means of homophony (e.g English *sea*, [now] pronounced like *see*).

Then, a mechanism called *fānqìè*, or ‘cross-cutting’, was developed: it consisted in a pair of characters, which indicated the initial and the ‘rhyme’, respectively (e.g a combination like *chē-huāng*, to express *chuāng*; or, in English, *she-lock*, to express *shock*). Hence the importance of keeping initials well distinct from the ‘rhymes’. The latter, moreover, were particularly important for classical poetry; not surprisingly, then, ancient poets used to refer to specifically rhyme-arranged dictionaries, which nowadays provide precious—but often vague and ambiguous—hints at the pronunciation of older forms of Chinese (among others, see PULLEYBLANK, E. G. [1991] *Lexicon of Reconstructed Pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin*).

The *Contemporary Chinese Dictionary* is also published in a Chinese-English bilingual edition, very convenient for those who are interested in Mandarin, but do not actually master the language up to the point of using a monolingual dictionary.

Again, *The Sounds of Chinese* also devotes quite a large amount of space to analyze the syllabic restraints determined by Chinese phonotactics (particularly, see § 5.3). We regard such an approach as slightly too theoretical, when compared to the highly pragmatic nature of our method; nonetheless, we would agree with many of the principles stated by the author, provided it is made clear that the phonetic reality—well understandably—is much more complex.
1. Vowels and vocoïds

**Vowels**

1.1. After thorough consideration, we have chosen to include eight vowels in our analysis of Mandarin phonology: /i, y, u, u; e, æ, o, a/. In fig 1.1-a, the various vocoids and their taxophones —ie contextual allophones— are grouped by phoneme; while, in fig 1.1-b, we present the same taxophones arranged in ‘functional’ groups: /V/, Vn, Vw/. Let us start, then, by systematically considering each phoneme and its most typical realizations in neutral pronunciation (further —less representative— variants will be shown below, as well, more for recognition than for actual use).¹

**fig 1.1-a.** Mandarin fundamental vocoids and their taxophones, grouped by phoneme.

```
[ɛi, ɛi, C, y,y, u, u; e, æ, o, a/; [u, -ui(u)], C, u, -ui(u)]
[bj, b, C, -i(0), m, v]\n[un, wun, an]
[j, j(a); qe, qe(a)]
[jei; uai]
[wair]
[a, a(a), an; ja, ja(a), jan; wa, wa(a)]
```

1.1.1. First, we find the close front phoneme /i/: /ɛi/, ɪj, ɪn/ [ɛi, ɪj, ɪn] (C/y)i, (C/y)in, (C/y)ing: [jī] / jī/ yī, [qī] / qī/ xī, [jǐ] / jǐ/ yǐ, [jǐn] / jǐn/ yǐn, [tɕǐn] / tɕǐn/ jǐn, [tiɛn] / tiɛn/ yīng, [tsiɛn] / tsìn/ jīng. Oscillations are possible and not infrequent: [ɛiûn, ɪjûn, ɪjûn] and [ɛiûn, ɪjûn, ɪjûn]. The approximant /j/ is seldom kept even when a proper initial is present: in that case, and where deemed necessary, the actual pronunciation of such syllables as lin(g), pin(g), ding could be better phonemized as /ljin, -n/, /phjìn, -n/, /tjìn/ (or even /ljìn, -n/, /phjìn, -n/, /tjìn/, which nevertheless should be reserved for more marked variants, as those found in heavily Peking-accented Mandarin).²

1.1.2. Secondly, there is the close front-central rounded phoneme /y/, with a fairly limited distribution: /ɛiy, ɪjyn/ [ɛiy, ɪjyn] (frequent variants with n: /ɛin,
The official orthography adopts \( \text{ü} \) only in \( nü \) and \( liu \), while leaving \( u \) unmarked in \( ju(n) \), \( qu(n) \), \( xu(n) \), \( yu(n) \). We see no advantage in doing so, and we prefer to keep that friendly \( \text{ü} \) anywhere: \( [\text{tɕy}] / [\text{tsy}] \ jü \), \( [\text{tɕʰy}] / [\text{tshy}] \ qū \), \( [\text{ɕy}] / [\text{sy}] \ xū \), \( [\text{ɥy}] / [\text{ɥy}] \ yū \), \( [\text{tɕy}, \ \text{tɕʰy}, \ \text{tɕu}] / [\text{tsy}] \ jūn \), \( [\text{tɕʰy}, \ \text{tɕʰu}, \ \text{tɕυu}] / [\text{tsy}] \ qūn \), \( [\text{ɕy}, \ \text{ɕu}, \ \text{ɕυu}] / [\text{sy}] \ xūn \), \( [\text{qy}, \ \text{qu}, \ \text{qυu}] / [\text{qy}] \ yūn \).

More conveniently, \( pinyin \) could have used \( zü \), \( cü \), \( sü \), \( yü \), by phonemically interpreting these syllables as we do: \( [\text{tɕy}, \ \text{tɕʰy}, \ \text{ɕy}] / [\text{tsy}, \ \text{tshy}, \ \text{sy}] \). Indeed, \( [\text{ɥy}] \) could even be \( /jy/ \); but since the phoneme \( /ɥ/ \) is there, it is more adequate to use it. In this way, as it seems convenient, it could also be used to show—even phonemically—the possible oscillations between \( [\text{ɥy}, \ \text{ɥu}, \ \text{ɥυu}] / [\text{ɥy}]/ \), and \( [\text{ɥu}, \ \text{ɥυu}] / [\text{ɥu}] / \). (Therefore, it could have been possible to write \( ziong \), \( ciong \), \( siong \), too, more internationally, instead of the official \( jiong \), \( qiong \), \( xiong \).)

1.1.3. Then we find a back rounded vowel, \( /u/ \) \( [u] \) (not to be confused with British English ‘\( /u/ \)’—better: ‘\( /uu/ \)’—whose various taxophones are rather mid-back, and


As far as //jan, ɥan// are concerned, we think they are better represented by /jen, ɥen/ (in spite of the official rendering with an): [ˈjen] /ˈjen/ yán, [ˈtɕɛŋ] /ˈtɕen/ dián, [ˈɥen] /ˈɥen/ yuàn, [ˈtɕʰɥan] /ˈtʰɥen/ qiuán.

1.1.5. The vocalic inventory of neutral Chinese continues with three mid vowels: /e, ɔ, ɔ/1. Since the interjection è is rather theoretical, the only environment where we find /e/ [e] alone (ie not combined in diphthongs, /ei, ae/) is in the two sequences /je², ɥe²/ [je, ɥe].

For the former, pīnyīn writes ye, when lacking a graphic initial; otherwise, ie /tʃje, ʨhje, sje, nje, lje/ [ʨje, ʨhje, ʨe, nʃje, lʃje] jiè, qiè, xiè, niè, lìe, and so on. Any combination is permitted, except fie and zhie, chie, shie, rie (since the post-alveolars cannot match with /j/).

As far as zie, cie, sie and gie, kie, hie are concerned, they have historically merged into the single series jie, qie, xie, seen above. Interestingly, some characters in the Peking Opera, Jīngjù /ˈtsiŋjù/ [ˈdʒiəŋjɪ], often chant their lines using a form of stage pronunciation which still retain archaic syllables, like zie and siang.

For the latter sequence, we have jue, que, xue and yue, but liè and nüè (no other combinations occur, even those –like mūe– that might be articulated without problems). As with ü(n), ‘our’ pīnyīn will always show üe, instead: /ʃuɛ, ʦhुɛ, ʨuɛ, nɥɛ, lɥɛ/ [ʨuɛ, ʦhуɛ, ʨuɛ, ɲɥɛ, lɥɛ] jìuè, qìuè, xuè, niè, liè.

1.1.6. The mid-back vowel /ɥ/ plays an important role in giving Mandarin its distinctive sound, and shows the following distribution: /(tʰ)x², (tʃ)x²; (tʃ)x, wʃn; (tʃ)x, wʃn/ [(tʃ)x(ʃ)], (tʃ)x, wʃn; (tʃ)x, wʃn/ e, er, en, wen/Cun, eng, weng. The last one has the variant [wɔn] (which, in case, could be indicated as /wɔn/). A
few examples: ‘kh[ǐ]+kh[ǐ]’ /’kh[ǐ]+kh[ǐ]’, ‘zhù’ /’zhù’/ Zhōu Énlái, ‘r̃ṽũh̃ẽ’ /’r̃ṽũh̃ẽ’/ éräié, ‘ˈj̃am.r̃p̃t̃ñ’ /’j̃am.r̃p̃t̃ñ’/ gēnbēn, ‘sw̃ñw̃ñ’ /’sw̃ñw̃ñ’/ Sun Wén, ‘f̃ñx̃ñ’ /’f̃ñx̃ñ’/ fēngshēng, ‘fĩw̃ñ’; ‘s̃ñ’/’fĩw̃ñ’/ fiwēng.

1.1.7. Finally, /ˈr̃o/ [ˈr̃o(ɔ)] ə, which can occur in the interjections /ˈo/, ə/ [ə(ɔ)], ə/ or be preceded by /w/ in the sequences /mwa/, pwa, phwa, (C)wa/ [̃m̃a(ɔ)], ̃b̃a(ɔ), ̃p̃a(ɔ), ̃a(ɔ), (C)wa(ɔ)], mo, bo, po, fo, wo/Cuo. A few examples: ‘mwa+mwa/ mōmò, ̃b̃a+b̃a/ bōpō, [lwɔ.twɔ(ɔ)] /lwɔ.tlwɔ/ luòtuò, [fɔwɔ+twɔ(ɔ)] /twɔ+twɔ/ zuò cuò, ̃hāñjɔ /̃lāñjɔ/ hāngyō, [ɔ(ɔ)] /ɔ/ ə!

As far as (i)ong is concerned, a traditional approach favours ‘/jʊm/’, and we consider it as perfectly legitimate (along with ‘/jʊm/’, for certain secondary taxophones). Nonetheless, we have observed that [ɵn] and [jʊn] are by far the most typical realizations, and the phonemic status bestowed on /ə/ suggests us to switch to the more realistic /jʊm/. Two examples will suffice: ‘jʊŋ+koŋ’ /koŋ+koŋ/ gōnggōng, ‘tʃjʊŋ+tʃjʊŋ’ /tʃjʊŋ+tʃjʊŋ/ qióngjíōng. 6

Phonemic conciseness vs realism: the interphonemic choice

1.2. We are perfectly aware that from an intraphonemic point of view, ie just within the Chinese language, [ɤ, ə, ə, ɔ, ɔ] and [e] could –or should– be considered as taxophones of the same phoneme. Besides, in contact with labials (mo, bo, po, fo, ou, wo/Cuo), [o] could be considered as a variant of /ɤ/ too, as many Chinese phonologists do (‘off’/wɤ/ or ‘/wɔ/’). In that case, a more ‘Solomonic’ symbol would be the ubiquitous ‘/ə/’, or –even better– the same ‘/ɔ/’ that we have employed in our simplified Pronuncia cinese per italiani (‘Mandarin Pronunciation for Italians’) as a practical passe-partout for the various occurrences of /ɤ/: [ɔ] (while keeping [e, ə] apart, for the reasons given above).

We are confident in our interphonemic choice to list /e, ə/ among the vowel phonemes of Mandarin Chinese. Anyway, more restricted (certainly, more abstract, because intraphonemic) inventories are possible as well, and –in fact– frequently adopted by many authors. Our approach is different and has been duly motivated, yet our readers might find it interesting to try out and compare various solutions, to evaluate the pros and cons of each one.

8
1.2.1. Thus, for the sake of completeness and *par condicio*, we present here one of such vowel inventories (coherent, nonetheless, with the same rigorous principles applied to our analysis throughout this book): /i, y; ə; a; w, u/.

Accordingly, our /je, ɋe, ei; wʊ, (j)ʊu, (j)ʊʊ/ should appear as ‘/jə, ɋə, əi; wə, (j)ʊə, (j)wʊʊ/’ (or ‘(j)un’, for the last one, as previously noted). In other words, our /e, ə/ would be treated as taxophon(ems) of the ubiquitous ‘/ə/’, when in contact with /i, j/ or /u, w/; inevitably, /æ, aʊ/ should revert to their traditional renderings ‘/ai, au/’ (though detached from phonetic reality as they can be).

Up to this point, one may feel content with this, and put up with the double or triple mental passages needed to switch from such a highly sketchy phonematography to something loosely resembling actual pronunciation, not to mention a passably accurate phonography. But some problems arise when trying to render å! and o!, two very common interjections. In our analysis, we do not have to think twice: /e, ə/ (usually: ɇ, ɈE, /æ/ [ɇ, Ɉ; ɇ, Ɉ, Ɉ], but Chinese dictionaries permit any toneme for both). The point is that interjections and onomatopoeia are written down in Chinese characters too, as any other lexeme, which may lead to think that they can be lexicalized and phonemicized accordingly. This is debatable, since ‘unorthodox’ phones appear sporadically in any language, precisely in interjections and onomatopoeia, but are not generally included in the phonemic inventory.

1.2.2. A glaring example: in Italian, laughter is represented by *ah ah ah!* (or something like that); the basic pronunciation should be something like [ʷhɑːhɑːhɑː], but we can actually hear anything ranging between [h, ɦ, h, ɬ, x, ɭ; ɑ, ɒ, ɑ, ə, ʊ, ɔ, ʌ, ʊ], just to list a few. None of these belong to neutral pronunciation, of course, except [a], ie the phoneme /a/. Now, how should we treat this? In fact, /h/ is a xenophoneme in Italian; as such, it is frequently omitted, and often it has to be: an English loan like *hi-fi* is currently pronounced [aɪˈfəi], then, while [haɪˈfəɪ] would be regarded as pedantic or affected. But *ah ah ah!* is purely Italian; thus, the only solution to represent it is accepting [h] as an incidental phoneme, /ð/ – to say, a phone that deserves to be exceptionally phonemicized, without actually being a fully functional component of Italian phonology. 7
1.2.3. If we want to stick to the ‘/i, y; ə, a; u, u/’ scheme rigidly, there is little choice but retaining /e, œ/ as *incidental phonemes* in Chinese, like /u/ (which, more exactly, is rather a *xenophoneme*). To make it even more captious, one may go as far as treating [e, œ] as *incidental monophthongizations* of ‘/əi, ai/’ and ‘/œu/’ respectively, still failing to transcribe them unequivocally. We presume that our readers have done the maths and drawn their own conclusions: a ‘skeletal’ phonology may appear logical and easy to remember, but it is not necessarily functional and easy to use.

We prefer, instead, to regard ‘/Ə/’ as an *archi-phoneme*, and to split it into three different *semi-phonemes*: /e, x, ə/. It has to be underlined that these are not the same as *incidental phonemes*: they are true phonemes, instead, but simply not ‘crucial’ ones. In fact, native speakers may perceive no difference between [œu] and [xu], for instance, or [æə, əə] and [AI], as well as between [un, un] and [on, on]; but we insist —once more— that phonology has to be internally coherent, but also as useful and explicit as possible.

That is why we have departed from the affIPA tradition of employing only ‘primary’ symbols at the phonemic level; and, according to the most adamant traditionalists, not merely ‘primary’, but as near to their graphemic counterparts as possible, too! Consequently, we should have even used ‘/e/’ instead of the nearly-passable ‘/ə/’, since there is no opposition between [e, e, e, ə, x] &c!

Fortunately, the latter practice has been dismissed, but the former still rules the land: as a consequence, most English dictionaries continue to employ ‘/e/’ to represent modern British [e], for which ‘/e/’ would be far better (but it is finally —and happily— gaining popularity among scholars). 8

1.2.4. In some analyses of Chinese, we find sober compromises like ‘/e, x, o/’ or, better, ‘/e, x, ə/’, but also masterpieces of schizophrenia, such as ‘/e, x, ə/’ or ‘/e, x, o/’ — and even ‘/e, ə, x, o/’, mixing phones and phonemes! This is highly confusing for foreign learners, who often rely on these poor examples of affIPA misuse to build their own Mandarin phon(em)ic inventory, when first attempting to learn *pǐnyīn* and try to pronounce the language.

For example, native speakers of French, Italian, Portuguese and German consider /e/ as actually [e], not certainly ‘anything in between [e] and [æ]’ (and /ɛ/ is [ɛ] as well, in those languages; neutral German even distinguishes between ‘/ɛ, ɛ,
'ε/ [ε; e; e']

What foreign learners really need is a phonemic scheme devised according to interphonemic principles: realism and multilingual applicability. In practical terms, this means choosing, for each phoneme, the symbol that encompasses all the taxophones (or may be able, at least, to suggest most of them), without discriminating ‘primary’ symbols against ‘secondary’ ones.

1.2.5. In canIPA, there is no such thing: a symbol is a symbol, and if it is best for its task, we employ it. So, even at first glance, Mandarin (but also English and Spanish) /e, œ/ suggest that the learner is dealing with something different from [e, o] and [e, ɔ], while at the same time implying a certain auditory and articulatory resemblance with both pairs. Besides, while our /e/ is able to cover successfully [e, ε, ι], ’/e/’ makes it necessary to revive the unrealistic ‘/jan, ʔan/’, to explain [jen, ʔan] yan, yuan.

Multilingual applicability is the other advantage of the interphonemic approach, in that it helps polyglots (or those who aim at being one) and phoneticians build a sort of personal, constantly growing ‘phon(em)ic mindset’. They can resort safely to it, when attempting to learn—or analyze—another language, or improve their own pronunciation of one they already know.

Just mastering [e, ε, ι] and [o, σ, ɔ] helps us avoid gross faux pas in most European languages and Japanese (and Chinese, as we have seen!), but also improves our understanding of different accents. For instance, English awe /σ/ is [σi] in modern Standard British, but [σo] in Mediatic British, and [ɔ] in most American accents; to an untrained ear accustomed to hear and speak only American English, [σo] may easily suggest owe /σo/ [σo] at first, but once the listeners are able to detect [o] in isolation, they would promptly recognize a mono-timbric diphthong like [σo], and distinguish it from [σo].

On the phonemic status of ɨ.

1.3. This last vowel phoneme requires a longer, but necessary, explanation, almost a chapter on its own. This is due to the fact that many Chinese scholars keep a conservative approach to Mandarin phonetics, being largely influenced by the assumptions of some pioneers in the field, the Swedish sinologist Karlgren among them. It is nearly given for granted, thus, that Mandarin has two retroflex—or apical—vowels. In our view, both belong to the same phoneme.
We strongly support ‘/u/’ as the phonemicization of what pinyin transcribes as i after z(\(h\)), c(\(h\)), s(\(h\)) and r. As explained below –cf Consonants– we adopt i instead, in order to mark, even at the graphemic level, and in parallel to /u/, that we are dealing with another fundamental component of Mandarin phonology, not with taxophones of /i/. Our choice falls upon /u/, as this appears to better represent all the taxophones emerged from long and accurate analyses of numerous recordings of neutral Mandarin voices (especially those of professionally trained speakers).

1.3.1. Many –if not all– Sinologists favour two symbols not enclosed in the IPA inventory: ‘ɨ, ɨ’; the former represents i in zi, ci, si — the latter, i in zhɨ, chɨ, shɨ, rɨ. Though called ‘vowels’, both are actually presented as ‘syllabic’ approximants (we prefer ‘intense’), totally assimilated to the point of articulation of the contoid that comes before them. They are often –but not systematically– considered as two distinct phonemes, though they clearly are in complementary distribution.

A very similar view is shared by the International Phonetic Association, which, in the 2003 issue of its Journal, proposed ‘ɪ’ as a unified expedient to represent both ‘ɨ, ɨ’. The official descriptions of these two phon(emes), then, are ‘(apico)laminal dento-­alveolar’ approximant and ‘apical postalveolar’ approximant, respectively. None can be entirely accepted.

1.3.2. Correctly, no groove is assigned to the first contoid, which nevertheless cannot be described as a full approximant (in our system, the proper symbol would be [ʒ]), nor is it totally assimilated to /ts(h), s/: in that case, we should perceive a continuous ‘buzz’, though lighter than that of [z].

More realistically, we have a vocoid, [ui], with negligible assimilation to the preceding segment: in fact, just a very short portion of the vocoid –its contiguous part to /ts(h), s/– exhibits a certain degree of co-articulation, but this is absolutely automatic, and need not be indicated in phonetic transcriptions, let alone in a phonemic one! We admit that an intense semi-approximant, [ɨ], may result auditorily quite similar to [ui], but the articulatory nature of this phone is primarily vocalic.

To sceptical traditionalist readers, we suggest a simple experiment, provided they can pronounce Chinese at least decently: articulate a long, slow and emphatic sɨ, paying attention to any transition from s to ɨ; if this is done correctly, the result should resemble something like [tszəzɿɨtuɹ]. 11
This being said and recognised, yet we have to determine what really matters – ie what is important and ‘typical’ – and what is simply the outcome of inevitable (but secondary) adjustments. From this point of view, it appears reasonable to skip even intense (or ‘syllabic’) [ʒ, z, ʒ, ʒ, ɕ, ɕ] and to keep only [u] as the real distinctive element of ɕi [su], which we therefore phonemicize as /su/ (with /ts(h)u/, too), discarding ‘i’ and similar expedients.

1.3.3. A slightly different criterion must be applied to /jual/, ie when /u/ follows the post-alveolar initials: zhı̊, chı̊, shı̊, rı̊. In this case, the automatic assimilation affects the entire articulation of /u/, and cannot be regarded as a merely transitional event: even an untrained ear can detect a continuous ‘buzz’, incomplete – or even absent – in /u/; auditorily, the latter has a very different timbre too, even if the mid-dorsum generally remains in the same ‘box’ as [u].

Much debated is the true nature of this /jual/. As we explain hereby, neutral pronunciation actually admits more than just one realization: it is surprising that, instead of simply recognizing this fact and providing all the relevant transcriptions, every single analysis that we have examined insists in presenting only one. Inevitably, very different phonic renderings have been proposed so far, depending on which sources the authors took into account, and how restrictive their concept of ‘neutrality’ is.

For example, in his Hànyǔ yǔyīn jiàochéng, Cáo Wén often refers to how ‘Peking speakers’ would pronounce something, as if the pǔtònghuà spoken in Peking were to be assumed as a model for neutral pronunciation. Undoubtedly, Pekingese speakers have the best chances to conquer a purely neutral pronunciation, considering that their local dialect served as the basis for Modern Chinese; yet, without proper training, anyone with an even slight Peking accent would be spotted immediately, sometimes by foreigners as well.

Nonetheless, the Peking accent is considered a ‘prestigious’ one, due to its connection with the capital city and its presence in the media, thus qualifying as what we call a ‘mediatic’ accent: not neutral, for sure, but accepted by most as nearly-neutral, and no doubt more widespread than a real neutral accent, which has to be learnt expressly.

With few exceptions, those handbooks where ‘ɻ’ is still employed describe it as something of the same nature as ‘ʃ’, apparently, but ‘narrower’... whatever this
means. Nevertheless, when one examines the relevant illustrations proposed by the
same handbooks, they usually show a curled-shaped tongue, with very little space
between its tip and the post-alveolar ridge. It is clear that, by ‘retroflex vowel’, the
authors actually mean an intense post-alveolar contoid.

As far as the exact nature of the contoid is concerned, opinions vary, but the
dominant position is that summarized by Cáo and others as ‘a less fricative [ʐ]’. 12
Provided ‘less fricative’ does not stand for a semi-constrictive phone (can[ʐ]), we can
safely infer that an intense post-alveolar approximant is actually meant (can[ʃ]), in ac-

cordance with the officially sanctioned description.

1.3.4. In fact, such a realization is very common in neutral Mandarin (as well as
in mediatic Pekingese) and, in general, in all northern China. It can be adopted
safely by foreigners learning the language, provided they refrain from that most
tiresome habit of curling their tongues so backwards, that they do not produce a
true post-alveolar contoid anymore, but rather an apico-palatal one: [ʈʂ].

It is a speech defect – or, at best, a form of hypercorrection– that even a certain
number of Chinese exhibit, especially when their native dialect contemplates no
[ʒ], and they are willing to ‘speak proper’: in doing so, they literally retroflect their
apex, instead of raising it gently behind the alveolar ridge, ‘just to be on the safer
side’. The resulting phones [ʈʂ, ʂ, ʐ] &c are normal in many Dravidian languages, for
example, but not in Mandarin, where they arouse an auditory impression of ‘jab-
bering’ instead, if used as initials.

Another peculiarity to be avoided is splitting [ʈʂ] into a sequence like [ʈʂɻ] (or
worse, [ʈɻɭ], even elongated in [ʈɻɭɭ], which might be misunderstood as /ʂɻ/). This
trait is pretty common among native speakers of ‘rhotic’ English accents, typically
those from the United States and Canada, in particular when they have just started
learning Mandarin. Not rarely, instead of [ʈʂ], these speakers would actually use any
one of their most familiar r phones, namely [ʂ, ʂ, ʐ, ɻ] &c (with their ‘intense’ coun-
terparts, [ʂ, ʂ, ʐ, ɻ] &c, for /ʃ/). These contoids act quite well as surrogates, but pro-
duce a fairly recognizable accent: those who aim at a more convincing pronunciation,
then, should choose [ʈʂ]. 13

1.3.5. The intense post-alveolar approximant, though, is not the preferred choice
of professionally trained speakers. Instead, we have noticed a consistent use of a
somewhat ‘milder’ phone, precisely [u], which—for the sake of conciseness—we shall describe here as a laterally contracted [u]. Since we have detected an entire set of such latero-vocoids as natural realizations of the so-called érhuà phenomenon, we suggest our readers to refer to the relevant §2, where we describe all the features involved in this—quite complex—co-articulation process.

Here, suffice to say that the presumable reason why professional voices favour [u] is both of an articulatory and euphonic nature. By comparing the orograms of [u] and [u], it is quite clear that the latter phone requires little—if any—movement of the apex, yet the auditory result is not very different from that of [u]. Therefore, in connected speech, the apex may pass from one articulation to another with relative ease, especially when the front half of the tongue is crucial (as in /i, y, j, q, t(h), ts(h), s, l, n/ &c).

Subjective and questionable as it can be, euphony certainly does play a role, even if [u] was not selected on purpose, nor was it ‘genetically engineered’ to improve the Chinese phonetic inventory. It is a natural taxophone, instead, and it has always been there: but those for whom a beautiful and mellow voice is the essence of a job well done—ie dubbers, actors, broadcasters, singers, and the like—know perfectly what sounds better… and instinctively adopt it.

It is worth noticing that using [u] suggests no far-fetched ‘stage pronunciation’ at all, like—for example—[r, r] in old-fashioned French and German theatrical delivery (even though an alveolar r did not belong to the neutral pronunciation of those languages anymore). For this reason, we argue that foreign speakers should at least attempt to learn and use [u] instead of [u] (or alternating between them).

There is a certain timbric difference between [u] and [u], though mostly as a side effect of the co-articulation phenomena occurring in the latter (lateral contraction, in the first place, primarily responsible for the characteristic ‘buzz’ typical of any [u]; not so much the post-alveolarity, which plays just a marginal role). Apart from this, [u] and [u] share almost the same mid-dorsum position.

1.3.6. Nonetheless, we have noted a more advanced taxophone, not very common, but adopted by professional voices, as well: some examples of it can be heard at the beginning of the Chinese dubbing of Hiroshima mon amour (Alain Resnais, 1959), which indeed sounds somewhat awkward to native listeners. Scattered occurrences of the same taxophone are found in various sound files attached to the
late 70’s *Practical Chinese Reader*; a few present-day tv anchormen use it occasionally too. A separate symbol is not needed, but, for truly fronted realizations, something like ‘[i]’ might be useful. In fact, kinesthesia reveals that it is possible to articulate a fully functional [Y] as front as in the area of [i], but no fronter than that, otherwise it would be very hard to contract both sides of the tongue and, at the same time, raise the apex.

**Diphthongs**

1.4. Modern Mandarin phonology permits only four diphthongs (fig 1.2): /ei, ae, ao, ou/ [ei, æː, əo, ʊə]. The former two may be preceded by /w/, the latter two by /j/, with slightly different vocalic traits: /wei, waɛ, jɑo, jʊu/ [wɛi, waɛ, jɑo, jʊu]. We insist that /i, u/ (and /y/) are always vowels, while /j, w/ (and /ŋ/) are always approximants, thus consonants, not miraculous ‘semi-something’; consequently, we cannot help disapproving of such renditions as ‘/ej, aj, aw, ow/’ or ‘/uei, uai, iau, iou/’ either (see our reference to /j, ŋ, w/, § 3.5.3).

![Diphthongs](image_url)

1.4.1. But, then, we find absurd inconsistencies: instead of changing *wei, you* into *uei, iou*, the official orthography prescribes *ui, iu*, with no real advantages for...
the learners. In all honesty, it is odd even at the typographical level: lacking the real nuclei˚
we are forced to put the relevant diacritics on the syllabic codae instead, i.e uì, iú, uî, iù, rather than the more friendly and correct ‘uêi, ióu, uêi, iîu’!

Besides, ui and iu easily lead foreign students to think that /wêi, jòu/ are actually to be reduced to ‘/wi, ju/’ in composite syllables. We can personally testify that even some Chinese teachers at the Beijing Language and Culture University are convinced of such a ‘reduction’, so strong does pinyin influence their understanding of Mandarin phonology. To prove their point, they would refer to certain secondary variants, where –nonetheless– we do not hear anything like ‘[wi, ju]’, but rather [wì, jòu] (and [xi, xu] for /ei, ou/, see fig 1.3).

Such variants belong to (nearly) neutral pronunciation as well: foreign listeners should be able to recognize them readily, but they are no way necessary for a good pronunciation, nor do they represent a true exception to the rule. Only if we really needed a different phonemicization (merely of academic interest, by the way), a compromise like ‘/wii, juu/’ would be somewhat worth consideration.


fig 1.3. Mandarin diphthongs (variants).

Vowel reductions

1.5. Mandarin vowels may undergo some timbric reductions %r shortenings, mainly when the relevant weakly-stressed or unstressed syllables occur in positions different from word-final ones, or in intonemes. As a matter of fact, in actual language, apart from in slow and accurate speech, some neutralizations occur elsewhere too. It is fundamental to know such reductions, at least to be able to understand spoken Chinese well. Foreigners need not use such reductions actively, al-
though an appropriate use of them would decidedly improve their pronunciation. In order to adequately show the behavior of different unstressed syllables (for vowel, consonant, and pitch reductions), sometimes we also had to use some rare words, rather than omitting something significant.

1.5.1. Fig 1.4 shows the most important reductions which may occur in fast casual speech, and in connection with the so-called qìngshēng, /$/, or ‘zero toneme’. This does not imply ‘absence of tone’, as many Chinese themselves think (and teach foreigners, too!), but simply a feeble tonal contour as that found in syllables which, being unstressed, lose their original (or theoretical) ‘full’ toneme.

The examples given are to be interpreted as in a sort of decreasing scale, starting from a pre-pausal position, in an intoneme (for the first variants), up to modifications produced within a sentence, in a pre-intoneme (for the other variants). For /VN, Vn/, the variants with [¥] have been included in the vocograms, but not in the examples; it would have been a sterile repetition, after all, as the vocaloid remains in its ‘box’, though nasalized. But it is important to note it, because the vocalic timbre remains the only discriminant, when [VN] and [Vn] merge into [¥].

Anyway, in order not to make our transcriptions more complex than necessary, we will not employ such reductions ordinarily. On one side, they do not occur automatically, nor universally; on the other, foreign learners—to whom our ‘normalized’ transcription is primarily aimed—can safely use normal vocaloids anywhere (but they should be able to recognize reduced forms, when native speakers use them).

Lastly, we have: [ɛ/ ɛ] in /wɑ(n); (w)ɛ/: [pə, ɻə; ɻ] bābā, [mə-ma; -mə] māma, [ɕ nIndex: ɥʉn] xìhuān, [tʰɛŋ, tʰɛŋ; -tɛŋ] tàitài; and [ɻ/ ɻo] in /jɪan; (j)ɔ]: [wɜŋ.ɻən: r] ɻɛn] wǎnshàng, [ɻʉ ɻʊ; -dɻ] zhīdǎo.

fig 1.5. Further reductions in syllables with ‘zero’ tones (in even quicker speech, not in an intonation; then, in an intonation, the realizations of fig 1.4 are also possible).

| /ɪn/ | /ɪn, ɣ] |
| /æ/ | [ʌ], [ɻ] |


Notes:

1 It is indispensable to read through and carefully consider CanePari, L. (2007) Natural Phonetics & Tönetics – articulatory, auditory, functional: G 1-7, to understand the basics of anindo and of the ‘natural’ method; and G 8, entirely devoted to vowels (vocalic phonemes) and vocoids (vocalic phones).

As in Cao, W. (2008) *Chinese Pronunciation Practice*, p. 12. In this booklet, the author seems to have resorted to such a kind of ‘improved’ pinyin, precisely to avoid IPA as much as possible, in teaching beginners the fundamentals of Chinese pronunciation. During a conversation with him, in 2008, the author frankly admitted that many foreign students were somewhat distracted by the ‘cohabitation’ of pinyin and phonetic symbols on the page. Therefore, given the tiny amount of time devoted to the methodic learning of pronunciation, and the huge amount of time and energy required to assimilate Chinese characters, he had to make a choice, and favoured pinyin over IPA.

In our view, this does not prove at all that phonetic transcriptions are superfluous, or even detrimental, but that they must be the logical outcome of a serious preparatory stage supported by a scientific approach, not a purely decorative embellishment, ‘just to make it look better’. In this sense, Cao’s choice was right, but not the idea that pronunciation can be really learnt through, without the aid of a well meditated and realistic transcription with phonetic symbols.

In Canepari, L. & Cerini, M. (2008) *Pronuncia cinese per italiani*, § 7.21, we proposed a similar romanization scheme (more for scientific interest than for practical use).

Native speakers of many English dialects, and German (just to mention two major languages) may find it easier to approach [ʌn] than [on, ʌn]. What really matters is that the speaker keeps a certain consistency.

Cf Cao, W. (2004) *Hanyu yuyin jiaocheng*, p. 84: the author adopts ‘[iuŋ]’ as the basic rendition, adding nonetheless that ‘in Peking, some speakers would also use different pronunciation, such as [yŋ] and [iuŋ]’, ie our can/ɔn, jun/ respectively.


In this time of majestic technological achievements, many instrumental (‘acoustic’) phoneticians underestimate the importance of kinesthesia and proprioceptive abilities. To those interested in ‘redeeming’ themselves, we may suggest Catford, J. C. (1988) *A Practical Introduction to Phonetics*.


We strongly suggest to read through Canepari, L. (2009) *English Pronunciation*, § 26 ‘The whole truth’ on English r’.
2. The \( r \) grammeme and the so-called \( \text{érhuà} \) phenomenon

2.1. As a result of an evolutive process, which lasted for centuries, the northwestern dialects (Pekingese and Standard Chinese itself among them), developed a series of monosyllabic unstressed suffixes, still in use today, like \( zi \) and \( li \)… \(^1\) Being very frequently used, these suffixes were progressively eroded, until they reduced to a mere \(-r\) coda; at first, it retained a certain syllabicity (probably as an intense contoid, \([Ç]\)), but its precise articulation is still under debate. Later, that coda merged with the ‘rhyme’ of the preceding syllable, modifying some of its components too, eg its vowel quality. Historically, such a process had been deeply rooted in the language well before the seventeenth century, when European missionaries described Chinese pronunciation in detail, for the first time. \(^2\)

This phenomenon is called \( \text{érhuà} \) in Chinese, literally ‘\( r \)-ization’; we could render it as ‘rhotacism’, which is partially adequate, or ‘syllabic retroflexion’, which is partially inadequate; it might be better to keep the original term (as with \( stød \), when describing Danish pronunciation), or to resort to something like ‘\( r \)-colouring’ if really needed. \( \text{Érhuà} \) not only influences the phonic outline of a lexeme; moreover, any \( r \) has its own semantic and even grammatical relevance: it is a true grammeme, then, with a phonic signifier and various realizations, as we will see right away.

2.1.1. A small number of verbs (including ‘attributive’ verbs, corresponding to our adjectives) become nouns by adding \(-r\):

- \( bāng \) /\( pān/ [\( ð\acute{\text{a}}n] \) ‘to help’ vs \( bāngr \) /\( pān\gender{\text{r}}\)/ [\( ð\acute{\text{a}}\text{ŋ}\)] ‘gang’;
- \( gān \) /\( kān/ [\( ð\grave{\text{a}}n] \) ‘to be dry’ vs \( gānr \) /\( kā\gender{\text{r}}\)/ [\( ð\grave{\text{a}}\text{ŋ}\)\( r\)] ‘dried food’;
- \( huà \) /\( h\grave{\text{a}}\)/ [\( ð\grave{\text{a}}\)\( ũ\)] ‘to depict’ vs \( huàr \) /\( h\grave{\text{a}}\gender{\text{r}}\)/ [\( ð\grave{\text{a}}\text{ŋ}\)\( ũ\)] ‘a painting’;
- \( huó \) /\( h\grave{\text{o}}\)/ [\( ð\grave{\text{o}}\)\( ō\)] ‘to live’ vs \( huór \) /\( h\grave{\text{o}}\gender{\text{r}}\)/ [\( ð\grave{\text{o}}\text{ŋ}\)\( ō\)] ‘activity’;
- \( shù \) /\( ð\grave{\text{u}}\)/ [\( ð\grave{\text{u}}\)] ‘to stand vertically’ vs \( shùr \) /\( ð\grave{\text{u}}\gender{\text{r}}\)/ [\( ð\grave{\text{u}}\text{ŋ}\)] ‘vertical brush stroke in Chinese calligraphy’.
2.1.2. Besides, a few noun minimal pairs are distinguished according to the same criterion:

tou /ˈθou/ [ˈθou] ‘head’ vs tōur /ˈθouə/ [ˈθouə] ‘chief’ (interestingly, chief derives from Latin caput ‘head’, through French chef);

xin /ˈxin/ [ˈxiŋ] ‘letter’ vs xìn/ /ˈɕin/ [ˈɕin] ‘information, news’;

bāntiān /ˈpaŋtʰjən/ [ˈpaŋtʰjən] ‘a long time’ vs bāntiānr /ˈpaŋtʰjəɾ/ [ˈpaŋtʰjəɾ] ‘half a day’;


2.1.3. Finally, an r-suffixed word may easily suggest a diminutive, reductive /ə/ affectionate acceptation:

bīnggûnr /ˈpiŋkʷɜɾ/ [ˈpiŋkʷɜɾ] ‘ice lolly’;

nánháir /ˈnaŋˈhəɾ/ [ˈnaŋˈhəɾ] ‘baby boy’;

mēnfèng /ˈmeŋfəŋ/ [ˈmeŋfəŋ] ‘a tight crack between a door and its frame, or between two doors’;

xiāo Zhāng /ˈɕjoˑz̚ang/ [ˈɕjoˑz̚ang] ‘young Zhāng’;

lǎo māo /ˈlɔˑmao/ [ˈlɔˑmao] ‘old dear kitten’;

àiqìng /ˈɛiˑq新时期/ [ˈɛiˑq新时期] ‘unimportant love’;

méi shír /ˈmeˑʃir/ [ˈmeˑʃir] ‘no problem’;

xiāoshuō /ˈɕjoˑsʰuײ/ [ˈɕjoˑsʰuײ] ‘novelette’.

2.2. With all its taxophones, érhuà provides one of the most prolific fields of research on neutral Chinese pronunciation: if a simple addition of [ʔ] to any syllabic rhyme –V(V)˚ Vn and Vng– might be sufficient for an ‘International’ kind of pronunciation to sound understandable and convincing, phonetic reality is—as usual—quite more complex.

Only certain phones are able to coexist with /ʔ/, which is primarily a postalveolar approximant, [ʔ]; any foreign speaker of Chinese may feel content with it, and employ it everywhere safely, including for érhuà. This is a wise and practical choice, totally loyal to the natural method. Nevertheless, audio files have revealed such a wide range of taxophonic oscillations, that it is no longer possible to passively accept what the ‘phono-sinological’ tradition has nearly passed off as a sacred dogma.
2.2.1. For decades, nonsensical and frankly embarrassing transcriptions like ‘[Vr]’ have been proposed in handbooks and essays, nothing being said about that curious mixture of pinyin and IPA... and sometimes the former prevails over the latter! Luckily, some ‘enlightened’ scholars at least resort to less shameful solutions, eg ‘/Vr/’ or ‘/[Vi]/’, but still inadequate both phonemically and phonetically. Such inadequacy derives both from the conceptual and typographical limits of the offIPA, and from the fact that ērhuà is still, too often, treated vaguely.

What we hear from professional voices (actors, dubbers and broadcasters) shows just that problematic sequences are systematically avoided, and adjusted to the few ones allowed. Furthermore, an actual [ʔ] rarely occurs in neutral ērhuà, since it would require a considerable effort if methodically applied to continuous speech. Instead of a vocoid plus a truly postalveolar approximant (with potential nasalization, as well), [V₄, Õ₄], in truly neutral and spontaneous pronunciation –neither emphatic nor mannered– /V₄/ sequences are more often realized as [Ṽ₄, Ŋ₄], ie a simple vocoid followed by a provelar (: midback) vocoid with lateral contraction.

2.3. As shown in fig 2.1, the apex is slightly raised behind the alveoli, certainly not ‘curled’, and that is why we disagree with the so-called ‘retroflex vowel [ʂ]’ frequently encountered in traditional descriptions of ērhuà. A certain degree of postalveolarity in any [ VARCHAR_1257] is indisputable, but the essential feature to be underlined is the lateral contraction, which inevitably causes the apex to move back and raise.

Concisely, any [ VARCHAR_1257] is a coarticulation of vocalic traits, ie a specific quality or ‘timbre’ (determined by the mid-dorsum position and the degree of lip rounding), inherent voiceness and potential nasalization; and of consonantal traits fairly similar to those of a slightly postalveolarized prevelar rounded semi-approximant with lateral contraction, [t]; the corresponding approximant is nothing but the famous ‘American /ɹ/’, [ɹ] (while Southern British /ɹ/ is, rather, a postalveolar rounded approximant, [ɹ], pace what most phoneticians still seem to think). 3

The very symbol ‘[i]’ suggested us the presser-foot-like diacritic in ‘/[ VARCHAR_1257]’. Yet, there are some differences worth noticing: while pǔtōnghuà [ VARCHAR_1257] are generally pro-velar, [i, ɨ] are pre-velar; the latter two are also intrinsically rounded, instead any [ VARCHAR_1257] owns its potential roundedness to that of its basic [ VARCHAR_1257]. In case a separate symbol for the ‘presser foot’ were needed, almost ‘distilling’ it from the actual coarticulation, that
would logically be a *provelar unrounded semi-approximant with lateral contraction*, [u]. Therefore: [V] + [u] → [V].

fig 2.1. The orograms of the laterally contracted vocoids.

![Orograms of laterally contracted vocoids](image)

fig 2.2. The vocograms of the laterally contracted vocoids for the suggested pronunciation.

![Vocograms of laterally contracted vocoids](image)

2.4. Before analyzing how these peculiar vocoids behave in *érhuà*, it is important to note that, even in /ζw/, we normally find a lateral articulation, [u], rather than an intense approximant like [z], as various *IPA* notations —eq *[i, i]— would suggest. Many sinologists use ‘[u]’, a nonstandard symbol whose *IPA* equivalent...
should be ‘[ ngàn]’, described as a ‘syllabic retroflex fricative’ or even ‘retroflex vowel’ (though written with a contoid glyph!). Undoubtedly, [ ngàn] belongs to neutral pronunciation as [ ụn] does, but it is associated with a tenser articulation and a certain emphatic component (if not with a marked Peking accent).

2.3. The vocograms of the laterally contracted vocoids with no simplification (but without secondary variants).

2.5. Starting with /a, a/, /a, o/ and /a/ interact without phonemic adjustments: /a, a, wa, o; a/. Chinese taxophonics does not allow /i, y/ as such, and adjusts them to /i, y/. Exceptionally, /i, y/ present oscillations that might suggest quite peculiar phonemicizations: [i, y] /i, y/; [y, y]/; but all the sound files that we have examined clearly show that these oscillations are always related to intentionally slow careful speech. In addition, /u, u/ change into /u, u/, in order to be better distinguished from [Cu, Cu] and especially from [Cu, Cu]. Herein we list the most frequent taxophones, including those with ạ (where ‘ạ’ suggests a less spontaneous pronunciation, often used by native speakers with a regional background and an intentional effort to approach a neutral pronunciation.)
2.5.1. Let us notice that ‘<…>’ means Suggested Pronunciation, that is a sober and sufficient simplification meant to employ as few taxophones as possible: in our opinion, those who are willing to acquire a native-like –yet manageable– kind of pronunciation are recommended to choose it. Our readers will not fail to notice that the suggested variant does not always coincide with the first variant shown on the list, and it needs not to: for example, [euy] is the most typical for /aɻ/, yet we ‘suggest’ [ɻ], because it sounds as much as natural to natives, and it employs what we consider to be a very versatile taxophone, [ɻ].

\[ar; yar, iar; war, uar /aɻ, j, w-/ [euy, j, w-] \& [ɻ] \& \$[ɻ] – τ[euy, j, w-]
\]
\[er \& \$r /ɻɻ/ (//uɻ//) [euy] \& [ɻɻ] \& \$[ɻɻ] – τ[euy]
\]
\[yir, ir; yür, iur /jɻ, ûɻ// [jɻɻ, û-] \& [ɻɻɻ, -] \& \$[jɻɻ, û-] – τ[jɻɻ, û-]
\]
\[wor, uor; (m-b-p-f-)or /wɔɻ// [wɔɻɻ, wɔς; ɻɔɻ, ɻɔς] \& \$[wɔɻɻ; ɻɔɻɻ] – τ[wɔς]
\]

2.5.2. Some sequences present such a wide range of taxophonic oscillations, that more than one phonemicization would be possible (and even advisable, for scientific coherence), though beginners need not use all the taxophones. For yer, ier; yüer, üer //jɻɻ, ûɻ//, the sound files suggest three groups of realizations: less frequently, \$[jɻɻ, ûɻɻ], which plausibly lead to ‘/jɻɻ, û-’; more often, [jɻɻ; ûɻɻ] /jɻɻ, û-/, totally assimilated to //iɻ, ûɻ//; yet the most typical realizations are [jɻɻ; ûɻɻ], which we still prefer to assign to /-er/ (in a more useful interphonemic analysis, according to the principle of Natural Phonetics, rather than to /-ər/, in a supposed ‘more economical’ –but more abstract, and less useful– intraphonemic analysis). So, we recommend the last taxophonic sequences shown, [jɻɻ; ûɻɻ] (or, at most, \$[jɻɻ, ûɻɻ], which risk to sound too precise, because of the use of the contoid [ɻ], instead of the more typical laterally contracted vocoids, cfr fig 2.1):

\[yer, ier \& yüer, üer /jɻɻ, ûɻ// [jɻɻ, û-] (\& \$[jɻɻ, û-]).
\]

2.5.3. Similarly, //uɻ// provides two choices: /wvɻ/, which we consider primary; and /uɻ/, certainly not indispensable, but useful to improve one’s understanding of spoken Chinese:
2.5.4. The second elements of the diphthongs /ei, ae/ are incompatible with the place of articulation of /a/; they are therefore reduced to a ‘zero’ element, leaving /ey, ae/; in theory, this causes homophony with er and ar, though the very few cases where this happens are substantially negligible. For //ei//, [(e)ae] is not totally uncommon, but it is better to adopt [(e)ae] as the ‘i’ norm, in order to keep an advantageous consistency with other /a/ sequences.

2.5.5. Regarding /ao, jaor/, both vowels are preserved, though a slight difference in /a/ can be observed: with /j/, the /a/ phoneme is prevented from being realized as far back as [a], which otherwise would be very natural, considering the inherent midback quality of [♀, ox]. A similar behaviour distinguishes /ou/ from /jou/, as clearly shown (cf fig 2.3), in spite of the high number of variants currently occurring for this diphthong: we would drag the reader’s attention to the high number of taxophonic oscillations registered for this diphthong: some of them would not justify a fully /jou/ phonemization; /jou/ surely suffices for an ‘International’ pronunciation, nonetheless it cannot properly represent secondary realizations like [ou, wou] (if really needed, ‘/jou’ might work well as a practical shorthand for both /jou, jou/).

2.6. In -nr, /n/ disappears, too. Consequently, gër ‘song’ and gënr ‘root’ are pronounced alike, /kvr/ [kr]; as well as bër ‘handle’ and bënr ‘obstacle, petal’, /pr/ [pr]. It is important to note that pînyîn is certainly not the actual phonetic structure of Chinese (as many are lead to think), but just an official romanization, which does not exactly correspond to the phonemes of the language. As a matter of fact, pînyîn might be more similar to a hypothetical, more abstract, phonological level, more like our possible // symbolization. For instance, we might have the sequences //in, yn, wvn, for inr, ünr, enr; but /jv, pvs, wvn/ are certainly better.

(w)ur /wn/ [wn] & t[wn] − i[wn] (& /um/ [um]).
As for /jen, ãen/, it must be remembered that /jen, ãen/ are simply more realistic phonemicizations for [jen, ãen] than ‘/jan, ãan/’; when /n/ disappears, the syllabic nucleus resumes its original theoretical vowel quality: /ja, ça/ (//jan, ãan//), with the same values as described above for /a/, e.g. [çãr ‘wɔpɔ₁] /’çen ‘ventions/ yuánquānr.

2.6.1. Also -ng /ŋ/ is dropped before /ŋ/, but it does not vanish totally; instead, it nasalizes the preceding vowel and the following /ŋ/, with a certain number of taxophones. It must be considered that [ŋ] are quite complex and unstable co-articulations: beside a slight raising of the apex and a lateral contraction, they involve the lowering of the velum, too, while having to keep the timbre consistent with that of the original [V]. In short, the actual realizations may vary: we have chosen the most frequent ones, including those with [ŋ].

The neutralizations and readjustments explained for ordinary /V(n)ŋ/ largely apply to /Yŋ/, as well: ingr /jvŋ/ (with a rare [iŋ] realizatiion, too, //Cŋ//); ongr can be explained either as a variant of wengr, thus /wŋŋ/ (with a fully audible [w] or, better, a milder [o]), or as an autonomous ‘rhyme’, /σŋŋ/. This requires a concise explanation: by and large, realizations referable to /wŋŋ/ instead of /σŋŋ/ are fairly common and certainly a good choice for foreign speakers; but with the labial phonemes, /m, p, f, v/, the glides [w, o] merge with the preceding /C/, producing a [C] articulated with rounded lips:

yingr, ingr /jvŋ/ [jʃʃ, jʃ] (& t[iʃ, jʃ], //Cŋ//) – j[ʃʃ];
angr; yang, iangr; wanger, uangr /anŋ/, j, w/- [aʃʃ, j, w-] & t[ɔʃ, j, w-]
engr /xŋ/, [ʃʃ] (& t[ʃʃ])
wengr, ongr /wsŋ/ [wɔʃʃ, wɔʃ, wɔʃʃ] (& t[wɔʃ, wɔʃ]) – w[ɔʃʃ]
ongr /σŋ/ [oɔ, ɔʃ] (& t[ɔʃ]) – o[ɔʃ]
yongr, iongr /wɔŋ/ [pɔʃ, pɔd] (& t[pɔʃ]) – p[ŋd].

2.7. We now present a few examples, transcribed according to our Suggested Pronunciation, though our readers will not find it difficult to prepare more transcriptions on their own, with all the –often capricious– taxophones of a purely ‘native’ pronunciation. Besides, once the few ‘suggested’ taxophones are automatized and the speaker’s fluency improves, thanks to intensive listening and speaking exercises (and to the influence of co-articulation, as well), the other taxophones very often come by themselves.
2. The r grammeme and the so-called érhuà phenomenon

A. /ær/: [ær] ar, air, anr; [jeə] iar, ianr; [ær] ianr; [wær] nar, uair, unanr:
/ær/ (/ə(a)(i)ə, anr/): háu huar [əваяwəř];
/jær/ (/jə(r), jən/): chuângliánr [τʂʰwə́nˈjər];
/ŋær/ (/ŋən/): yuângquârn [ŋə́ŋtɕʰpə́ŋ];
/wær/ (/wə(ə)́n, wən/): háowânr [ɕʰqə́wə́n].

B. /ɛr/: [ɛr] er, eir, en, iər; [ɛr] ur, ünr; [wɛr] ur, uir, unr:
/ɛr/ (/e(ə)r, ə(ə)́n, ən/): biêkər [bɊjŋkʰər];
/jɛr/ (/jə(n)/): shëjınr [ɕɊtɕʰjın];
/ŋɛr/ (/ŋən/): qūqùr [tʂʰjɊtɕʰjʊ́];
/wɛr/ (/wə(ə)n, wən/): kǒuweĩr [ʁkʰʊwə́n].

Again, yer, ier & yüer, üer might be given as /ʃɛr, ʂɛr/ too, yet we prefer to interpret them precisely as /ʃɛr, ʂɛr/, thus:
/ʃɛr/: shëjìr [ɕɊtɕʰjì];
/ʂɛr/: miǔjìr [ɱɊtɕʰjɪ].

c. /oʊr, jouɬ/: [oʊ] our, [joʊ] iur; /wɔɬ/: [wɔɬ, ɬu] (u)or; /jəʊɬ/: [aʊ] aor, [jaŋ] iaoɬ:
/ʊoɬ/: tǔdōur [tʰuɬɊʊ];
/ʃoɬ/: zhíliūr [ʃɊɬɊjʊ];
/wɔɬ/: shānpòr [ʂamɊpʊ];
/jəʊɬ/: nǎimáor [ŋɊɬˈmɕo];
/jaʊɬ/: màimiáor [mɊɬˈmɕa].

d. /aŋɬ/: [aŋ] anɡ; [jʊŋ] ianɡ; [wʊŋ] uanɡ:
/aŋɬ/: piâncâŋr [pʰjɛnɬɊkʰʊŋ];
/jaŋɬ/: dǒuǐiâŋr [dɊɬɊɬjǎj];
/waŋɬ/: tiânc̀uâŋr [tʰjɛnɬɊkʰwʊŋ].

e. /ʊŋɬ/: [ʊŋ] engr; [jʊŋ] ingr; [wʊŋ] uengr:
/ʊŋɬ/: bândèŋr [ʂɊɬɊtʃ̥];
/jʊŋɬ/: qî míngr [tɊɬɊmɕ];
/wʊŋɬ/: xiāo wēŋr [ɕɊɕɊwʊŋ].
2.8. Compared to the huge number of the Chinese-speaking population world-wide, it is just a tiny minority that can use érhuà natively and properly, being it a peculiarity of Peking and of the Peking area: many geolects, including some of the ‘north-eastern’ group, do not contemplate the érhuà phenomenon at all, or lack those phones necessary to realize it correctly. In other dialects, something comparable to érhuà has crystalized in a more conservative phase than that of Peking érhuà, or it has taken a different course.

Beside dialectal interference, the Chinese script, too, influences speech habits of those for whom érhuà is an (often erratically) acquired feature: the r grammeme, when explicitly registered, is represented by the character ㄜ ‘son’ added to the primary character. Because of that, many non-Pekingese speakers, especially when poorly educated, may indeed pronounce two distinct syllables, each one with its own segments; other speakers would instead articulate only one, but possibly with minor % non-standard neutralizations... not to be imitated, such as:

* [ʂɨ́ʐ] [ʂɤʂ]/shùr ‘number; vertical stroke in Chinese calligraphy’;
* [ɕi̯fuɭ]/sìʃur ‘wife’;
* [sja̯tʃʰɭouɭ]/ xiàο qiúr ‘small ball’;
* [tʃʰoʊɭ]/cuór ‘mistake’.

2.8.1. Proper érhuà enjoys prestige in speaking pǔtōnghuà, but its use is hardly coherent outside Peking; even there, ‘Pekingese’ speakers (treated as such for being born in the city, or for having acquired the local accent after a long stay there) are uncertain whether to use érhuà in formal contexts, considering it too colloquial. Not surprisingly, when bound to a formal register, most broadcasters and reporters, as well as government spokespersons, read aloud official texts as they are written: without the character ㄜ and –consequently– with no érhuà. Instead, a markedly colloquial register shows the opposite trend: érhuà abounds, sometimes overwhelmingly –if not anarchically– in vernacular Pekingese.
2.8.2. From a general perspective, the formal vs colloquial opposition is not the best criterion to explain érhuà, as many scholars still do. In fact, we are dealing with the result of a historical process inherent to the language, from which the new r derives: a true grammeme, not a mere embellishment. Regardless of any artificiality, érhuà is as much a part of the spoken language as the aforementioned suffixes -zi and -li are. To put it bluntly, certain words must have -r, others may have it, while others are not entitled to. The most authoritative reference work in the field is, by and large, the Contemporary Chinese Dictionary, where any lexeme requiring or allowing érhuà is clearly marked. Both native speakers and foreign learners could simply resort to it and look up the relevant word(s) they doubt about, but unfortunately few people actually do so. Therefore, it is not easy to determine a universally shared rule in cultivated speech, not even among ‘Pekingese’ speakers.

2.8.3. In an ‘International’ kind of pronunciation, slightly –but reasonably– simplified, when compared to the ‘normalized’ neutral one, just three sequences might suffice, namely [aɻ, ɻ, ɻ], to which any other /V(V)ɻ, Vnɻ/ shall adhere and conform. For /Vnɻ/, [ɔɻ, ɔɻ, ɔɻ] are fair enough to make oneself understood; but attention should be paid not to omit this nasalization, given that [aɻ, ɻ, ɻ] are not to be merged with [aɻ, ɻ, ɻ]. Let us examine a few examples: bänr [paɻ] ‘job; scar’ vs bängr [pɑɻ] ‘band, gang’; bèir [pɻ] totally; generation’ vs bêngr [pɔɻ] ‘separation; little coin’. After all, such nasalization is not difficult at all, provided the speakers are able to lower their velum, as illustrated in § 2.1. For yingr, we recommend [jɔɻ]; while, for yongr and iongr, the learner could usefully evaluate the alternatives, [jɔɻ, ɻ, ɻ, ɻ], though we recommend iongr [jɔɻ] and ongr [ɔɻ], for being the most practical ones: the important thing is to be coherent in the course of time. It goes without saying that an approximation to native vocoids produces a more authentic result and is, of course, highly recommendable (and appreciated).

A. Related to i[ɑɻ], we have: i[ɑɻ] ar, air, anr; i[jɑɻ] iar, ianr; i[ɲɑɻ] üanr; i[waɻ] uar, uair, uanr.

i[jɑɻ]: chuanglianr i[ʃɯanɻɪjɑɻ], ɻ[iʃɯanɻɪjɪɻ];
i[ɲɑɻ]: yuanquanr i[ɻɻəpɪɻɪɻ], ɻ[iɻɻəpɪɻɪɻ];
i[waɻ]: houwánr i[ʊəwɒɻəɻ], ɻ[iɻəwɒɻəɻ].

31
b. Related to \( i[\sigma] \), we have: \( i[\sigma] \) er, enr, ir; \( i[j\sigma] \) ir, ier, inr; \( i[\pi\sigma] \) ür, üer, ünr; \( i[w\sigma] \) ur, uir, unr.

\( i[\sigma] \): \( shj\text{inr} \), \( s[\sigma painfully] \); \( i[\pi\sigma] \): \( q\text{iquiri} \), \( s[\pi painfully] \); \( i[w\sigma] \): \( k\text{ounweir} \), \( s[k\text{houghway}] \).

c. Related to \( i[\sigma] \), we have: \( i[\sigma] \) our; \( i[j\sigma] \) iur; \( i[w\sigma] \) uor; \( i[\alpha\sigma] \) aor; \( i[j\alpha\sigma] \) iaor.

\( i[\sigma] \): \( t\text{udour} \), \( s[t\text{houtog}] \); \( i[j\sigma] \): \( zh\text{iliur} \), \( s[zh\text{illion}] \); \( i[\alpha\sigma] \): \( n\text{aimador} \), \( s[\text{nashman}] \); \( i[j\alpha\sigma] \): \( m\text{aimiaor} \), \( s[m\text{ashmaq}] \); \( i[w\sigma] \): \( sh\text{anpor} \), \( s[\text{shamphoq}] \).

d. Related to \( i[\á\sigma] \), we have: \( i[\á\sigma] \) angr; \( i[j\á\sigma] \) iangr; \( i[w\á\sigma] \) uangr.

\( i[\á\sigma] \): \( p\text{ianch\&n} \), \( s[p\text{hien\&n}] \); \( i[j\á\sigma] \): \( d\text{ong\&n} \), \( s[d\text{ong\&n}] \); \( i[w\á\sigma] \): \( ti\text{anch\&n} \), \( s[th\text{ien\&n}] \).

e. Related to \( i[\á\sigma] \), we have: \( i[\á\sigma] \) engr; \( i[j\á\sigma] \) ingr; \( i[u\á\sigma] \) uengr.

\( i[\á\sigma] \): \( b\text{\&nd\&n} \), \( s[\text{ban\&n}] \); \( i[j\á\sigma] \): \( q\text{in\&n} \), \( s[q\text{in\&n}] \); \( i[u\á\sigma] \): \( x\text{i\&w\&n} \), \( s[x\text{jowwax}] \).

f. Related to \( i[\á\sigma] \), we have: \( i[\á\sigma] \) angr; \( i[j\á\sigma] \) \& \( i[\pi\á\sigma] \) iangr.

\( i[\á\sigma] \): \( t\text{\&w\&n} \), \( s[\text{tho\&khoq}] \); \( i[j\á\sigma] \): \( k\text{\&qi\&n} \), \( s[k\text{hur\&khoq}] \).

Notes:


2 cf also Lu, Y.-Zh. (2001) Qingsheng be ěrhua, which, besides ěrhua, provides abundant examples of words with a 'zero' tone, too.
2. The r grammeme and the so-called érhuà phenomenon

Again, see CANEPARI, L. (2009) *English Pronunciation*, ¶ 26 ‘The “whole truth” on English r’, as it helps avoid any further misunderstanding (and misuse) of the relevant phonetic symbols, especially in consideration of the noticeable difference—in this regard—between offIPA and canIPA.

Besides, LIN, H-Y. (2007) *The Sounds of Chinese*, § 2.1.3.2, on /çV/; and § 3.4.5, where ‘rhonatized/retroflexed vowels’ are concisely described. Though the symbols adopted by the author are inevitably quite generic, yet it is quite clear that a certain degree of taxophonic variation has been duly noticed (which is remarkable, when compared with the bewildering superficiality found elsewhere).

3. Consonants and contoids

3.1. The consonantal system of Mandarin Chinese shows some interesting peculiarities, including some alternative possibilities of phonemic analysis. For example, the phonemic status of [tɕʰ, ʨ, ɕ] is debatable; in fact, pinyin has chosen their ‘official’ acknowledgement—as j, q, x—as to the detriment of vowels, especially /y/: this is rendered as ū, even if its phonemic values are indisputably /y, ū/. On the contrary, it would have been sufficient to systematically use ū, instead of fluctuating between yu, ju, qu, xu and nǚ, lǚ— all the more so because, with nǚ, lǚ, both an Umlaut and a tonal mark are to be used anyway. If no ū had to be used at all, the most reasonable—and typographically coherent—solution would have been keeping iu and iou as the post-consonantal counterparts of yu and you, without switching schizophrenically from one to another. Then, simply yu, jiu, qiu, xiu for [ɥy, ɻy, tɻɻy, ɕy], and you, jiou, qiou, xiou for [ɻjou, ɻɻjou, tɻɻjou, ɕjou].

Fig 3.1. Table of Mandarin contoids.

\[
\begin{array}{cccccccccccc}
& \text{bilabial} & \text{bilabial rounded} & \text{palatalized bilabial} & \text{labiodental} & \text{labiodental rounded} & \text{dental} & \text{alveolar} & \text{postalveolar} & \text{bilabialized prepalatal} & \text{palatal} & \text{palatal rounded} & \text{velar} & \text{velar rounded} & \text{uvular} & \text{laryngeal rounded} & \text{laryngeal} \\
N & \text{m} & [\text{m}] & [\text{m}] & [\text{n}] & [\text{n}] & [\text{n}] & [\text{n}] & [\text{n}] & [\text{n}] & [\text{n}] & [\text{n}] & [\text{n}] & [\text{n}] & [\text{n}] & [\text{n}] \\
K & \text{p}^{h}[\text{b}] & [\text{p}^{h}][\text{b}] & [\text{p}^{h}][\text{b}] & [\text{t}^{h}[\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] \\
K & \text{p}^{h}[\text{b}] & [\text{p}^{h}][\text{b}] & [\text{p}^{h}][\text{b}] & [\text{t}^{h}[\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] & [\text{t}^{h}][\text{d}] \\
S & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] & \text{s}[\text{z}] \\
J & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] & \text{j}[\text{ɻ}] \\
L & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] & \text{w}[\text{w}] \\
\end{array}
\]

Considering the vowel phonemic inventory that we have adopted, it appears more reasonable not to consider [tɕʰ, ʨ, ɕ] as phonemes, but simply as taxophones of /ts, tsh, s/ when followed by /i, y; j, ū/. All this, of course, is irrespective of both morphological and lexically-derived considerations, since they are not at all absolute. In fact, [tɕʰ, ʨ, ɕ] historically derive from two distinct series: /ts(h), s/ + /i, y;
j, ŋ/ and /k(h), ə/ + /i, y; j, ŋ/. The ancient ‘Phags-pa inscriptions, as well as comparisons with modern dialects, testify that both series still existed in Early Mandarin (around the XIII century AD), but then gradually merged. Nevertheless, we insist that phonology must be explicit, in order to be useful. The simple reason why we have chosen to assign [tʃʰ, ŋ, ç] to /tʃ, ts, s/ only, then, is that this series is nearer to the phonetic reality than /kh, k, ə/ is, both at the auditory and articulatory level.

**Nasals**

3.2. There are three nasal phonemes (fig. 3.2): /m, n, ng/ [m, ŋ, n, ŋ, n, ŋ; n], which pinyin correctly renders as m, n, ng. In syllable-final position, only /n, ng/ occur, and the latter always maintains its uvular articulation, even before any consonants. Instead, /n/ regularly assimilates, except in very slow and accurate pronunciation; assimilation is only partial for /nj, ŋj/, though: [ŋ, ŋ]; for /n[w]/, [ŋw] is a possibility. Besides, let us notice the reduced forms of wōmen and tāmen, which become monosyllabic in very fast speech, with /m[w]/: /wom, –tam/. The tendency of /n[w]/ to assimilation is so strong that we can find cases such as: [‘dej⁻xe⁺wx⁻mə] & [‘wx⁻m] /‘tsj⁻xe⁺wx⁰ma?/, Jièhūn le ma? Mandarin has /n/ /i, y; j, ŋ/, as said, although [n] is possible as well, and can be safely used by foreign learners.

On the other hand, before the dentals, /t, th, ts, tʃ, s/, the more specific symbol would be [n], but this is not at all necessary. On the contrary, it is useful to clearly mark [m] in /m[w]/ [mö(a)],. Syllable-final and word-final nasals always have an inaudible release, even before a pause: [n’, n’]. (This is quite normal for other oriental languages such as Cantonese, Thai, Cambodian, Vietnamese, Tagalog, Malay, and Indonesian – even for final voiceless stops, now lost in Mandarin.)

![Fig 3.2. Nasal contoids.](image)


**Stops and stop strictives**

3.3. Our transcriptions aim at being as clear and realistic as possible; at the same time, they are to be meant as a convenient normalization of numerous taxophones, some of which tend to be more typical and recurrent than others. It is precisely such a statistical recurrence criterion that our readers should consider, when comparing normalized transcriptions with those of some specific sentences uttered by given speakers. In this last case, all the peculiarities of a single speaker must be taken into account, and transcribed accordingly (especially at the hyperphonetic level, '[]', where necessary).

![fig 3.3. Stop and stop-strictive continoids.](image)

Traditionally, Chinese stops and stop-strictives (fig 3.3) are said to distinguish minimal pairs through ‘aspiration’: /ʃ/  /ɕ/. In fact, ‘aspirated’ consonants are, rather, /Ch/ sequences, where /h/ is therefore a true phone and phoneme, indeed. Since voicing is not phonemic, many Chinese phoneticians —simplistically— explain ‘/Ch/, /ɕ/’ exactly as ‘[Ch, ŝ]’ anywhere, as if there were no variation. Others, at least, introduce ‘ CONDITIONS’ for /C/; again, without further distinctions, regardless of their phonetic reality, which—as usual— is a little more complex than that. 5
3.3.1. In genuine pronunciation, we have observed—at least—three degrees of
variation, depending on whether a syllable has a primary stress, [§], a secondary
one, [§], or if it is unstressed, [ो]. For stops and stop-strictives, we have /Kh, KSh/
[צח, צח, צח], and /K, KS/ [צ, צ, צ]. This is absolutely automatic for native
speakers, who instinctively give priority to fully stressed syllables, while producing
much laxer articulations in unstressed ones.

Yet, for /Kh, KSh/, native speakers might also employ [צח] instead, with a semi-
constrictive (obviously weaker than its constrictive counterpart [ח]), or even [צח],
with the lenis laryngeal approximant more commonly found in /Kh, KSh/. That
depends on many factors, including enunciation speed and paraphonic superstruc-
tures, like emphatic or (on the contrary) muddy speech, not to mention those vocal
traits that good actors often adopt for theatrical purposes. For example, it is clear
that an ancient military hero giving orders to his soldiers will prefer a distinct, cut-
glass, somewhat stentorian delivery, while a Tâng lady-in-waiting is supposed to
talk in a most refined, mellifluous voice, coherent with her social status.

Regarding [ोצ(h)], we shall employ [ोצ] in general transcriptions, but it is impor-
tant to underline that a residual 'aspiration' may be present, here and there, espe-
cially in relatively slow speech: if not [ח], possibly something even lighter and
shorter, e.g a semi approximant, [צח]n].

3.3.2. Similar flexibility must be applied to [צ, צ, צ], respectively voiceless lenis,
partially voiced (actually, a mixed-phonation contoid) and voiced contoids. First of
all, it is hard to present Mandarin /K, KS/ as simply voiceless [צ]: in that case, na-
tive speakers of languages with a true /צ, צ/ opposition would notice no difference
between—say—their /t/ and what d is supposed to represent in pînyîn. Instead, to
Italian speakers, for instance, a sequence like dàodà médîd /tta9'ta "muntûtû/ would
almost certainly sound like a hypothetical Italian sequence /taodamu diiti/, or even
/daodamu dîdî/, depending—among other things—on the hearers’ own speech hab-
its, which always influence their perception of unfamiliar phones.

Since /t/ is a truly voiceless [t] in Italian, such a misinterpretation suggests that
the Chinese /t/ produces an auditory impression slightly 'milder' (or 'laxer') than
its Italian counterpart, even when fully stressed.

In our normalized scheme, in fact, we would have [t'q'da 'mu’dâ[­] -di]: never-
theless, as good recordings easily demonstrate. [צ] occurs too, as well as [צ, צ]. Os-
cillation, again, depends on many factors and on the characteristics of single speakers. Anyway, very few Chinese hearers perceive any difference, especially among stop phones (while voicing is somewhat easier to detect in stop-strictive phones, thanks to the /S/ component that makes voicing much more evident).

Even if /K, 'KS/ were occasionally articulated as purely voiceless [Ç], and not as voiceless lenis [Ç], a degree of distinctiveness between ‘aspiration’ and ‘non-aspiration’ would be guaranteed anyway. In fact, only /o Kh, ’o Ksh; ’K, ’KS/ would actually overlap, coinciding phonetically (but not functionally, from a phonemic point of view).

3.3.3. Native speakers of languages like English and German will find that acquiring Chinese taxophones is a relatively easy task for them: their mother tongues primarily oppose /Ç/, Ç/, but ‘aspiration’ and mixed phonation are very important components, at the phonetic level, and can be employed intentionally, once perceived adequately; for those who prefer an easier approach, we regard /Kh, Ksh/ [Çh] and /K, KS/ [Ç] as a versatile compromise. This is certainly ‘safer’ for foreign learners: first, it ensures a larger auditory distance between /Ch/ and /C/ than simply [Çh; Ç] (which is not wrong anyway), and it avoids unintentional ‘aspirations’ by those speakers whose native language—or accent—features this phenomenon: besides English and German, many Southern Italian regional accents, for instance (while neutral Italian has no ‘aspiration’, neither phonemic nor phonetic).

To acquire [Ç], one may learn to isolate and consciously articulate the latter conoid in Bob [b], dad [d] (alveolar, though, not dental as in Chinese, [d]), gag [g], before a pause or an unvoiced conoid. In fact, if we listen carefully, we shall notice a subtle difference between Bob comes [b ‘kh-÷ -z] and Bob goes on [b ‘q-; -z].

English has no unitary /ts/, yet the sequence ds as in dads [d] offers a rough approximation to [d]; for [d, d], one should start from English judge, first comparing j- [d] and -dge [-d], then gradually approximating to the correct points of articulation (fig 3.4 shows the orograms of [t, d; t, d], for a useful comparison between English [d, d] and Mandarin [d, d]).

fig 3.4. The orograms of [t, d; t, d], for a useful comparison between English and Chinese.
3. Consonants and contoids


3.3.5 As far as stop-strictives are concerned, therefore, we have /ts, tsh/ [ tì, dì, sì ]; /tsh, tsh, ôs] z, c; before /i, y, j, ɨ/, they are realized as [tì, dì, tì; tì, tì, ô] j, q. Finally, /tś, tsh/ [ tś, ô, ôd; tś, tś, ôs] zh, ch. As previously said, [ôs, ôc, ôt] must be intended as [ôs, ôc, ôt] ie with a possible residual ‘aspiration’. 


Constrictives

3.4. There are three constrictive phonemes, all phonemically voiceless: /f, s, ʂ/; they are also phonetically voiceless in syllables with a primary stress, [ $ ] , as well as in those with a secondary stress preceding another [ $ ], ie /($)$/, [ $ ] & [ $, $ ] (see 6.4): fànsì [ fànsì ], sàn sì ['sànsìv'], shìzhì [ 'shìzhìv'], fànsìnshù [ fàn 'sìnshùv], sànsì sì ['sànsìssìv'], shòushìzhì [ 'shòushìzhìv']. It is to be remembered that, in a typical [ $, $ ] tri syllable, the former [ $ ] is auditorily less prominent than the latter, due to the influence from the rhythmic structure.

3.4.1. When the most prominent syllable comes first, as in the /$o$/$($$/,$$/)/ [ $o$ ] & [ $o,$$/,$$/ ] sequences, the following less stressed syllables are articulated with much less energy, too. They are of course very weakly stressed per se; but, regardless of phonological considerations, in fact two degrees of ‘weakness’ exist: in relatively slower speech, we actually have /$o$/ [ $o$ ] instead of the [ $o$ ] found in normal conversation. Similarly, as enunciation speeds up considerably, it is easy to observe that all trisyllabic sequences tend to converge to one similar ratio, corresponding to a [ $o,$$/,$$/ ] pattern, in terms of phonatory energy. It is precisely such syllable-strength attenuation that produces a kind of ‘voicing gradation’ similar to that observed in stop and stop-strictive contoids: [ Cì, Çì, Gì]. Thus we have: /f/ [ f, ɨ, ə ] f, /s/ [ s, ə, ə ] s and [ Cì, ə, ə ] x, /ʂ/ [ ʂ, ə, ə ] sh.
A few examples: [ˈma-yan], ə-maven] máfan, [kʂʊ̃.ʐʊ], -zʊ] gàosu, [kwaŋ-ɻi], -ʐi] guānxī, [ɻjan-ʈʂu], ɻjan-ʂu] liángshì. This is obviously referred to fluid and moderately fast speech; while, in very slow % caref ul speech, the speaker may want to avoid any voicing at all, producing voiceless contoids anywhere. Many Chinese teachers adopt such a style of enunciation to teach beginners the fundamentals of Mandarin pronunciation.

3.4.2. Nevertheless, we do not support the [ɻɻ, ɻɻ, ɻɻ] pattern, for it might sound artificial to native speakers. Secondly, it would require the same amount of phonatory energy on every syllable, regardless of stress. But stress is quite relevant instead, in order to produce tones and intonation properly.

As speed increases, a simple two-way gradation spontaneously emerges, with for-tis [ɻɻ] and lenis [ɻɻ] as extremes; syllables with a secondary stress freely oscillate between the two (for, as usual, par phonics and the position in a rhythmic group do play a role): [ɻɻ], [ɻɻ, ɻɻ], [ɻɻ, ɻɻ]. Due to its relative ease and flexibility, such a pattern seems very suitable indeed for teaching purposes.

At an even higher speed, the taxophonic scenario becomes a bit more complicated, with the pronunciation of professional broadcasters and dubbers diverging from that of neutral (or nearly-neutral) native speakers in spontaneous conversation. The ’professional’ choice favours [ɻɻ], [ɻɻ], [ɻɻ, ɻɻ] (again, with smooth and flexible transitions between the edges), from which we have devised the normalized scheme ‘[ɻɻ, ɻɻ, ɻɻ]’ presented above.

Statistically, [v, z, ʐ, ʐ] are much more common in a preintoneme than in an intoneme. Besides, ‘laxer’ articulations abound in informal and colloquial environments: more [ɻɻ], then, together with ’reduced’ vocoids. However, we will keep the three-level gradation, in transcribing isolated words or short sentences, in order to highlight the possible phonetic ‘extremes’, at all times. In transcribing audio files (see 6), the symbols relevant to actual realizations will be employed, of course.
3.4.3. Even lenitions—both for voiced and voiceless phones—can be heard sometimes: [f, y; s, z; ã, ẑ; ñ, ẑ]; however, it is hard to define an explicit criterion for this last point, as it appears to be a side effect of many combined factors. Such phenomena interact with appropriate paraphonic features, to convey a sense of straightforwardness or familiarity; in more formal delivery, however, the same style could be regarded as sloppy and, if excessive, even regional (e.g., too ‘Pekingese’). Foreign speakers should therefore pay attention not to make a bad imitation of native speakers, just for the sake of sounding ‘cool’, or simply more fluent.

3.4.4. Though not precisely a constrictive, /h/ must be included here as well, since we have chosen to describe it—primarily—as a uvular semi-constrictive, [ã]. Our choice has a decisive advantage, hidden behind this unitary symbol, free from annoying diacritics: as a matter of fact, it is extremely difficult, if not impossible, to intentionally produce such a phone in current speech, with the absolute certainty of not articulating a full constrictive, [χ], or an approximant instead, [œ]. In syllables with a primary stress, native speakers themselves oscillate airily, probably more than with any other consonant. Thus, [ã] appears to be the ideal ‘normalized’ choice, capable of suggesting both the main taxophones, at the same time; secondly, by attempting to articulate [ã], foreign learners will happen to automatically reproduce the same—very natural—taxophonic variation observed in neutral pronunciation. 4

Statistically, [œ] is most common in syllables with a secondary stress, while we frequently find a velar approximant, [ŋ], in unstressed ones (be noted: a change in place—not manner—of articulation). Then, for héhào /ˈxɤˌækɞ/ and hâohê /ˈækɞˌxɤ/, we actually have [ˈxɤˌækɤq, ˈxɤqˌxɤ]; while, for huǒhù /ˈhɤˌækʌu/ and còuhe, /ˈtʃouˌxɤ/, we have [ˈxɤqˌxɤu, ˈhɤu; tʃɤuˌxɤ, ˈhɤu]

In fact, any of the taxophones of /h/, [h, h], would work as well, at the phonemic level; but phonetically, only [h] is auditorily ‘strong’ enough to replace [ŋ], not so much [œ] (while, certainly, [ã] would not be fit). More examples: [ˈxɤɪsɨ] běshi, [ˈʂɪəˌxɤ] shíbê, [ˈxɤɨˈxɤcɤ] bōhù, [ˈpɤˌxɤ(ˈsɭ)], ˈhɤs bōhe, [ˈsɤˌxɤu, ˈpɤs; ˈhɤ] shìhou(ɹ), [ˈsɭnˌxɤ] shânghài, [ˈxɤpɤˈxɤ면서] huânghuà, [ˈtɕɭfəxɤʍa] Xîzânghuà, [ˈxɤtʃɤɭ] bê cú.
Approximants

3.5. In our analysis, Mandarin Chinese has five approximant phonemes (fig 3.6). The first one is postalveolar, [ɻ] [ɻ]: [ɻɛ], [ɻɛn] rén, [ɻɛn] rún, [ɻon] róng, [ɻon] rì, [ɻɑnɻɔnɻɔnɻɔnɻ] rénɡrán róuruán. The various occurrences of [ɻ] at the end of a syllable –ie the érhuà phenomenon– are treated in depth in § 2, since most speakers actually realize /rphasis]< which is not as sequences ending in [ɻ], but rather as peculiar vocoids featuring both lateral contraction and a slight post-alveolarization.

fig 3.6. Approximant and semi-approximant contoids (with [ɢ], for the theoretical /u/ [ʊ])

3.5.1. Besides [ɻ], the corresponding post-alveolar constrictive [ɹ] is an important taxophone of syllable-initial /ɻ/; in fact, there is a certain degree of free variation, with the possibility of an intermediate articulation too, ie the semi-constrictive [ɹ] (though there is no real need to employ a distinct symbol, in ordinary transcriptions). However, the constrictive is not the primary rendition of Chinese r, contrary to what nearly all phoneticians sinologists insist on presenting in their analyses, where only ‘/ɻ/’ and ‘[ɹ]’ are generally employed.

From a certain point of view, that is an understandable choice: for one thing, ‘/ɻ/’ looks like the logical counterpart of /ɹ/, in terms of phonemic opposition (which theoretical linguists and acoustic phoneticians consider to be more than enough); secondly, the IPA inventory lacks a unitary symbol for /ɻ/, and the only two ‘candidates’ worthy of consideration, /ɹ/, /ɹ/, are not entirely adequate… not to mention that there is little agreement, among IPA enthusiasts, on what ‘/ɹ/’ precisely stand for. As far as /ɹ/ is concerned, we employ it univocally for British English r, which is certainly post-alveolar, but also laterally contracted, and it has a
certain amount of lip rounding, thus it is different from [ʐ] — of British English [ʒaʊ] /təʊ/ row ‘dispute’ vs Chinese [tʂʰaʊ] /tʂao/ nào ‘coil’.

Nevertheless, another problem arises with 'ʐ': how to represent ērhuà phonemically? No doubt, '/kʂj/' would seem a bit exaggerated as a phonemicization for gér. The sources consulted so far show a variety of solutions, from '/-ʐ/' or '/-u/' (with the possible ambiguities mentioned above), to an embarrassing '/-r/'. Sometimes, the question is ignored altogether, and the ērhuà ends up being banally transcribed with a mixture of pīnyīn and a few scattered 拗音 symbols, not always with a clear distinction between phonemes and phones (even [ʐ]! It is self-evident that /ʐ/ has the great advantage of being adequate anywhere (just one symbol), and of suggesting the most common real pronunciation (with no diacritics).


3.5.3. Some might even argue that /j, ɥ, w/ are not truly necessary, and could be replaced by their ‘vocalic counterparts’ /i, y, u/, at least at the phonemic level. But that could hardly be accepted; however, instead of simply saying ‘no’, we prefer to offer our readers a chance to judge by themselves. According to such a minimalist approach, then, /tʃ(C)V(V), ɬ(C)VN/ should be rendered as */tʃ(C)V(V), ɬ(C)VN/’.

A few concrete examples will instantly underline what the basic fault is, in this kind of analysis: ‘/tʂʰiɛn/ qian, */syɛ/ xüè, */tʂuan/ zhuàng. Apparently, there is no major difference with respect to our /tʃiem, ʂɛ, ʂwan/; but, let us examine the phonetic reality and listen carefully: in normal speech, we would normally encounter [tʃɬɛn, ʂɬɛ, ʂɬan]. Surely, the most prominent segments are the vocalic nuclei, namely [i, ɤ, ɒ], where most of the phonatory energy insists and focuses on.
Nevertheless, should we rigorously accept something like ‘/tʃiɛn, sye, tʂuan/’, the phonetic outcome would inevitably be [tɕʰiɛn, ɕyɛ, tʂuan], with major prominence on [ɻ̃ ỹ ŋ u]. It could not be otherwise, because /ie, ye, ua/ are certainly diphthongs, as much as /ei, ae, ae, ou/ are. It is self-evident that such an absurdity could be avoided only by means of a (no less problematic) phonemicization as ‘/oʃiɛn, syɛ, oʂuán/’. But, even so, we would expect, as a primary rendition, something like [tɕʰiɛn, ɕyɛ, tʂuˈan]; in other words, *three disyllabic sequences containing hiatuses*, while all native speakers perceive and treat them as *purely monosyllabic* words, instead!

Of course, that similar sequences can occur in extremely slow % emphatic speech, often with an intentional *staccato* enunciation, does not prove at all that /j, ɥ, ʍ/ are superfluous; rather, the contrary is true, provided everyone accepts that phonology must represent what is typical and frequent, not just incidental.

**Laterals**

3.6. For the last manner (of articulation) necessary for Mandarin, we find just one phoneme, /l/ [ɻ] (which remains unchanged even before /i, y, j, ɥ/) (Front/ɑɛ] /ɑɛ/ lāi, [lɔn] /lɔn/ lóng, [ɕʃən] /ɕʃən/ lián, [lʏ] /lʏ/ lù. Native speakers of English should pay attention, moreover, not to pronounce Mandarin l as the so-called English ‘dark’ ɻ, found in preconsonantal or prepausal position: *velarized lateral*, [ɻ], or—less conspicuously— *velarized semi lateral*, [ʃ]. Though neutral English /l/ is [ɻ] too, when preceding a vowel, there are some English speakers—especially Americans and Scottish— who happen to produce slightly velarized phones, if a Chinese /ɻ/ occurs next to back vowels, or to the back approximant /ʍ/. It is a speech defect with no phonemic implications, but it is particularly annoying in terms of phonetic authenticity and acceptability, and must be avoided at all costs.

---

**fig 3.7.** Lateral contoids (with a useful comparison with phones occurring in English).

---

Marco Cerini — *The Pronunciation of Mandarin Chinese*
Casual speech

We present hereby some supplementary information, regarding casual speech, i.e. the kind of pronunciation normally employed in informal situations, where native speakers not only talk faster than usual, but also most frequently tend to ‘indulge themselves’ in relaxed articulations.

What has been said so far, with respect to vocalic and consonantal reduced forms in ordinary pronunciation, largely applies to casual speech too, of course (as well as the typical paragogic setting with raised larynx, (·), not only for women; other Chinese languages, instead, use a normal—or modal—laryngeal setting, or even one with a lowered larynx). In addition to that, non-native speakers should be able to recognize two less common types of reduction, affecting vowels and consonants in fast speech and in a pre-intoneme (ie not in an intoneme), for they can make Chinese less easy to understand. They are not at all indispensable to acquire a good pronunciation; however, should foreign learners want to apply them in an appropriate manner, naturalness would certainly be improved.

First, the partial devoicing of close vowels after /p(h), t(h), k(h); ts(h), ç(h); f, s, ç, h/, i.e. not after sonants, and with the fourth toneme, particularly its ‘long’ taxotone [Y], (see below): /i, y, u, u/ [i, ı, ɨ; y, وية; u, ۇ, ۇ]. Be noted that partially devoiced vowels do not feature any ‘creaky voice’, laryngealization, (%, as their fully voiced counterpart would, when associated with low taxotones.


Besides, in unstressed syllables with ‘zero’ toneme, again not in an intoneme, approximant realizations are possible for /ʊs/: [z] instead of [z], and [j] instead of [z]; for /oɞ/, [z] is seldom reduced to [z]. In very fast and uncontrolled pronunciation, stop-strictives may feature the same reductions: then, /ts(h)/ [z] also for [ts, ç], and [j] for [ç, ç]; /çi '(h)/ [ç], also for [ç, ç].

Notes:

1 In Lin, H-Y. (2007) The Sounds of Chinese, the author seems to share our view on j, q, x.

2 cf Pulleyblank, E. G. [1991] Lexicon of Reconstructed Pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin, p. 6-8.

3 For instance, ‘[צ]’ are used in Cao, w. (2000) Hanyu fayin yu jiuyin (bud, sadly, not in Hanyu yuyin jiaocheng and Chinese Pronunciation Practice), as well as in Bruce, J. P. (1930) Chinese, a pioneering Mandarin course by the Linguaphone Institute, and entirely transcribed in IPA (included a tonetic notation which even clearly marks the various taxotones!).

   Sadly, by using ‘[צ]’, the authors normally mean ‘unvoiced unaspirated contoids’, not our partially devoiced contoids. In other words, ‘[ח, ד, ג]’ &x should visually prevent foreign learners—especially native speakers of English—from realizing ‘[פ, ט, ק]’ as [ph, th, kh], due to the interference of common orthography. In fact, the diacritical circlet simply means devoicing, in the official system, with no further indication whether devoicing is partial or total; thus, ‘off[צ]’ may correspond to ‘can[צ]’, but also ‘can[צ]’ (and ‘can[צ]’, too). Consequently, ‘off[ח, ד, ג]’ &x can be legitimately—though very questionably—used, in place of ‘[פ, ט, ק]’.

   However, we cannot help regarding this expedient as a ‘not-so-clever trick’. Surely, it avoids one mistake, but it easily leads to make another one, and for the same visual factor mentioned above: who can misinterprets ‘[פ, ט, ק]’ as [ph, th, kh] could certainly misinterpret ‘[ח, ד, ג]’ as [b, d, g], too, which is simply not correct in [§, $].

4 Most handbooks on Chinese pronunciation simplistically resort to ‘[צ]’ (as they do with ‘[נ]’, instead of ‘can[נ]’), with no gradation or further taxophones. We made a different choice, but we admit that ‘[צ, נ]’ can be perfectly legitimate, if used by non-native speakers, for purely communicative purposes.

5 ‘off[צ]’ corresponds to a number of different ‘canIPA’ symbols, but could be a decent replacement for ‘can[צ]’, in simplified transcriptions. However, it is the questionable use of ‘off[צ]’ to render the American English r, whose articulation has little to do with a true [צ], that forces the ‘offIPA aficionados’ to resort to ‘צ’, nonetheless always implying something like ‘[צ]’, ie an approximant, or a semiconstrictive phone.

4.1. One of the most remarkable features of all Chinese languages is that they employ different tonemes, besides vowels and consonants, as phonemic features: a variation in pitch, then, is capable of distinguishing syllables that otherwise would be perfectly homophonic. It is a fascinating challenge for foreign learners, but also a typical source of endless frustration. That is why it is paramount to practice all tonemes—and their phonetic counterparts: the tones—with as much care and perseverance as possible.

Whether ancient Chinese was a tonal language since the beginning is still under debate, but it appears plausible that tonemes developed at a relatively early stage of its long history. However, what can be targeted as the ‘ancestral’ dialects spoken by Hán tribes some three-thousand years ago, very presumably, possessed quite a large syllabic inventory, and the mere combination of vowels and consonants would have been more than enough to avoid ambiguous homophonies.

4.1.1. Tonemes are very common in the East-Asian language area, to which Chinese belongs: from Tibetan to Thai, from Burmese to Vietnamese, we invariably find quite restrictive syllabic inventories enriched by the use of tonemes. Otherwise, the lexemic monosyllabism so typical of those languages would simply be impossible, as it is the case in Japanese: though the Japanese akusento has a tonal nature, that language has no tonemes, in the strict sense of the term. No surprise, then, that polysyllabic words abound in purely indigenous Japanese vocabulary: atarashii /atəɾəʃi/ [aɾəʃii] ‘new’, kokoro /koʊkoɾo/ [koɾokoro] ‘heart’ vs shin, the Sinic pronunciation of both the corresponding Chinese characters (as in shinbun /ʃiŋbuŋ/ [ɕin⁴bun⁴] ‘news’, and shinri /ʃinɿɾi/ [ɕin⁴ɿɾi] ‘mentality, psychology’).

4.1.2. As far as Chinese ‘dialects’ are concerned, some of them count as many as eight tonemes; others, like Shanghainese, are more concerned with the tonal contour of entire words, rather than that of single syllables, and the five traditional tonemes nowadays interact in a simple two-way opposition. Pǔtōnghuà and the Peking dialect lie somewhat in between, with only four marked tonemes, or shēngdiào
collectively called 四声 [sìshēng] ‘four voices’. They are indicated by their ordinal number –T-1, T-2, T-3, T-4– or can be named after four of the eight tonemes of traditional Chinese phonology (since these names do not necessarily reflect modern pitch contours, it is better to avoid them):

- yǐnpíngshēng [jīnˈphūnˌʂın], ‘lower level’, nowadays rather a steadily high and level toneme, in neutral Chinese;
- yángríngshēng [jānˈphūnˌʂın], ‘higher level’, nowadays a mid-to-high rising toneme (in its basic form);
- shāngshēng [ʂānˌʂın], ‘ascending’, nowadays nearly always a low and slightly descending toneme, only seldom with a rising coda;
- qùshēng [ʨùhʊŋˌʂın], ‘departing’, apparently coherent with the modern steep, abruptly high-to-low falling contour.

Numerous expedients have been devised to indicate tonemes in Latinized transcriptions; that adopted by pīnyīn marks with a diacritic the most prominent graphic vowel in each ‘final’ (which not always coincides with the phonic vowel):

- T-1, qīn /tʃin/ [ʨhɪɛn], and qiān /tʃhen/ [ʨhɛn];
- T-2, lóu /lou/ [lʊʊ], but liú /ljoʊ/ [ljoʊ] (instead of ‘lióu’);
- T-3, hēn /hён/ [hʊŋ], but gǔn /kwʊŋ/ [kwʊŋ] (instead of ‘guên’);
- T-4, huì /hweɪ/ [hwʊi] (instead of ‘huèi’).

4.1.3. The marked tonemes only occur in syllables with a primary stress, [§], or a secondary one, [$]. However, since syllables are rarely uttered in isolation, some of them may easily undergo a further accentual reduction and become destressed syllables, [□§]; their phonatory energy is not strong enough to articulate a full tonal contour, and what is left is a shorter tone –with reduced intensity and no peculiar movements– whose actual pitch depends on that of surrounding tones, especially the immediately preceding ones.

The same criterion can be applied to unstressed syllables, i.e., those which never take a primary stress, not even when considered in isolation. Actual pitches –again– are determined by the tonal environment; but, unlike in destressed syllables, they cannot be linked to any of the marked tonemes. Therefore, another tonal group must be considered: a true unmarked toneme. The Chinese call it qīngshēng [ʨhɪŋˌʂın] ‘light voice’, another popular term is ‘neutral tone’, as if it had no tonal identity at
all. We, on the other hand, agree with the simple definition ‘zero tone’, which—of course—has to be interpreted as ‘zero toneme’: \( t-0 \), underlining that it is none of the other four (and it is the least prominent, too). \(^3\)

Dealing with Chinese tonemes—and all the relevant taxotones—is decidedly hard, not just for foreigners whose mother tongue does not use pitch for lexical purposes, but also for many Chinese whose native dialect has a different tonal behaviour. Very often, the more similar this is to the neutral scheme, the harder it is for the speaker to avoid cross-dialectal interference. For example, in the small borough of Lónghúzhèn, near Zhèngzhōu, Hénán Province, the locals pronounce \( t-1 \) and \( t-2 \) in a very similar way to \( t-3 \) and \( t-4 \) of \textit{pǔtōnghuà} respectively (and the other way around): thus, they have the right tones, but these are not assigned to the right tonemes, and a sentence like \textit{tā xǐé Hànyǔ} ‘he/she is learning Mandarin’ becomes a puzzling \textit{*tâ xìè Hányǔ}.

However, controlling the pitch to use tones correctly is not an option: practicing tones ‘the hard way’ from the start is the only way—if not to master them—at least to attain a decent level of accuracy. It will not be easy, and a certain skill on the side of the learner is certainly required, but any effort is worth the price.

\textit{Taxotones and stress}

4.2. Many students learning Chinese often fail to realize that the four tonemes are not fixed and unchangeable entities; in an attempt to ‘sound proper’, then, they tend to produce ‘full’ pitch contours under any circumstances, just like in uttering isolated syllables. Too many teachers do not sufficiently insist on the importance of linking the proper (taxo)tones in a natural speech \textit{continuum}, provided their pupils are able to put ‘the right shēng (toneme) on the right zì (character)’. Surely, the speakers will manage to communicate unambiguously, but such a kind of Chinese will sound emphatic and unnatural.

Instead, each toneme features at least two contextual variants, or \textit{taxotones} (better than simply \textit{allotones}), which have to be mastered, and employed in connection with all the relevant vocalic % consonantal reductions, where possible (not to mention, of course, that grammar and vocabulary must be kept under control, as well): quite a challenge!
4.3. For each one of the marked tonemes, there is a primary ‘full’ taxotone (fig 4.1), to be used in [§] (and—of course—in emphatic enunciation): [], ['], [], [']; the last two are further limited to pre-pausal position (or, to be more precise, when—at the end of a rhythmic unit not semantically and structurally linked to the next one—an interruption in the tonal chain intervenes).

Consequently, in normal speech, the various tonemes are expressed by different taxotones capable of interlinking more smoothly, thanks to a number of continuous and reciprocal small adjustments (or sandhi, should we want to show-off by using that technical term borrowed from ancient Sanskrit linguists), which produce a rather complex and composite pitch contour. Taxotones (see fig 4.2) are definitely not an unnecessary nuisance: without them, it would be uneasy and tiresome for both the speaker and the listener.

In syllables with a secondary stress, the primary taxotones are somewhat reduced, both in duration and in intensity; we have, then, [§]: [], ['], [], [']; (not considering other taxotones modified by intonation). Since the ‘zero’ toneme is regularly associated with [-§], the actual pitch contour is so feeble that it can be represented by a single dot, and only three symbols are necessary: [], [], [] (see fig 4.8-9); but, in slow speech and in an intoneme, [–] may receive more phonatory energy and substantially merge with [§]: [], [], [].

To be entirely rigorous, a notation like ‘[]’ would suffice to express a secondary stress with medium pitch, but we prefer ‘[-]’ to keep consistency with [], []; besides, in non-tonal languages, [] may be completely unrelated to pitch. For example, in English, Mexico /mɛksˈkəʊ/ is actually [mɛksˌkcəʊ] (American and ‘international’ pronunciation; while neutral British English would have [-kəʊ]); but what is the actual pitch of the three syllables, when the actual intonational patterns are applied? Considering three British marked intonemes only, we would hear.
4.4. In order to fully understand the mechanism of stress in Chinese, we must be able to isolate and analyze its fundamental stress units, usually called *rhythmic groups*. These do not necessarily correspond to *words*, no matter which particular language is concerned, but to syllabic strings within actual utterances, with at least one prominent syllable – bearing most of the phonatory energy – and a number of less prominent ones; all the more so in Chinese, where even the very concept of ‘word’ is particularly vague (and words of more than two syllables have more than one primary stress, although, in these cases, the dominant stress is the last one).


A useful comparison between the basic tones and three their important taxotones is clearly shown in fig 4.2: attention should be payed to [\$] /\$ á/, a T-3 that closely resembles a T-2 (though not identical, as too many textbooks still describe it), and occurs only before another T-3; then, the so-called ‘half T-3’ is [\$] /\$ á/, when preceding some other marked tonemes. Finally, there is even a ‘half T-4’, [\$] /\$ á/, which only occurs before another T-4. Quite evidently, the pitch contours of these taxotones make tonal adjustment easier, by avoiding inconvenient sequences; as a matter of fact, [\$] and [\$], in particular, [\$] and [\$] are nothing but ‘condensed’ versions of their corresponding full forms.

4.5. It may be argued that secondary taxotones are not fundamental, and therefore not actually necessary to foreign learners. On the other hand, if one aims at pronouncing Chinese convincingly, not just at mimicking its sounds, it is clear that four tonal contours are simply not enough. Besides, even if taxotones were not used by non-native speakers, they should anyway be able to recognize those taxotones, when listening to the Chinese. Besides, other peculiarities found in complete utterances become far less cryptic if one starts from a seven-element inventory.
4.2. Tones and taxotones.

![Tone and Taxotone Diagrams]

4.3. Fundamental taxotones in dissyllabic sequences.

A better explanation of how the taxotones behave is provided by 4.3, as well as by the following examples, where secondary-stress reduced taxotones are also employed. It is clear that [.] → [.] (& []), allows for a smoother transition from /./ to /\/, or /\/, or /\: huôché [χwɔŋ-ɕɪhʊ], děngyì [dəŋyì-yy], kâosbì [kʰɔsɔbì]. Similarly, a sequence like /.$$/ would be unpractical, if realized as such, or even with the taxotone just seen, ie [.,$]; instead, [\] decidedly improves the enunciation: shōubiào [ʂɔu-ɲjitone] (again, with a secondary stress and a reduced duration: [ʂʊ] → [ʂʊ], instead of [ʂʊ]).

In /.$$/, there is a generally descending contour: of course, the theoretical [\$\$] would require an uncomfortable ‘zig-zag’ movement, should we actually try to
make our voice ‘climb up’ (from [.] to [\]) before falling down again. Instead, with [\], a very straightforward {[} is obtained: fàng jià [fan\Œja]. In short, seven tones and as many symbols are required for {[}: [\], [\], [\], [\], [\], [\], [\]; but only five, for {[}: [\], [\], [\], [\], [\] (because [\] & [\] → [\]; [\] & [\] → [\]); and three, for [\$]: [\], [\], [\].

4.6. Generally speaking, trisyllables composed by marked tonemes follow the [\$] pattern, which is substantially coherent with that of disyllables, as it can be analyzed as ‘[\$]+[\$]’. It is easy to observe that, unlike English, no syllable is entirely deprived of stress, especially if speaking at a normal speed and the word occurs in an intoneme: again, Mexico [\œks\koo] vs Mòxïgë [\œ\ï\k]\; moreover, both [\œ] and [\k]\ bear a primary stress, though the former is auditorily less prominent, due to the influence of rhythm.

Other examples: yöümògân [\j\ou\m\œ\k\] ‘sense of humor’, féngrèn\jì [\f\œ\\j\] ‘sewing machine’, huôchëzhàn [\w\œ\\œ\] ‘train station’, fùmù shuô [\f\œ\m\œ\] ‘(my) parents say’. It is interesting to notice that the primitive lexemes, if taken in isolation, would switch from [\$] to the [\$] pattern: yöumò [\j\ou\m\œ] ‘humor’, féngrèn [\f\œ\\] ‘to sew’, huôchë [\w\œ\\] ‘train’, fùmù [\f\œ\m\œ] ‘parents’.

4.6.1. Besides the norms explained hitherto, which occur inevitably and nearly irrespective of enunciation speed, there are other particular cases, involving [\], where a certain degree of fluency and speed is required for the tonal adjustment to take place: [\] → [\].

In fact, native speakers are rarely conscious of converting a t-2 or a t-3 into a t-1 (actually, the overlapping is taxotonic, not tonemic): for them, that is a perfectly natural outcome, in their speech habits; but, if asked to repeat the same sequence at a slower pace, paying attention to its taxotones, they invariably avoid [\] in favor of the original taxotones.

Foreign learners must pay attention to this last point, and are even prone to omit the ‘flattening’ presented below, if they do not pronounce the language fluently enough, or are not entirely proficient in linking words together: otherwise, the use of [\], though in the right place, might result puzzling and be judged (paradoxically!) as wrong.
First of all, let us examine §4.4, presenting the behaviour of /$$/, ie /$$/ followed by any other marked toneme: [$$] → [$$], eg xianrenzhang [xiɑnrɛnʐʰæŋ], dongnanfeng [tɔnɲan盼], sanmianjie [saɲɲjeŋʦɿi], xihongshi [ʨiˊusahaan]. As said, in slower speech, a slightly less spontaneous but ‘safier’ realization would resume [nan], [ɲɛŋ], [ʦɿ], [ɲɔn].

§4.4. Taxotones in /$$+/$$ trisyllabic sequences.

Besides, §4.5 presents the behaviour of /$$/ [$$] → [$$], eg Mei Lanniang [meiланʃIan], wanquan dong [wɑnɿtʃɥanɿ], hai mei lai [清算meilai], xinyangjian [ɕyɹjanɿ].

§4.5. Taxotones in /$$+/$$ trisyllabic sequences.

Finally, §4.6 describes what happens when the /$$/ sequence follows one of the first three marked tonemes: [$,,$] → [$$,$] → [$,$] (with [$] ≠ [$]), eg ta henshao [ʦʰaˊxɿŋ], hanshubiao [ɿɿʂuˊpɿʃɿʊ], wo ye you [woˋ], (be noted: [wo], not ‘[wɔ]’; depending on rhythm and emphasis, the alternative reading
[wó:je:jìː] is possible as well). Even if a rhythmic unit includes more than three syllables with marked tonemes, it would have a primary stress on the last syllable, anyway, and another on the first one; ie with a strong tendency to keep most of the phonatory energy on the ‘outer ends’, and to reduce it in the middle. This means that, in a four-syllable sequence, the usual distribution is [$$,$$], rather than [$$$], which would not sound very smooth; but with five syllables, there is no reason to subvert the [$$,$$$] pattern resulting from the simple juxtaposition of [$$,$$$] + [$$], or the spontaneous adjustment of [$$] + [$$,$$$] (lexical factors play their role too).

A practical example: taken separately, *pinyin* ‘to spell words out (phon[et]ically)’ and *zìmù* ‘alphabet’ are naturally [$$]: [pʰin̪ jìn], [dʒu:mû]. By combining them into *pinyin-zìmù* ‘phono-alphabetic script’, ‘romanization’, the theoretical (yet possible) rendering [pʰin̪ jìn dʒu:mû] normally changes into [pʰin̪ jìn dʒu:mû]. In English we have examples like *the seven-year itch* [ðə sævn̩ ɪtʃ] or *fifth generation computers* [fɪfθ dʒeɪnəˈteɪʃən kʌmpərətər], which combine their plain elements: *seven* [sævn̩], *year* [ɪtʃ], *itch* [ɪtʃ] and *fifth* [fɪfθ], *generation* [dʒeɪnəˈteɪʃən], *computer* [kʌmpərətər].

4.6. Taxotones in /$$/,$$/,$$/ trisyllabic sequences.

It has to be underlined, both from the learner’s and teacher’s point of view, that rhythmic groups are not detached from intonation and paraphonics; instead, they influence one another: therefore, and not surprisingly, it is frequent that more than one stress arrangement is possible for similar utterances. Let us observe Mao Zédōng *zhùxì* [mao dzu: tson dzu:qi] ‘chairman Mao Zedong’, with simple juxtaposition: [$$,$$] + [$$]. But, in the basic sentence *tā shì zhùxì* (he/she) is (the) chairperson’, we have [$$,$$$]: [tʰalzhuː dzu:qi]; in fact, [tʰalzhuː dzu:qi] (omitting sentence intonation symbols) would not be a mere variant –as in *pinyin-zìmù*— but would im-
ply that 'he/she is (the) chairperson', though with less assertiveness than \[\text{'th~æ~[\text{yu}¥\text{ci}]}\] ‘(he/she) definitely is (the) chairperson’, with an extra-strong, emphatic stress on \textit{sh}i, the copula.

On the other hand, one might say \textit{Li fu~zh~ù~xí} \[\text{sic~\text{yu}¥\text{ci}]}\] ‘(the) vice-chairperson Li’, where two contiguous syllables retain a primary stress, for they logically belong to different semantic entities: rhythmically, however, they can be regarded as part of one sequence (even if auditorily— with slightly different prominence).

4.7. A Chinese disyllable may feature a ‘zero’ toneme on its second syllable, which is therefore considered as \textit{intrinsically un-stressed}: /o$/ [o$]; in that case, the primary stress obviously falls on the first syllable. \textit{Pinyin} systematically annotates unstressed syllables by leaving them without tonal marks: \textit{ẉóde, nàli, ěrzī, rènshī, luòbo, zhàogū, méiguī, etc.} Such typographical convention is mandatory, and (good) teachers would reject such misspellings as *ěrzī, *rènshī, and *méiguī.

There are also some \textit{de-stressed} syllables, \textit{/$/} [$/] → [o$], ie those with a marked toneme, which—under specific circumstances— may be ‘weakened’ up to the point of acting, \textit{de facto}, as T-O syllables. Though such a distinction is far from being indispensable, from a practical point of view, nonetheless it proves useful to thoroughly understand the complex mechanisms of Mandarin prosody. The official orthography is somewhat inconstant, on this respect: de-stressing is sometimes explicitly annotated, more often it is not (pretty much like tonal adjustments, which are given as implicitly understood by the reader/speaker). A few examples, duly transcribed, are given below.

As a general rule, the actual pitch of a ‘weak’ syllable is determined by the tone of the syllable that comes before it; in fast connected speech, the pitch of the following syllable may exert a certain influence as well, but always on a secondary level. As previously said, three symbols are enough: [], []; but, phonetically, \textit{four} different pitches may be detected, as shown in \textit{fig 4.7}: \textit{jùzī} [\text{[ẉi~dùu]}] ‘chicken’, \textit{jízī} [\text{[ɕi~dùu]}] ‘collected works’, \textit{jùzī} [\text{[ɕi~dùu]}] ‘muntjac’, \textit{jízī} [\text{[ɕi~dùu]}] ‘bun’.

In slow speech and in an intoneme, as said in § 4.3, the last syllable may receive a secondary stress, [$$], so that it becomes slightly more prominent than it usually is: eg [\text{[ẉi~dùu]}], [\text{[ɕi~dùu]}], [\text{[ɕi~dùu]}], [\text{[ɕi~dùu]}]. The difference is noticeable, even
to an untrained ear, thus foreign learners should pay attention not to abuse of this possibility, out of the contexts in which it is allowed.

By referring to fig 4.7, again, there is clearly a mid pitch after both T-1 and T-2; more precisely, a lower mid pitch, after T-1, and a slightly raised mid pitch, after T-2: *bipo* [piː-pɔ], *piqi* [piː-tɕi], *piqi* [piː-tɕi]. After a T-3, which is automatically 'shortened' to [], we have a lowered high pitch; finally, after a T-4, the 'dot' lies just in the middle of the low band: *bijang* [pɔ̚-vən, -vən], *bilui* [pɔ̚-wəʂ, -wəʂ].

fig 4.7. Taxotones in /$+$$/.$/$ trisyllabic sequences.

\[
\begin{array}{cccc}
\text{f} & \text{of} & \text{[".] [".]} & \text{f} & \text{of} & \text{[".] [".]} & \text{f} & \text{of} & \text{[".] [".]} & \text{f} & \text{of} & \text{[".] [".]} \\
\end{array}
\]

What above mentioned, of course, works well with single rhythmic groups pronounced in isolation; but, when affected by intonation, such a kind of 'ideal' pronunciation is inevitably modified % compressed, and many nuances may become less peculiar, especially in current speech: even more reason, then, to 'get by' with just three punctiform symbols (the three-level dots we have already seen [···]).

4.8. When a marked toneme is followed by two unmarked ones, /$+$$/.$/$, the phonatory energy is redistributed as [+$+$/], akin to the general trend of reducing stress in the middle —as observed in /$+$$/.$/$— while keeping most of it on the edges. As fig 4.8 clearly explains, we have kāilai le [kʰaː-læ̚-lɪg(ʂ)], nālai le [nɑ̚-læ̚-lɪg(ʂ)]. zōulai le [tʂʊq-ɻæ̚-lɪg(ʂ)], sōnglai le [ʂʊq̚-ɻæ̚-lɪg(ʂ)], with /læ/ destressed to /lɛ/ (see below).

fig 4.8. Taxotones in /$+$$/.$/$ trisyllabic sequences.

\[
\begin{array}{cccc}
\text{f} & \text{of} & \text{[".] [".]} & \text{f} & \text{of} & \text{[".] [".]} & \text{f} & \text{of} & \text{[".] [".]} & \text{f} & \text{of} & \text{[".] [".]} \\
\end{array}
\]

Between two marked tonemes, a T-0 (fig 4.9) is practically realized as a 'linking taxotone', whose pitch is directly influenced by the following tone, too, as well as by the preceding one: chēlè fàn [tʂʰuɻ̚-lɪən], pā-bu-shàng [pʰabuɻ̚-ʂən], wōde shū
The examples presented here lead us to say a few words more about the behaviour of ‘destressed’ syllables, which we usefully consider to be distinct from purely unstressed ones, at least conceptually: *yínweǐ* ‘because’ is regularly [ˈjɪnˌwɤt], but the alternative pronunciation [ˈjɪnˌwʊi] is permitted (with a different spelling, too: *yínwei*). The authoritative *Contemporary Chinese Dictionary* signals these alternatives by putting a dot before a syllable with a marked toneme, as in ‘*yín·wěi*’.

**fig 4.9.** Contextual pitch of a ¼-o in some trisyllabic sequences.

![Contextual pitch of a ¼-o in some trisyllabic sequences.](image)

Resultative complements are most often destressed. Taking *kàn* [ˈkhɑŋ] ‘to see’, and *qǐ* [tʃʰi] ‘to start; to lift’, as main verbs, we can add an enclitic verb to them, like *jiàn* [ˈlɯjɛn] ‘to see’, or *lài* [ˈlɑa] ‘to come’; we obtain *kànjiàn* [ˈkhɑŋʃiʒjɛn. ˈdʃjɛn] ‘to see’ and *qǐlài* [tʃʰiˌlɑa, ˈlɑa] ‘to rise up, to commence’ (and other meanings). But, if we insert *dē* [tʃʰ] ‘to reach’ or *bǔ* [pʊ] ‘not’ between the verb and its resultative complement, a potential construction is obtained; *dē* and *bǔ* lose their stress, [d̚s], while the other verb resumes the original toneme: *kàndejiàn* [ˈkhɑŋtʃʰiˌʃjɛn], *qǐbulài* [tʃʰiˌbɿˈlɑa].

Let us also compare *shǒu + qǐlái + le* → *shǒuqǐlái le* [ˈʂʊtʃʰiˌlɑaˌlɤ(ɿ), ˈlɑaˌlɑa] ‘to have collected scattered things’: if *le* comes after the main verb, the sequence *shǒu + le + qǐlái* gives *shǒu le qǐlái* [ˈʂʊˌlɑaˌtʃʰɿˌlɑaˌtʃhɿˌlɑa]. A monosyllabic pronoun like *tǎ* ‘he, him; she, her; it’ might be inserted too: if fully stressed, we would have *shǒuqǐ tǎ lái le* [ˈʂʊtʃʰiˌtʃʰɿˌlɑaˌlɑa], or *ˌlɑaˌlɤ* [ˈʂʊtʃʰiˌtʰɿɿɿˌlɑaˌlɤ] (more than one arrangement is possible, depending on what the speaker wants to underline). *Tǎ* can be used as an enclitic too; but, since *pínyīn* does not permit such spellings as ‘*tǎ*’ (unlike *lái, qǐ, jǐan, ㄆ*), we are compelled to write *shǒuqǐ tǎ lái le* anyway, for /ˈʂʊntʃʰɿʔtɕʰaˌlɤɿɿɿˌlɑaɿɿɿ/ [ˈʂʊtʃʰiˌtʰɿɿɿˌlɑaˌlɤ].

**4.9.** Lastly, there are some peculiar uses to be learned ‘as they are’, concerning the negative adverbs *bǔ* and *mēi*, and the numerals *yì* ‘one’, *qǐ* ‘seven’, *bā* ‘eight’. Before a ¼-4, *bǔ* /pʊ/ changes into /pʊ/, producing a smoother combination: *wǒ bǔ*

qù ‘I don’t go’ /wɔIpʊtʃhʏ/ → /wɔIpʊtʃhʏ/ [wɔIpʊtʃhʏ] (≠ [wɔIpʊtʃhʏ], with loss of stress); for a moderately stronger assertion, like ‘I do not go’, the spontaneous pronunciation would rather be [wɔIpʊtʃhʏ]. However, if the speaker wants to emphasize that he or she has ‘absolutely no intention to go’, bù may retain its toneme, possibly with two distinct rhythmic units and a potential pause, eg /wɔIpʊ. tʃhʏ./ [wɔIpʊ. tʃhʏ.] (the intonation symbols will be explained in the next chapter). The official orthography permits the written variant bù, as in wǒ bù qù, but this expedient is not systematically used.

In x-not-x interrogative patterns, bù and méi are generally destressed (this could be indicated by their romanization, too: bu, mei): dōng bu dōng [tɔŋbu,tɔŋŋ]; chi̍h mei̍h [tʃhu̍hmei̍h,tʃhu̍h, -mei̍h].

The numeral yi ‘one’ keeps its original T-1 when the lexeme is used alone, or at the end of an utterance: shíyī [ʂu̍hjí] ‘eleven’, yī yī dē yī [jíjí: ‘çī-jí:] ‘one multiplied by one gives one’ (again, please, refer to the next chapter for an explanation of our intonation symbols). Before a T-4, yī becomes /jī/: yīgòng [jíkɔn̥] ‘altogether’, yī bān [jípɔn̥] ‘a half’. Before other tonemes, instead, yī changes into /jǐ/: yǐdiānr /jǐdjaɾ/ [jǐtʃɤŋk] ‘a bit’, yī tiān [jǐtʃen̥] ‘a day’, yī nián [jǐpʃen̥] ‘a year’. For teaching purposes, more realistic spellings are found, though they are not encountered in official orthography: yīgòng, yī bān, yǐdiānr, yī tiān, yī nián.

Traditionally, qī ‘seven’ and bā ‘eight’ used to be pronounced as /tʃhi/, ‘pa/ before a T-4: qīyuè [tʃʰjɨjɛ] ‘July’, bā-wèi lāošī [bɔwɛjˌlɑɔʂɨ] ‘seven teacher’. However, such peculiarity is now almost lost in contemporary neutral Mandarin, and could even be regarded by some as regional, or too ‘old·Pekinesesquē’. Official spelling does not record it, by the way, and today we only have qīyuè [tʃʰjɪjɛ], bā-wèi lāošī [bɔwɛjˌlɑɔʂɨ].

Notes:


Here, too, we refer to *Canepari, L.* (2007) *A Handbook of Pronunciation,* § 11.3.3.1 et seq, with important enrichments and improvements.


Despite remarkable differences, in terms of tonetic notation and iconic representation of intonation, our approach has many points in common with what Cáo Wén presents in his *Hányǔ Yùyīn Jiàochéng,* especially § 4.4.1-4 (he even introduced the notion of ‘transitional tone’, *guòdú shēngdiào,* for the occurrences of ‘‘ instead of ‘‘). It is sadly surprising, then, that his most recent *Chinese Pronunciation Practice* has simplified everything, ‘downgrading’ the whole treatment of tonemes and tones to nothing more than a sketchy resumé.

5. An introduction to Chinese intonation

5.0. This chapter is presented as an introduction to Chinese intonation, and not as a comprehensive description of what we have observed so far. There are good reasons for keeping such a prudent approach, considering the highly peculiar mechanisms determining intonation in a tonal language. Thus, as a mark of intellectual honesty, we prefer to describe what appears to be frequent and truly useful to foreign learners, instead of diving into all the fascinating—but sometimes puzzling—phenomena of the ‘Chinese lilt’. ¹

Intonation surely is what conditions the most how a language sounds, i.e. its melodic flow and its rhythm. Unfortunately, while children find it easy to pick it up just out of spontaneous imitation—sometimes even before mastering contoids and vocoids—many adults often never attain a decent approximation to the right intonational patterns of the tongue they are learning. Obviously, there is no need to sound exactly as a native speaker, but improving intonation really makes conversation smoother and also more enjoyable, not to mention that it avoids some curious misunderstandings too, e.g. ‘are you asking me, or are you telling me?’—‘hey, dude, don’t speak to me in that tone!’ and the like.

To put it bluntly, learning intonation is a hard work. In fact, an adult brain is not as physiologically retentive as a younger one: very little can be done about that; however, a lot could be done by teachers and authors of language courses, to provide the student with—at least—a clear and reliable guide to intonation. Regrettably, a lot more should be said about how most of them parously fail to do so, provided they do not omit the matter at all in the first place (which is not necessarily an evil, since a good number of handbooks just mislead the student). ²

5.1. First of all, one has to bear in mind that, no matter which language we are dealing with, intonation is primarily the product of tonality: in other words, when the pitches of the syllables in an utterance are put together in rapid succession, a precise melodic curve is generated. Linguists analyze those curves and extract what is recurrent and typical in their shape; once conveniently normalized, then, a small
number of representative ‘shapes’ are classified as the intonational patterns of that particular language.

Intonation plays two very important roles in spoken language. First, it helps distinguish similar sentences with different communicative purposes; let us consider this minimal conversation:

– Really?
– Really.

It is clear that, from a phono-syntactical point of view, there is no difference between these two one-word sentences, apart from the way we ‘sing’ them, respectively with a rising or a falling pitch contour. According to canIPA Natural Tonetics terminology, we prefer to say that an interrogative intoneme /?/ and a conclusive intoneme ./ have been used.

Changing the pitch of a sentence does not affect its word order, ie its (syntactical) linguistic arrangement (though, in English, most interrogative sentences and their affirmative counterparts may have a different structure); however, it cannot be denied that intonation plays a ‘syntactical’ role in the aforementioned examples, as it determines the nature of a given utterance, pretty much like a phoneme is capable of determining the meaning of a single word. In this chapter, then, by saying intonation, we primarily mean linguistic intonation.

Secondly, tonality may also work to express the speakers’ feelings and mood, as well as their point of view about the matter being discussed, regardless of whether the sentence is a question or an assertion. Such a use of tonality for ‘emotional’, non-linguistic (or paralinguistic) purposes is better called paraphonic intonation, or simply paraphonics (although this last term also includes non-tonetic paraphonic features, such kinds of voice, and articulatory settings), and it is important to distinguish it from intonation proper, in order to devise a realistic and normalized picture of both. Regrettably, that is what most language handbooks tend to mix up.

All we have just said, with due differences of course, can be applied to any language. However, the tonal nature of Mandarin triggers an obvious question: how does intonation work in a language where most of its syllables already have a distinct pitch contour of their own? Well, that is where things get really tough.

5.1.1. In learning intonation, the first step is to perceive how we modulate our voice when we speak our own mother tongue. Surely, it would be better if everyone
knew and could use the neutral type of pronunciation, at least when it is required
for advisable to do so. Anyway, the natural approach is equipped with everything is
needed to analyze one’s own peculiarities and compare them with the actual neutral
pronunciation. For that purpose, we provide hereby the intonational patterns relevant
to modern neutral British pronunciation (see below).

Clearly, this is not the easiest of tasks, since English is a worldwide, multi-centric
language with more than one accepted ‘norm’... not to mention those readers of
ours who are not native speakers of English, and –though fluent in the language–
will almost certainly speak it with a more or less clear foreign accent. In this last
case, it will take longer to master neutral English intonation, but the process is sub-
stantially the same applied to native speakers.

Besides what we concisely report here, a number of useful reference books are
listed in the Bibliography. For a thorough, clear and practical treatment of English
intonation (though relatively concise, in comparison with so many whole –but in-
complete– books on the subject), the most realistic and satisfactory texts are English
PronunciationS (‘EPs’) and The Pronunciation of English around the World (‘PEW’,
particularly, § 36-49. They follow the Natural Tonetics Method, which is not mingled
with paraphonics, by clearly separating it, though –of course they are closely
related). They also include the intonation patterns of all the accents described there.
It is very important to clearly know that the NT method has nothing to do with
such unsatisfactory things as ToBI and other merely acoustic treatments.

Native speakers of Italian will find their language described with the same level
of accuracy (and regional variants, as well) in Manuale di pronuncia italiana, while
A Handbook of Pronunciation provides only seemingly less extensive analyses (be-
cause they must be considered together with § 1.1.5-8 – or, better still, with § 12.14
of Natural Phonetics & Tonetics) also for French, German, Spanish, Portuguese,
Russian, Arabic, Hindi, Japanese, and Esperanto. Finally, a large number of other
languages are more concisely treated in Natural Phonetics & Tonetics.

5.2. Once the goal of our study has been clearly determined, an iconic represen-
tation of the intonational patterns is needed. We adopt a four-lined, three-spaced
‘staff’, akin to that employed in the Gregorian Chant, and based on very similar
(but relative, not absolute) principles. This is called a tonogram (fig 5.1).
A ‘graphemic-iconic’ representation of British English intonation.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>See you on Saturday.</td>
</tr>
<tr>
<td></td>
<td>(Will they) see you on Saturday?</td>
</tr>
<tr>
<td>2</td>
<td>(If they don’t) see you on Saturday... (it’ll be a total disaster.)</td>
</tr>
<tr>
<td>3</td>
<td>(If they don’t) see you on Saturday... (don’t worry about it.)</td>
</tr>
</tbody>
</table>

Each syllable with a primary or a secondary stress—and a marked toneme, as far as Chinese is concerned— is represented by a line, which, like the ancient neumes (signs), symbolizes a relative pitch and strength, according to its position, height, orientation and shape, on the tonogram. Unstressed syllables—featuring the unmarked toneme, in Chinese—are better represented by dots; though, in fact, these are to be meant as very short, nearly ‘punctiform’ segments, not extended enough to produce any peculiar contour shape.

At this early stage, however, we shall employ the alphabetic graphemes themselves, instead of line and dots, in order to make our approach even more intuitive and easy to understand. Please observe the examples in fig 5.1.

5.3. No matter which language is taken as a starting point (here, we consider English, of course), it is important that any example is repeated several times, until the speaker is able to enunciate each syllable with the right pitch. If necessary, good recordings of neutral speakers, like dubbers and experienced actors, may provide an excellent source to start from. At this point, the speaker’s own pronunciation shall be recorded and then compared to that of the neutral speakers available.

Following, one must develop a satisfactory level of proprioceptive ability. As in learning tones, ‘a good ear’ helps, but it is not indispensable, provided the learners can analyze their own voice and determine whether a certain syllable has a high or a low pitch, possibly with a falling or a rising contour. This is not only necessary, but also useful, in order to modulate one’s own pronunciation with full awareness and flexibility.
The intonation of English

5.4. From the very intuitive—but not entirely rigorous—depiction of intonation as presented in fig 5.1, let us now become familiar with more methodical tonograms: fig 5.2-3 show the four pre-intonemes, and the four intonemes, of British English (as per EPs and PEW, where neutral American English and a well-meditated ‘International’ English are described as well, together with a great number of regional accents). Each tonogram is presented twice: on the left, the strokes and dots are connected in a quite realistic depiction of the speech continuum (for a smoother transition from fig 5.1), while the tonograms on the right—those currently used—are more schematic, but totally sufficient for anyone who knows how to interpret them. In fact, the exact collocation of our tonographic symbols results from the ‘arithmetic mean’ of numerous utterances, once their intonational behaviour has been accurately analyzed and normalized.

fig 5.2. British English preintonemes (realistic and schematic).

fig 5.3. British English intonemes (realistic and schematic).

5.5. Our approach to intonation implies a short introduction to what intonemes and preintonemes are. Roughly speaking, both can be described as the intonational units in which a sentence can be ‘broken down’ and analyzed, irrespective of syntactical and lexical boundaries (mirrored by traditional punctuation marks, in the

65
A preintoneme and an intoneme taken together are usefully called an intonation group.

In non-tonal languages, an intoneme is the intonational unit whose ‘tonal core’ coincides with the pitch contour of the last fully stressed syllable of an intonation group, therefore called a tonic syllable. A number of unstressed (or weakly stressed) syllables, each with a relative height, can surround the tonic; in that case, stressed and unstressed syllables interact to create the tonal contour, distinctive of that intoneme.

An ‘ideal’ sequence would feature one pretonic syllable before the ‘core’, and two posttonic ones after it (further classifiable as internal and terminal posttonics). Of course, monosyllabic intonemes exist, too; if so, the tonic is often slightly lengthened, in order to accommodate the entire contour, which—in turn—tends to be compressed, both in its length and extension.

As a general rule, there are four intonemes in any language, whose symbols end a transcription: conclusive ./ — interrogative ?/ — suspensive ;/ — and continuative ,/ (each will be explained below). Sometimes, and especially in tonal languages, the difference between certain intonemes can be slight, and even partially blurred by paraphonics. However, it is advisable to keep using all four intonemes anyway, since they remain functionally distinct, even when there may be a tonetic overlapping.

As its name suggests, a preintoneme is the intonational unit before an intoneme in the same intonation group, more or less with the former introducing the latter. Obviously, their symbols begin a transcription (and, as they are different from those of the intonemes, there is no risk of confusing them). For example, in English and many other non-tonal languages, the ‘foreshadowing’ nature of the interrogative preintoneme, ?/, is well demonstrated by its shape, which appears as a compressed and shorter version of the interrogative intoneme, ?/. Similarly, the unmarked preintoneme, logically represented by a ‘zero’ symbol, /, is somewhat reminiscent of the conclusive intoneme, ./, and the continuative intoneme, ,/ — with which it is most often associated.

Finally, there are the emphatic preintoneme, /\, and the imperative preintoneme, /\: some may argue that these are not purely intonational units, but rather something more on the paraphonic side. Surely, /\; /\ are employed where paraphonics naturally plays a heavy role, yet there is an unmistakable linguistic feature to be considered, and that we regard as prevailing.
Now, let us examine a sentence like his cousin's name is Bartholomew /bæθoʊməw/ The intoneme corresponds to the full name of Bartholomew, while the preintoneme is everything prior to it: his cousin's name is... If the example were his cousin's name is Dick /dɪk/, the intoneme would be is Dick. The tonic, then, would absorb part of the tonal contour that, under normal circumstances, is 'carried' by the posttonic(s). In consequence, when only one syllable is present (as in the answer to what is his cousin's name? – Dick), the result is a fusion of the expected pitch patterns, which nevertheless maintains the characteristic movements, in an attenuated form (fig 5.4).

fig 5.4. Intonemes with different number of syllables (from tonic to posttonics).

Once again, we remind our readers that the parts of an intonation group do not necessarily respect lexical or syntactical boundaries, as the following examples clearly show:

That's my favorite dictionary [θæts mæθfərɪŋp ˈdɪkʃəni];
That patient thinks he's Giuseppe Verdi [ðærp ˈpiːʃənt θəɪks ɪzˈdʒuːspɛpi ˈvɜːrdi].

In fact, the intonemes in these utterances are, respectively: [θæts mæθfərɪŋp] (rite dictionary) and [ɪvərdi] ([p]e Verdi); while the preintonemes are [θæts mæθfərɪŋ] (My favo-) and [ðærp ˈpiːʃənt θəɪks ɪzˈdʒuːspɛp] (That patient thinks he's Giusep-).

5.6. Let us observe, then, the sequence see you on Saturday in fig 5.1, where the graphemic text is arranged in the form of an intonation curve (instead of lines and dots, which will be the norm elsewhere). The conclusive utterance contrasts with the interrogative one (of a 'total' question); then, we find a suspensive utterance, to be opposed to a continuative one.
5.7. In the last two sentences, the semantic importance of what follows (given in parentheses) is fundamental, whether it is expressed out loud, or it remains implicit, instead. Evidently, the role of syntax itself is irrelevant, in this case.

The crucial point is that a **suspensive intoneme** is characterized by an element of **suspense**, as its name implies, i.e. a decidedly greater and more immediate anticipation; while a **continuative intoneme** precisely aims at smoothing down the transition from one unit to another. As a matter of fact, a **continuative intoneme** can be viewed as the ending part of an **unmarked preintoneme**, which is ‘promoted’ to an intonemic status for being comparatively more prominent, thanks to a slightly larger amount of phonatory energy.

5.8. In neutral British English, the **conclusive intoneme** is falling, /\.\r (see § 5.3, as well as eg 1, 3, 4, in § 5.1): **Bartholomew** [baθələmj ju:] (which, to be meticulous, means \[baθələmj ju\]). At the very opposite, the **interrogative intoneme** is rising (/?/ [\.
\.\r]), as in the one-word question **Bartholomew?** [\əbaθələmj ju:]. 3

The **suspensive intoneme** is falling-rising, /\r/ [\.
\.\r\.\r]: **Bartholomew...** [baθələmj ju:] (ie \[baθələmj ju\]). It is to be underlined that a certain resemblance between [\.
\.\r\.] and [\.
\.\r\.\r] is undeniable, as the posttonics are rising in both cases: however, this happens at different heights, and with a different contour. In fact, to have a totally similar pattern, /\r/ should produce something like [\.
\_\_\_\.] or [\.
\.
\_\_\_\.\r], which is clearly not the case. That is why we disapprove of the classical British School approach –shared by many, alas!—of reducing [\.
\.\r\.] and [\.
\.\r\.\r] to ‘two sides of the same coin’, merging both in a very broad & general ‘rising intonation unit, ‘\r’, or ‘\r’, or ‘\r’.

5.9. § 5.2 and the second example of § 5.1 also feature the **interrogative preintoneme**, /\?\r/, whose characteristic movement anticipates (as said, in an attenuated shape) that of the following interrogative intoneme. Thus, /\?\r/ is precisely the typical intonation pattern found in ‘total’ questions, those which one would normally reply to by simply saying ‘yes’ or ‘no’, ‘maybe’, ‘I do not know’, and the like: **Another cup of tea?** – Yes, please.

But /\?\r/ does not occur anyway. In fact, /\?\r/ is peculiar enough to ‘stand alone’, as in the so-called **partial questions**, whose interrogative nature is self-evident, due to the presence of at least one ‘question word’, such as why, when, who, how... An an-
swear is required for that ‘part’ only, not on the whole content of the utterance. Then, by asking *What would you like to drink?*, the speaker expects us to name some kind of beverage, like *A cup of tea*, not to confirm our ‘willingness to drink’.

Generally speaking, basic partial questions end with a *conclusive intoneme*, /¿ ./; but, as this might sound too blunt, or even rude (depending on the communicative context, the relationship between speakers and listeners), a ‘gentler’ version exists, employing a *continuative intoneme*, /¿ ,/ –or, more ‘colorfully’– an *attenuated interrogative* one, /¿ ?/.

Plain /¿ ?/ should be carefully avoided in partial questions, as it would sound redundant, or convey another meaning, e.g. *How much?* /¿hao’men/. (or /’men,/, /’men?/; asking for a price) vs /¿hao’men?/ (implying: *Are you really asking me ‘how much’? It’s a gift!*). It is clear that the last mentioned *How much?* is a total question, after all, and such a reply as *Yes, that’s what I asked* would be perfectly normal. If said in a surprised and unbelieving way, the structure would become /¿hao’men?מו/ (possibly with an emphatic stress, and –no doubt– with a suitable paraphonic setting, such as *raising* or *expansion* of pitch).

A lot more should be said and explained; instead, this short –and purposely incomplete– introduction precisely aims at spurring our readers to read up more on the intonation of their own mother tongue (by referring to the various books suggested above) as we finally introduce the intonational patterns of Mandarin Chinese.

**Fig 5.5.** Difference between total questions (1) and partial questions (2).

<table>
<thead>
<tr>
<th></th>
<th>(Will they) see you on Saturday?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(Why won’t they) see you on Saturday?</td>
</tr>
</tbody>
</table>

**The intonation of Mandarin Chinese.**

5.10. At this point, we give it for granted that our readers are fully acquainted with the characteristics of the *tonemes*, and with their essential role as ‘semantically pertinent components’ of any Chinese syllable (including those with ‘no tone’, or better with a ‘zero toneme’), just like vowels and consonants. However, as in nontonal languages, pitch modulations are also employed for intonational purposes.
This double use of tonality is actually very complicated for non-native speakers, including the Chinese whose native ‘dialect’ is remarkably different from *pǔtōnghuà*: inevitably, intonation overlaps and intertwines with the tonal flux, affecting the height and duration of the basic tones and their taxotones. More importantly, the intonation patterns can modify the actual *shape* of syllabic pitch contours, but always within an acceptable range of variation, so that their ‘tonemicity’ is preserved, and all elements remain recognizable as such. In other words, while a toneme’s *relative pitch* does not change, with respect to other tonemes in the same intonation group, its *absolute pitch* must conform to the relevant intonational ‘grid’. Such ‘grids’, which obviously correspond to Mandarin *preintonemes* and *intonemes*, are clearly depicted in fig 5.6.

However, compared to the English tonograms seen above, these ones feature a striking—and, apparently, inexplicable—difference: they are *empty*! Actually, the intonation patterns are still there, visible and distinguishable: to put it simple and straight, the intonation does not lie in the ‘content’ of the tonogram (strokes and dots, as in non-tonal languages), but in the… ‘box’ itself, so to speak. The tonograms *can* be filled with actual tonal symbols, but only when *actual tones and taxotones are determined*, each with its distinctive shape (which will be subsequently placed in appropriate height, and slightly ‘reshaped’, where necessary).

5.11. The peculiar nature of Chinese intonation, then, makes it necessary to clarify how to interpret our *Mandarin* intonation symbols and diagrams. In non-tonal languages, they generally represent exact tonetic values, as we have shown by accurately transcribing various occurrences of *Bartholomew*, where [·] and [:] represent the actual pitch of the two posttonic syllables (respectively, in the *interrogative* and *suspensive* intonemes).

In tonal languages, instead, the intonation symbols merely suggest what sort of placement and reshaping will affect the tones and taxotones. Let us shortly review our tonemic and tonetic symbols (and the corresponding tonograms, which are more explanatory than any description): [‘]/ [·], [:]; [‘]/ [:]; [‘]/ [:], [:]; [‘]/ [:], [:]; [‘]/ [:], [:]; [‘]/ [:], [:]; [‘]/ [:], [:]; /s/ [:], [:], [:]. It is self-evident that adopting something like [‘], [:], [:], [:] to show intonation would be highly confusing and misleading.

After all, what matters is having a generic indication of the pitch *range* (raised vs lowered, compressed vs expanded) % for the *movement* (horizontal vs ascending vs de-
scending). Therefore, [·] or [:] do not refer to steep movements from up to down, and vice versa; nor should [:] be misinterpreted as an absolute and mandatory expansion for *everything*. On the contrary, [:] simply means that the tones and taxotones preserve their fundamental characteristics of pitch, duration and shape, for there is no compression towards the middle, as in [:] (not to be confused with [\$], a taxotone of ‘zero’ toneme, /\$\!/), nor any gradual rising, [:], or falling, ['].

Numerous handbooks, including many of those published in China, resort to arrow-like symbols to show intonation: ‘[t], [i], [\t]’; however, we are convinced that such ‘darts’ are not as intuitively self-explanatory as our [:], [:], [:] (not to mention that canIPA better employs ‘[t], [i], [\t]’ for sociophonetic remarks). 

![fig 5.6. Mandarin preintonemes and intonemes.](image)

5.12. This being said, let us now examine the Mandarin *preintonemes*: the generic symbols employed in our phonemic transcription – [·] / (for the ‘unmarked’ preintoneme) & /\z\/, /\i\/, /\i\/~ can be safely employed at the phonetic level too: [·], [\z\], [\i\], [\i\/~]. Wherever it is required and convenient to be more explicit, preintonemes can be represented by tiny circlets –actually, empty dots– in *double square brackets* (those reserved for ‘hyperphonetic’ transcriptions), in order to keep them consistent with –but also clearly distinguishable from– the black dots symbolizing intonemes.

The *unmarked preintoneme*, / [·], features a moderate compression towards the lower range of the tonogram, [\e\], which makes tonal transition easier, by avoiding too large tonal ranges.

The *interrogative preintoneme*, /\z\/ [\z\], is compressed upwards, which means that the lower edge of the tonal band is evenly rising, [\e\]. As a result, all tones will be risen too, and their typical contour will change accordingly. As in / [·], the tonal range is usefully moderated; but, thanks to a relatively (but significantly) higher
register, the listener knows it is a question, not an assertion. Indeed, the actual (ie ‘measurable’) difference can be minimal, and paraphonics may even almost nullify it; but a native speaker is still capable of telling a lowered high pitch from an elevated low one, as the human brain processes spoken language, not merely as sound waves (this being the main limit of the acoustic approach), but as the hearable counterpart of a mental act.

As far as foreign learners are concerned, no single recipe can work for everyone. Those who already speak a tonal language are accustomed to pay attention to such subtle differences in pitch range, while those who speak a non-tonal language tend to be more focused on pitch movements, as determined by a series of monosyllabic pitches (again, not just their height, but –more importantly– their shape, ie their form and orientation).

Listening to good recordings is the only and indispensable way to assimilate intonation at the passive level; but practicing intonation just through plain imitation is of little use, and potentially sterile (unless one is particularly gifted). Our tonograms represent an invaluable aid to understand what learners should pay attention to, when listening to Mandarin audio files, so that ‘rehearsing’ intonation may be a meditated and conscious act of speech, not a random and erratic one.

5.12.1. The imperative preintoneme, /反复/ [j], used in giving peremptory orders, features a gradually falling movement and a certain compression (upwards, at the beginning; then, downwards, finally): [.Gradient].

The emphatic preintoneme is generally not compressed at all: /反复/ [反复]; or, if we chose to consider the unmarked / / [ ] as the preintonemic ‘benchmark’, [Gradient] could be regarded as an expansion of [Gradient]. Honestly, it is nothing but a matter of point of view. What matters is that all tones and taxotones in an emphatic preintoneme will be articulated full and clear, as in single words when read aloud from a list (which is typical of the innumerable vocabulary lists that abound in Chinese language courses). Such distinctiveness is obviously coherent with the communicative purpose of emphasis.

All taxotones occurring in the various preintonemes are shown in fig 5.7, printed in gray; while the four ones printed in black, in the emphatic preintoneme, correspond to the ‘basic’ tones, as pronounced in isolation. These should be studied carefully, and taken as a reference to understand the changes affecting other taxotones in real, extended utterances.
5.13. Before introducing the four intonemes, we remind our readers that their tonetic notation employs a system of solid black dots, and therefore it is different from the tonemic one (while, for the preintonemes, some ‘circlets’ can be used, but are confined to extremely accurate transcriptions, as previously said).

The conclusive intoneme—or assertive intoneme, though not used exclusively in assertions—gradually compresses its tonality downwards, thus producing an overall falling profile: /./ [:]. In any language, that is the most frequent and versatile of all intonemes; even more so in Chinese, where concise and straightforward utterances are the norm, and intricate constructions are virtually absent. In fact, even if intonation is not directly related to syntax, it is undeniable that, the wordier a sentence is, the more frequent /./ and /;:/ will be.

5.13.1. The interrogative intoneme is rising, /?/ [:], as its tonality is progressively raised, and the lower edge compressed upwards. As such, Mandarin Chinese conforms to a sort of ‘international’ trend, where a rising intonation is generally associated with a question, and contrasts with falling—or level—non-interrogative intonemes. (To be true, almost half of the world’s languages have rising-falling interrogative intonemes, /?/ [:], where a rising part is still there, anyway; this can happen especially when the suspensive intoneme is a rising one, /;:/ [:].)

The Mandarin suspensive intoneme is expanded, without any distinctive movement: /;:/ [:]. This helps express the needed suspense, that may be due to a sense of excitement, or fear (though not with paraphonic relevance yet), or strong opposition between two concepts, or between a condition and its outcome, and so on, depending on the context and content.

Finally, the continuative intoneme keeps a middle level tonality: /,/ [:]. It can be considered as the unmarked one (similarly to the ‘normal’ preintoneme, / / [ ] [ ]), though on a slightly higher register, as it makes tone articulation easier, and adds no emphasis to the utterance.

Fig 5.8 shows ‘isolated’ tones in black ink, as pronounced in vocabulary lists, or in /;:/ [:], as well as in /;:/ [:]. Besides, in gray ink, all the taxotones of the four marked tonemes are shown, as they are realized in the conclusive (/./ [:]), interrogative (/?/ [:]), and continuative (/,/ [:]) intonemes. We urge our readers to carefully observe all the—often subtle—tonal alterations in their shape and height, due to different intonational constraints.
fig 5.7. Taxotones in preintonemes.
5. An introduction to Chinese intonation

Fig 5.8. Taxotones in intonemes.
5.14. It would have been pretty absurd to present the Chinese intonation structure, without describing its general nature, in the first place, and providing practical tools to reach a satisfactory awareness of the different intonational phenomena. Clearly, Chinese tonograms will mean little, at first, if not methodically compared to those of one’s own language, considering some variants as well.

Undoubtedly, such comparison is easier when dealing with non-tonal languages; but, again, our tonograms offer a sure helping hand, provided there is an auditory feedback, which learners can rely on. In fact, listening to good recordings is important to get acquainted with rhythm and most recurrent syllabic structures, and intonation too, in a broader sense. But combining tones and taxotones with intonation is so delicate a task, due to continuous register changes, that it would be impossible to make any progress without the reassuring support of good native neutral voices, properly recorded. In addition to the audio quality, the quality of learning is also important: if necessary, difficult passages should be listened to, bit by bit, at lower speed, since the very beginning.

5.15. Let us deepen our understanding of how total questions are build up. Typically, they will feature the same syntax as that of the corresponding assertive utterances, plus an unstressed ‘modal’ particle, like ma and ne. While ma is generally quite neutral, ne implies that the asker is seeking confirmation to an hypothesis or, more broadly, is expecting the answerer to express an opinion. /¿/ is regularly used, but, with ne, /¿./ (& /¿./) can be used as well, expressing different nuances:

_Tā bù lái ma?_  
/¿θaːlpuːlæʰmaʔ/  
[¿θaːɻʌ̃lɐːmaː]  
{he/she not come *MOD}  
‘Is he/she not coming?’

_Tā bù lái ne?_  
/¿θaːlpuːlæʰnxʔ/  
[¿θaːɻʌ̃lɐːnx]  
{he/she not come *MOD}
Tā bù lái ne? said as /tʰâlpʊɐlæŋx/ implies a certain surprise: 'Oh, so you mean he/she is not coming? Is that a fact?'; and it is a total question: to be precise, it is a request of confirmation to something that is clearly viewed as bizarre % unexpected. Instead, Tā bù lái ne? /tʰâlpʊɐlæŋx/ (or /ŋx/, with appropriate paraphonetic settings), is a total question just on the surface: in fact, it is a kind of partial question, which we could render as 'what (would happen) if he/she does not come?'. Clearly, we cannot simply answer that with 'yes' or 'no', but with something like 'no big deal', or 'it will be a trouble, then' &c. One could also introduce emphasis, of course: /tʰâlpʊ; ɐlæŋx/. /tʰä-hä-ˈbu; ɐlæ-ˈnx/ 'what if he/she does not come?'; /tʰâlpʊ; ɐlæŋx/ /tʰä-hä-ˈpu; ɐlæ-ˈnx/ 'what if he/she does not come?'.

Caution is advised when dealing with ne, as this particle is not primarily an interrogative one (like ma) and has got a wide range of possible applications. An assertive clause like tā xiùxi ne is perfectly legitimate, but it means 'he/she is studying', ie an action predicated while it is happening.

The same polysemy holds true for a [ə]a, ɥa, ʊa (and its further euphonic variants, with a separate spelling: ya [jə], wa [wa]), and ba. When used in questions, these two particles imply that the answerer is supposed to express, respectively, confirmation or denial (ie consensus or refusal). Please, pay attention to how the same particle is used both in questions and in answers:

Nǐ bù lái ya?

/ni, ɲpʊlæˈja/  
[ŋpʊlɛˈja]  
{you not come *MOD}  
'So you are not coming?'

Shè a.

/šuˈa/  
[ʃuˈa]  
{to-be-so *MOD}  
'Right.'

Zànmén qù ba?

/ʃəz âmˈmən ʈʃyɔpa/  
[ʃəz ɹəmˈmən ʈʃyɔpa]  
{we go *MOD}  
' Shall we go?'

Qù ba.

/ʈʃyɔpa/  
[ʈʃyɔpa]  
{go *MOD}  
'Let's go.'

In nǐ bù lái ya? (which could be rendered also as /niˌɲpʊˈlæˈja/ [ŋnɪbʊˈlæˈja]), the asker already has a rough idea about the answer, and is 'double-checking' it.
Zánmen qù ba? – instead – is a question and, at the same time, a kind invitation; that is why ba is perfectly logical when replying too, since the answerer is not giving an order, but accepting the invitation with a mild exhortation. Surely, the sole interrogative intoneme may suffice to ask a syntactically unmarked question, especially in very colloquial speech (as we can do in English):

Zánmen qù?
/ʔtsam-myn |tʃyʔ/  
/Qù.
/ʔʃhi/  
[we go]  
{go}  
'Shall we go?'  
‘Let’s go.’

One should be careful not to abuse such possibility, given the intrinsic non-neutral acceptance attached to it: in fact, zánmen qù? could imply ‘I did not expect we were going’, or ‘did I misunderstand you?’ (or ‘are we really going?’), just to mention two interpretations; many more are possible, thanks to the multi-faceted contribution of paraphonics.

5.16. As previously said, Chinese partial questions often feature a different syntactic arrangement, compared to English ones. But, as in any other language, they require a selective answer to a specific topic, signaled by an appropriate interrogative word: who? – what? – where? – when? – why? – how? – how many? – how much? and so on. Clearly, these pronouns/adverbs shall receive a noticeable stress (possibly reinforced by emphasis, whenever the actual communicative purpose requires it). The ‘hybrid’ nature of partial question is mirrored in their intonational patterns: as in English, we have an interrogative preintoneme, but not a plain interrogative intoneme. Moreover, we find the same intonemic gradation from an informal /ʔ/ [ʔ] (somewhat ‘cold’, potentially inappropriate with strangers), through /ʔ.ʔ/ [ʔ.ʔ] to a milder /ʔ./ [ʔ.:], and up to /ʔʔ/ [ʔʔ] (where o indicates the attenuation of a preceding intoneme). In the following examples, only the first one will be used:

Nǐ shénme shìhou qù Luómā?
/ʔni ʔʃæ̇m-mə ʂʊəɻou, ʔʃhi ʔtwɔ complimentary./  
[ʔni ʔʃæ̇m-mə ʂʊəɻou, ʔtʃyʔ ʔtwɔ complimentary.]  
{you which? time go Rome}  
‘When are you going to Rome?’
5. An introduction to Chinese intonation

Ni gēn shéi qù Luòmā?
/ʃi ni /kən ʃei ʃi tʃi lwo ma/  
[ʃi ʃei tʃi lwo maa]  
{you follow who? go Rome}  
‘Whom are you going to Rome with?’

5.17. The same is true for exclusive (or alternative) questions, which require a global reply, but focus their actual ‘interrogativeness’ on one portion only, the ‘x-not-x’ structure; from the intonational point of view, then, exclusive questions act pretty much as partial questions do. Therefore, an interrogative interrogative intoneme is not needed, sufficing an interrogative preintoneme. As usual, the intoneme can vary from /. / to /?/. Be noted that the negative adverbs bu and méi are often reduced to bu and mei:

Ni qù bu qù Luòmā?
/ʃi ni tʃi lwo ma/  
[ʃi tʃi lwo maa] (& [bu], ‘attracted’ by the following fourth toneme)  
{you go not go Rome}  
‘Are you going to Rome (or not)?’

Ni zuò méi zuò zuòyè?
/ʃi ni tʃi lwo méi lwo jie/  
[ʃi ni tʃi lwo méi tʃi lwo jie] (& [mei])  
{you do not do homework}  
‘Have you done your homework (or not)?’

5.18. In addition, there is also a kind of selective questions, made up by two (or more) clauses connected by hāishi/’ or...?’. In fact, we are not dealing with ‘individual questions’ proper, but with complex sentences whose interrogative nature derives from the opposition between different choices. Depending on the actual meaning, there can be ‘total’ and ‘partial’ selective questions; respectively: ‘x, or not-x?’, and ‘x, or y?’.
Either way, the first part usually starts with an interrogative preintoneme, /¿/ [¿], while the second part ends with a conclusive intoneme, /./ [\]. To express a stronger participation, we can have /¿ i/ [¿], and even /¿ ?/ [¿]—thus, separating more clearly the two parts—regularly followed by /./ [\].

With more than two alternatives ('x, or y, or z, or...?'), the same scheme repeats before any háishí, which itself normally bears a conclusive intoneme, /'léægu/. Depending on the speaker’s attitude, one alternative can be clearly underlined by ending with /./ [\], while other alternatives (but the last one) keep /./ [\]. Examples:

Nǐ qù háishí bù qù Luómà?
/¿nîtʃy, 'léægu, 'puʃtʃy, 'lwo,ma./
[¿nîtʃy, 'léægu, 'buʃtʃy, 'lwo,maa.]
{you go or not go Rome}
‘Are you going to Rome (or not)?’

Nǐ qù Bâlì háishí qù Luómà?
/¿nîtʃy 'pəli, 'léægu, 'tʃshy 'lwo,ma./
[¿nîtʃy, 'ba/li, 'léægu, 'tʃshy 'lwo,maa.]
{you go Paris or go Rome}
‘Are you going to Paris, or Rome?’

Nǐ qù Bâlì háishí qù Luómà, háishí qù Bēijing?
/¿nîtʃy 'pəli, 'léægu, 'tʃshy 'lwo,ma; 'léægu, 'tʃshy 'pei̯tsin./
[¿nîtʃy, 'ba/li, 'léægu, 'tʃshy 'lwo,maa; 'léægu, 'tʃshy 'pei̯tsin.]
{you go Paris or go Rome or go Peking}
‘Are you going to Paris, or Rome, or Peking?’

5.19. For a full treatment of Chinese intonemes, one final remark has to be made about the conclusive intoneme, /./ [\], which has two apparently opposite functions: (1) to draw less attention than a suspensive intoneme does; and (2) to highlight a word, a concept or an entire clause, even without employing any emphasis.

Should we decide to renounce any conclusive intoneme would result in much longer strings, with most words bound together in almost endless preintonemes.
apart from being unpractical, auditorily, it would produce a flat and 'litany-like' effect, animated only by marked intonemes. An even more unsatisfactory result would be produced if we decided to use /./ or /;/, wherever /./ is normally encountered. Not that this is wrong in se, but it would sound quite unnatural and heavy, unless the speaker intentionally implies a stronger focus on every and each segment of the utterance. Let us examine the following example:

\[
\text{Annà zàoshang gîle chuang xí-zāo, chi diànr záofàn zài chù-ménr.}
\]

Surely, /?an\(\text{na}./\) _tsao_\(\text{gan}._-_tsi-\(\text{le}-\text{tsh}\)wan _si_\(\text{tsao}._\) _\text{tsh}\)u,\(\text{tj}aaz_\)\(\text{tsao}\)\(\text{fan}\), _\text{tsae}-\text{\(\text{tsh}\)u-'m\(\text{y}\)\(\text{u}\)\}. /—with /./ both after \(\text{Annà} and \_\text{xí-zāo}—\) would be correct, as well; but /?an\(\text{na}./\)_tsao\(\text{gan}._-_tsi-\(\text{le}-\text{tsh}\)wan _si_\(\text{tsao}._\) _\text{tsh}\)u,\(\text{tj}aaz_\)\(\text{tsao}\)\(\text{fan}\), _\text{tsae}-\text{\(\text{tsh}\)u-'m\(\text{y}\)\(\text{u}\)\} would not be so (not to mention /?an\(\text{na}./\)_tsao\(\text{gan}._-_tsi-\(\text{le}-\text{tsh}\)wan _si_\(\text{tsao}._\) _\text{tsh}\)u,\(\text{tj}aaz_\)\(\text{tsao}\)\(\text{fan}\), _\text{tsae}-\text{\(\text{tsh}\)u-'m\(\text{y}\)\(\text{u}\)\}. /

5.20 Hereby we present a number of examples featuring various intonemes and preintonemes, depending on actual communicative purposes. We recommend our readers to examine these examples with all due attention, and to methodically compare the phonemic notation with the phonetic one.

\[
\text{Nà gānìăng, jiùsuàn bù bēshàn wō yè huì ài tā.}
\]

(as usual, with the possible variants: [ta „tha, „tha])

\{that girl even if not good I also can love her\}

‘Even if that girl were not good, I would love her nonetheless’.

\[
\text{Wō zhēn xìwàng nêng xìié-bào hǎnyù.}
\]

\{I truly hope I can learn Chinese\}

‘I truly hope I can learn Chinese well’
Wômen shì jìjiě ninde kànsà / kànsà.

{Tongshìmen dōu fèicháng gànxiè nin.}

By underlining fèicháng and nin, for example, the entire tonal structure should be reorganized around the syllables with stronger prominence:

Lían nǐde qīzi yě bú huì zhuō-fán.

Màkè jǐyuè jiù huì guó?

Weishénme bú dào Bèijíng zhǎo gōngzuò?
Púde gòngjià-bànjing shì duōshǎo?
/'pħu•tæ \kʊɐn\tʃjæ\pæ\ʃi\n, figcaptiontwɔ\zæ\s/.
[z̚'pħu•dæ 'kʊn\dæ\'hæn\tʃjæ\'hæn\tʃæ\ t'hæn\tʃæ\ t'su '(foo, sʊ)']

or, more spontaneously:
[z̚'pħu•dæ 'kʊn\dæ\'hæn\tʃjæ\'hæn\tʃæ\ t'hæn\tʃæ\ t'su '(foo, sʊ)']
{protactinium *DE covalent radius to be so how much?}
'What is the covalent radius of protactinium?'

Nǐ zěnme zào zhème shǎde jǔzì?
/¿ni\tsyn\ɔmy \tsa\s, figcaptiontʃy\ɔmy  sæ\tæ\tʃs\tsul./
[z̚'ni\tsæm\mæ \tʂʊ\ʂʊ\mæ ʂæ\dæ \tʃy\dzʊ\s/]
{you how? create this-way stupid sentence}
'How can you possibly conceive such silly sentences (as those)?'

Tā huì yīngyǔ ma?
/¿tæ\hwa\wei \jɪntɪŋ\mæ\s/.
[z̚'tæ\hwa\wæ\ 'jun\wæ\'ɪŋ\mæ\s/]
{he/she can English *MOD}
'Does he/she speak English?'

Nà nǐ bù qù?
/¿na\ni\pʊ\ts\ʃi\r/.
[z̚'na\ni\pʊ\'bi\ts\ʃɪ\r/ (ə [bʊ])]
{that you not go}
'Oh! So, you are not going!?'

Wōmen zōngli shuō-huāng le?
/¿wʊsæm\n \tsɔn\li, figcaptionz\wʊsæ\wæn\dæ\l/.
[z̚'wʊmæ \dʒʊn\lɪ\l figcaptionz\wʊsæ\wæ\wʊŋ\lɪ\l/]
{we Prime-Minister tell-lie *MOD}
'What!? Did our Prime Minister tell a lie!?'
Let us compare also the following occurrences of the same sentence, with emphasis on different words. Double-line symbols, \[\text{\textquotedblleft}'\text{'}, \text{\textquotedblright}'\text{'}\], very intuitively represent the reinforced intensity of emphatic tautosones, with respect to ‘normal’ \[\text{\textquotedblleft}'\text{'}, \text{\textquotedblright}'\text{'}\]. We chose to employ Canepari in these sentences, especially because it is a four-syllable surname, rendered as \[\text{/kha'ni-phali/}.\]
We omit the phonemic transcription, in this case, as it is of very little use to highlight emphatic phenomena; besides, it is a good opportunity for our readers to check their ability to derive a phonemic transcription from a phonetic one:

Kānpālǐ jiāoshòu zài Wēinīsī dàxié jiāng yǔyīnxué.
[Canepari professor to be in Venice university explain phonetics]

[keletal phonetic transcription of pronunciation]

'It is professor Canepari that taught phonetics (not another subject) at Venice University'.

'Professor Canepari (not a generic Mr Canepari) taught phonetics (not another subject) at Venice University'.

'Professor Canepari taught phonetics at Venice University'.

'Professor Canepari taught phonetics at Venice University'.

'Professor Canepari taught phonetics at Venice University'.

'Professor Canepari taught phonetics at Venice University'.

Parentheses & quotations

5.22. Lastly, fig 5.9 shows the diagram of *parenthetic phrases*, or simply *parentheses* (either low, [ [ ]; or mid, [ ] ), and of *quotations*, [ ' ] . In the following sentence, each of them occurs once:
‘First of all’—he said—‘let’s consider “Natural” Phonetics, as it’s properly called’.

 Parentheses typically feature an overall reduction of their stress and an increase in the rate of speech, while the pitch is compressed in the low—or mid—range of the tonogram. Quotations, instead, are quite the opposite, as their role is—precisely—putting one or more words in full evidence, by means of a slightly louder and distinct enunciation; thus, their stress is increased, the speech rate is reduced, and the pitch is raised (without compression).

Marking such prosodic subtleties in phonetic—let alone phonemic—detailed transcriptions is neither necessary nor recommendable. The symbols [, ], [ ′ ] are more than sufficient to bear in mind all these differences, with respect to ‘normal’ utterances. Then, even with regard to stress (and taxotones, as far as tonal language are concerned) there is no need to deviate from the usual notation.

Quotations must not be confused with direct speech. Let us go back to ‘First of all’—he said—‘let’s consider “Natural” Phonetics, as it’s properly called’. In that sentence, only he said should be excluded, because all the rest—and what may follow—is direct speech, indeed.

fig 5.9 Parentheses and quotations.

Chinese parentheses are low, [, ], after a conclusive intoneme; and mid, [ ′ ], after other intonemes; quotations, as said, are high, [ ″ ]. Such intonational superstructures are quite useful in tonal language, thanks to small variations in pitch (small, yet perceptible by native speakers, but not so much—at the very beginning—by foreign learners, who have already a hard time struggling with tones and taxotones in ‘normal’ utterances). A few examples:

Wǒ àì nǐ, qín’àiide. Nǐ yè àì wǒ; name, zánliǎ hǎi děng shénme?

[I love you beloved you also love I then the two of us yet wait what?]

‘I love you, my dear, and you love me. So, what are we waiting for?’
Jírán zánlià xiǎng’ài, qín’ài de, wò xiǎng gèn ni jié-bùn.
"given-that the-two-of-us love each other beloved I want be with you marry"

'Since we love each other, my dear, I want to marry you'.

Tà sī-kái shàngbiàn xièzé «Mākē shuàīgē shòu» de xìnfēng yě kàn, xìnli jiù xièzé: «Wǒ yìjīng jiébùn le, huìtōu zài shuō».

'He opened the envelope headed to handsome Mark, and the letter said I am married. Maybe next time'.

5.23. We think it is useful to present a comparison between natural perception and the acoustic measurements of intonation: Fig 5.10 shows a horizontal tonogram (in Chinese, it could be a / / [ ] "['", or a \a/ [ ] "\"] and -superimposed in dotted line- the tonal range which would be produced by such acoustic softwares as Praat.

The natural approach, as previously said, implies a useful normalization of numerous occurrences of the same phenomenon; the human ear is a selective organ, and only a small portion of the actual sound waves of speech (the one that 'matters') is eventually sent to the brain, which -in turn- discards the incidental and the non typical. This is done, in part, by the language oriented nature of human beings; in part, by the fact of their having pronounced tens of thousands of utterances, and even more of them.

However, no matter whether we are native or (fluent) non-native speakers, almost no-one can be confidently sure of being capable of actually describing intonation units, by simply looking at one of the particular tracings produced by acoustic softwares, if they are not accurately normalized. These tracing are of little practical use, because, paradoxically, they are too finicking. A machine will present a melodic curve as it records it, plain and simple, with everything put on the same level, in-
cluding what the human brain would downgrade at the paraphonic level, or even ignore altogether, for being nothing more than an automatic adjustment. In fact, in fig 5.10, the much broader range at the beginning, and the gradual frequency reduction towards the end, are absolutely normal: that is why our brain perceives a level intonation, instead of a steeply falling one.

fig 5.10. Natural (tonogram) and acoustic (dotted line) depictions of intonation.

Notes:

1 The complex mechanism of intonation is described in depth in Canepari, L. (2007) Natural Phonetics & Tonetics, § 13 et seq.

2 None of the books that we have examined so far offers a really convincing treatment of Chinese intonation, and that explains why the present chapter, though very complex and detailed, practically features no notes. Interestingly, it is an old-fashioned handbook, such as the 1981 edition of the Practical Chinese Reader, that has captured our attention: its pragmatic approach to intonation is reminiscent of the old British school scheme, where /¿/ and /?/ are often merged, and paraphonics is currently mixed up with linguistic intonation; however, the Reader has the merit of getting straight to the point, by coupling many sample sentences—basically, but clearly, annotated— with a number of relevant audio files.

Surely, that is not the most scientific way of analyzing pronunciation, but—at least—it does not muddle up the readers with a bunch of mysterious symbolic conventions, as those adopted in Cao, W. (2002) Hanyu yuyin jiacheng, p. 140-148 (which, frankly, seems quite random and empirical, whereas a good annotation system should always be founded on scientific and consistent elements, like ours preintonemes and intonemes).

3 Among the very few reliable works on the matter, we would suggest Kingdon, R. (1958) The Groundwork of English Intonation, although not completely free from paraphonic mixes.

4 We will note—en passant—the existence of a particular kind of pronunciation handbooks, those compiled for the Chinese themselves who aim at a post where ‘standard pronunciation’ is required: we have found one Henan sheng putonghua shuiping ceshi zbidaoshu ‘A Guide to the Henan Province Common Language Proficiency Test’, where intonation is shown, quite erratically, with a very similar system of dashes and arrows.

5 For example, Swedish and Vietnamese feature /¿/ [¿].
6. Phono-tonetic analysis of extended texts

*The North Wind and the Sun*

6.1. In this chapter, what we have explained so far will be summarized and put in practice, by accurately transcribing two extended texts. Customarily, many linguists choose the Aesopian tale *The North Wind and the Sun*, as a sort of ‘universal specimen’ for phonetic analysis. Actually, a tale is not always a good choice, since it usually contains a lot of descriptive passages, accompanied by a moderate amount of *direct speech*. The latter, in our view (and according to the natural approach itself), is the most natural and representative form of ‘spoken language’; in a tale, however, direct speech is often reduced to very short sentences, offering little room for complex intonational patterns and paraphonic features. Hence the need for descriptive passages, where longer utterances are more common, provided they are... *said* by the speaker, not mechanically read aloud, like a press release.

In fact, reading aloud is nothing but the *phonic rendition of a written text*. Good writers –journalists above all– aim at being as convincing and understandable as possible, while (hopefully!) striving for conciseness, in order to keep their readers from falling asleep. Natural speech, instead, obeys to different dynamics, paying much more attention to rhythm and ‘flow’, rather than simply delivering information. When these two different ways of using language are forced to coexist, results are rarely satisfactory.

In reading aloud –for example, from a newspaper, or a textbook– everyone must have experienced that uncomfortable feeling of ‘something missing’, as if the text lacked strength and balance. Typically, passionate and eloquent speakers have a hard time coordinating the pauses perceived by them as spontaneous and necessary, with those found in the written page. The main reason is that traditional punctuation works –almost exclusively– as a typographical aid to highlight the *syntax* of complex utterances, and only secondarily to mark expressive pauses and emotional features.

As such, punctuation evidently fails at reproducing the colourful richness of real speech, with all its changes in pitch, speed, and paraphonic nuances. That is why ‘verbatim’ transcriptions –like parliamentary and judiciary records– look so redun-
dant in some parts, and desperately elliptical in others. Omitting altogether what readers could not reconstruct by themselves is exactly the price to pay, in order to make written language understandable.

6.1.1. Our version of *The North Wind and the Sun* is a decent compromise, in this regard; yet it remains a step behind what would be judged as an ‘ideal target’ for our *can*IPA phono-tonetic notation. Besides, we added two total questions in *direct speech*, at the end of the tale proper, as it would have been otherwise impossible to introduce a sample of the relevant /ʔ/ intonation.

The Mandarin text is presented, first, in *pīnyīn* romanization: punctuation and hyphenation are somewhat ‘richer’ than what is found in common usage (as *pīnyīn* is usually subordinated to traditional Chinese script, which is very parsimonious with punctuation marks, and –of course— uses no hyphens between Chinese characters). However, such ‘enrichments’ will certainly help our readers parse longer sentences with ease, and will offer a useful reference to underline similarities and differences between grapho-syntactical punctuation and a true intonemic notation.

We provide quite a literal English translation, *not* the English version of the tale. However, those who know Mandarin should better avoid looking at any translation at all, since a residual cross-language interference is inevitable anyway. In order to render the communicative ‘logic’ of the original in a genuine way, any addition is put in square brackets. Verbal persons and tenses are not explicitly marked in Chinese, which rather focuses on distinguishing verbal aspect. We assume the traveller is a male, in order to get rid of such masterpieces as *he or she, him or her, his or her*... The Wind and the Sun, though ‘personified’ here, will be referred to as *it*.


Hǎo... Bēifēng jiù shì-qì dà jǐnr lái jǐnguā jǐnguā; kěshì tā guā de yuè lǐhai, nà-gé rén bā páozì guō de yuè jǐn. Dào mòliǎor, Bēifēng méile fāzī, zhīhào jiù suàn le. Yīhuī Tàiyáng jiù chūlái rěrèrde yì shāi, nèi zōudàorde máshàng jiù bā páozì tuō-le-xiàlai. Suǒyì, Bēifēng bù néng bù chénggèn, dàodì hāishì Tàiyáng bǐ tā běnshì dà.

Nǐ xīhuān zhè-gé gūshī ma? Wǒmen zài jiāng yì-biàn ma?
Once upon a time, the North Wind and the Sun were there disputing whose skill was greater. While still talking and talking again, a traveller came along, wearing a thick cloak on his body. The two, having discussed, said then: ‘The one who can first make this traveller take his cloak off... let us consider its skill as greater’

OK... The North Wind, then, began using great energy and repeatedly blew hard; however, the more violently it blew, the more closely did that man fold his cloak around him. Finally, the North Wind had no way out, and was forced to give up. In a while, the Sun came out and, as soon as it shone scorchingly, that traveller took off his cloak immediately. Consequently, the North Wind could not help admitting that, in the end, the Sun was the stronger of the two.

[Do] you like this tale? [Shall] we tell [it] once again?

A more natural translation:

Once upon a time, the North Wind and the Sun were disputing over which was the stronger, when a traveller came along, wearing a thick cloak. The two made deal and said: ‘Who first succeeds in making this traveller take off his cloak... shall be declared the stronger’.

OK... The North Wind, then, began blowing with full power; however, the more violently it blew, the more closely did that man fold his cloak around him. Finally, the North Wind was left with no other choice but giving up. In a while, the Sun shone out warmly, making the traveller took off his cloak immediately. Consequently, the North Wind could not help admitting that, in the end, the Sun was the stronger of the two.

Did you like this tale? Shall we repeat it?

6.1.3. And now, let us examine the phonotonemic transcription: this does not represent the actual pronunciation, but it is a reasonable approximation to it. As a matter of fact, native speakers and L2 speakers, who already know the language, should be able to read our phonotonemic symbols and mentally ‘translate’ them into their phonetic equivalents, pretty much as they do when reading aloud a passage in pinyin (which, more or less, can be viewed as a non-IPA phonemic transcription in se). With respect to pinyin, however, our phonemic transcription provides some extra—but, decidedly, not at all secondary—information.
First, it clearly indicates any necessary tonal variations, like that of  yī ‘one’:  yī-huí /jiˈɕweɪ/,  yī-ge /jiː-kɻ/,  yī-jiàn /jiˈts箪/,  yī-biàn /jɨpɻ/. Secondly, it avoids ambiguities in rendering ērhwà, as in  yīhuír /jiˈɕwɤ/, (with two tonal changes, too, typical of idiomatic use; while, in a more formal delivery, /jɨˈɕwɤ/ would be preferred). Finally, and more importantly, our readers will find a secure guide to the pronunciation of Mandarin with the right intonation: preintonemes, intonemes, parentheses and quotations (/, /‘/, /!‘//), pauses (of different duration: /|/, /||/, /|||/), and paraphonics, as well.

Regarding this last point, we have decided to be quite ‘sober’. We could not omit the raised larynx (\>\ ), and the backing of the body of the tongue, (\<\ ), so typical of a genuine ‘Mandarin voice’ (but not necessarily of all Chinese languages); and we put them together, (\>\<\ ), before the actual transcription, in order to mean that they are referred to the whole text (or, in this case, to any text in Mandarin Chinese). Then, we simply added (\>, indicating that the following unit (here, a monosyllable: /ɕɑʃ/) is uttered more slowly than usual, to increase the suspense; and (\>, implying that the entire sequence /ɕʊo=mxnɪʦæ, şɪʃian ˈjɪpjɛn-ma?/ is uttered more quickly instead (of course, the paraphonic symbol must be repeated before any new relevant rhythmic group).

But paraphonics can be much more intricate than that, being able to represent the lowering of the habitual pitch extent of a given speaker, (\<\ ); or its raising, (\>\ ); or an overall expansion, (\>\<\ ); or, finally, a compression, (\>\<\<\ ). Of course, these pitch settings can be employed for proper intonational purposes —ie linguistic intonation, between [ ]— as well, and that is why mere acoustic measurements are of no scientific relevance, if not combined with the discerning ability that only the human brain possesses.

Moreover, one speaker may feature a particularly nasal voice; or an exaggeratedly acute pitch, commonly said falsetto; and many more peculiarities. Some of them are typical of a certain language, and thus generally used; others may be limited to one single speaker, or are intentionally employed in that particular utterance, for emotional —if not ‘theatrical’— purposes. Multiple combinations are practically endless. As can be seen, the subject is exceptionally vast and multi-faceted, and we strongly suggest our readers to refer to Natural Phonetics and Tonetics, and to Phonetic Notation, for a more detailed treatment of paraphonics.

\>\>\>\ /\jou\ jɪˈɕweɪ\ /ˈpɛɪfyn\ /ʊxniθæjan\ /ˈtʃxniʦænzn\ ˈtʃxniˈwɪx\ /ˈtʃiɛtʃ\ /pʊnən ˈtæ\ /ˈtʃʊo=tʃ ˈtʃʊo=tʃ\ /ləpɛ ˈjɪʃkɪ ˈʃʊo=tʃətʃ\ /ˈtʃʊo=tʃ\ /ˈtʃhwan=tʃ\,
6. Phono-tonetic analysis of extended texts

It is true that, for such languages as Italian and Spanish, the distance between the phonemic theory and the phonetic reality is so small, that a good phonemic transcription is already enough for a convincing –yet non-neutral– pronunciation. Thus, for Italian *dice* /ˈdiːtʃeː/ ‘(he or she) says’, one could actually say [diːtʃe], instead of the expected [dɪ:tʃː], without any fear of being misunderstood or harshly criticized. In Chinese, the same distance is more relevant, and a word like *jiējié* /ˈʃiɛjɛ/ ‘younger sister’ should not be articulated exactly as [ʃiɛjɛ]; native speakers would certainly understand it, but would not accept it even as a ‘passable’ alternative to the possible neutral realizations: [ʃiɛjɛ] (as [ʃiɛjɛ, ʃiɛj], in fast speech, and in a preintoneme).

In other words, if *pinyin* can largely suffice to ‘transcribe’ Mandarin phonemically (provided intonation, pauses and paraphonics are adequately registered too, possibly via our own IPA symbols), a more detailed phonetic transcription must by provided anyway, in aid to anyone who uses it to learn the actual pronunciation. We recommend to read through the following transcription with utmost caution and attention: this concise phonotonetic *summa* contains a very large portion of what has been explained so far, in the former chapters. There are multiple examples of vocalic and consonantal taxophones, and as many taxotones. There is more than one occurrence of ērhuà and qīngshēng. Intonemes and preintonemes are duly represented, in combination –as said– with a small, but meaningful, number of paraphonic traits.

We remind, once again, that the consonantal gradation, /C/ [Ҫ, Ŭ, Ԩ] and /Ch/ [צ, צh, צʼ], represents the taxophonic extremes reachable in fast, fluid and connected speech, but oscillations are absolutely normal. In particular, for /oC, oCh/,
[\oC, \oC] are more common in preintonemes; while, in an intoneme, /\oC, \oCh/ [\oC, \oCh] (or [\oC, \oCh]), and even ‘full’ [\oC, \oCh]) are statistically frequent. Thus, for better realism, we will precisely adopt [\oC, \oCh] in intonemes, and [\oC, \oC] elsewhere.

6.2. We propose another short text, taken from Hé Ping’s Chinese Pronunciation Course vol.2, lesson 1: “Liùyōu yúdī”, which can be rendered as ‘to preserve a leeway’. In this case, our translation will be less literal, as some passages would be simply meaningless in English otherwise.

**Liùyōu yúdī.**


Yìxióng shàngqié zhīdāo lìuyōu yúdī, « jihuà » shíyòng zīyúán, rén lèi bù gǎi sān-sǐ-ér hòu-xín ma?
Always leave yourself a leeway

In the Amazon tropical rain forests, there is a kind of animal, called an anteater, which is the biggest myrmecophagid in the world, and eats up to sixteen thousand ants per day.

However, what is most surprising is that the anteater has a peculiar habit: when eating ants, it does not simply ‘wipe them out’; as it digs out an anthill with tens of thousands of ants inside, it eats but a small portion of them, no more than five-hundred, totally ignoring the others, and goes straight looking for the next target. Evidently, the anteater clearly knows that, in order for its own species to survive, it is necessary that the ant family lives on.

If even anteaters know how to leave themselves a leeway, and ‘plan’ their use of resources, should men not think thrice before acting, too?

6.2.1. For The North Wind and the Sun, we provided a normalized transcription ‘as it should be’, after collecting a number of recordings from various informants, who, though speaking Mandarin natively, presented different levels of proximity to neutral pronunciation. For Always leave yourself a leeway, instead, we referred to a single audio file in a pronunciation course, publicly available on the market; this obviously bound us to adopt a more realistic transcription, where oscillation from the ‘norm’ will emerge quite plainly.

Let us start, as usual, with the phonemic transcription, before presenting the ‘normalized’ phonetic one. It goes without saying that the reader should pay attention to the differences (often very minute) between the romanization and our canIPA transcriptions

\[\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldotted

6. Phono-tonetic analysis of extended texts
6.2.2. The actual audio file, instead, presented some interesting realizations: each will be presented along with its ‘normalized’ counterparts. Sometimes, the phonetic reality contrasts quite noticeably with what one would expect: thus, where necessary, a concise explanation will be provided.

Starting with ‘yǐzhòng jiào yìxióng de dòngwú’, the speaker produces [jɪɔŋ], instead of [jɪɔŋ], underlining the numeral yǐ /jǐ/ ‘one’; this implies a general reorganization of the rhythmic unit, [$$+$] → [$$+$], also under the influence of the primary stress of [jɪɔŋ] nearby.

In yìxióng de dòngwú, the audio file presents [jɪɛŋdɐdʊŋwʊ], with [r] before a front vowel (a permitted alternative to [ɪ]), and the primary stress on yì, as this is the lexeme –meaning ‘ant’– that distinguishes a yìxióng ‘anteater’ (lit. ‘ant-bear’) from any other kind of xióng. Besides, note [dʊ], instead of [dʊ], as [r] allows for an easier transition from [r] to [ɪ]. Finally, [tʊm, wʊ] reveals a semi-neutral pronunciation, dòngwú, which is nevertheless used by professional voices as well; the unvoiced, non-lenited [t] is as normal a taxophone, in [tʊ], as [r, ɻ] would be, too.
Then, there are two other cases worth noticing: *shǐyìshòu* [ʃi²⁴ji³³ʂou⁴⁴] ‘myrmecophagid’ and *mǎyī* [má³³ji¹¹] ‘ant’.

Thus, their two [ji] are easily explained, when a full toneme becomes a abbix one, according to normal Mandarin structures, producing /jix/ → /jix/ → [jix] for the first one, and /jix/ → /jix/ → [jix] for the second). For /má³³ji¹¹/, there is a clear stress shift: [sz$] → [sz$] → /sz$/ (where /sz/ is a de-stressed syllable, and so a ‘toneless’ one). In addition, of course [sz$] is subject to the strong influence of the conclusive toneme /j/ [j], which lowers considerably its final part.

In *jìngtàn de shì* ‘[what is] surprising is’, the copula *shì* obviously acts as a pivot between a long ‘relative’ clause (the logical subject of the utterance) and the following ‘[the fact that] the anteater has a peculiar habit’; consequently, such pivots are frequently de-stressed: /tsin³³thän³³tsou³³/ [tsin³³thän³³tsou³³].

For *mèi-yī-ge* and *bùfèn* the normalized renderings, /[mèi³³ji¹¹-t⁴⁴]/ and /[bù³³fèn⁴⁴]/, slightly contrast with /[mèi³³ji¹¹-t⁴⁴]/ and /[bù³³fèn⁴⁴]/, as found in the actual recording. Lastly, we have *jiāzú de* [ʨʰja¹³ sued⁴⁴], explainable through the following passages: 

$n[ʨʰja¹³ sued⁴⁴] → [ʨʰja¹³ sued⁴⁴] → [ʨʰja¹³ sued⁴⁴] → [ʨʰja¹³ sued⁴⁴]$ It goes without saying that any language admits a certain degree of –mostly random– divergence from what is ‘normal’, i.e. statistically more frequent, for the simple reason that the very act of speech, though extremely effective, yet is not a mathematically flawless mechanism. What matters is pronouncing correctly all that is crucial to mutual understanding; our brain, meanwhile, will naturally reconstruct what the speaker neglected. However, just to be on the safer side, and as form of courtesy to their listeners, foreign speakers should avoid ‘tricks’, and decidedly stick to the normalized scheme proposed so far (a courtesy that native speakers should better reciprocate, when talking to foreigners).
7. Phonetic transcriptions in current handbooks.

7.1. We are perfectly aware that such a highly detailed phonotonic transcription, as that presented in this work, would be quite challenging for beginners with no previous training in linguistics and phonetics. It must be remembered, by the way, that learning Chinese implies learning Chinese characters, too, and that requires a huge amount of time and energy on the part of the learners. It is not surprising, then, that so many students do not even think about studying Chinese pronunciation in a systematical way: in their view, the task is too complicated and demanding, to be undertaken since the very beginning, when any other aspect of the language (its script, as said, but also its grammar and vocabulary) is to be learnt from scratch.

We shall reply to that by saying, first of all, that being able to write Chinese quick and swift, while speaking it so bad that even native speakers do not understand what we say, is simply pointless, and disrespectful as well. Secondly, in our opinion, most handbooks are not very well balanced, with respect to the space devoted to the basics of pronunciation, though many ‘pronunciation drills’ are generally presented throughout the first one or two volumes of the series (as in the Practical Chinese Reader and similar contemporary textbooks).

The ordinary approach is that of letting students improve their pronunciation simply out of mere imitation and repetition of audio files. With few exceptions, that alone does not work. Hence the need of specifically prepared pronunciation handbooks, for those who want a more complete perspective, and aim at acquiring a native—or nearly-native—pronunciation.

This work, in fact, can be certainly used as a pronunciation handbook, as it provides a detailed description of all phonotonic phenomena that are relevant to the Mandarin pronunciation system. Of course, in an ordinary language course, such an abundant amount of information could not be easily condensed and presented in a readable form. Beginners would invariably find that cumbersome, and not as useful as it can actually be, after a certain degree of fluency and familiarity with the overall ‘sound’ of the language has been attained.
Therefore, a slightly simplified—but not inaccurate—version of our ‘normalized’ pronunciation is the most advisable and viable solution, for teaching purposes. This ‘soft’ approach has another advantage, that is leaving no student behind, even the least interested in phonetics, and pronunciation in general; while the more enthusiast—and gifted—ones will find what is needed to build a strong basis, which will enable them to deal with more complex analysis, with relative ease and proficiency.

7.1.1. Our proposal is simple and effective: for each group of [C] taxophones, just one contoid is used, thus avoiding the aggravation of rigidly respecting the three-way consonantal ‘gradation’, that is natural for native speakers, but not necessarily for foreign learners; namely, our choice fell on [ץ, צ], for stops and stop-strictives, and on [ץ], for constrictives. Here follows a complete list:

<table>
<thead>
<tr>
<th>Letter</th>
<th>Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>[b]</td>
</tr>
<tr>
<td>d</td>
<td>[d]</td>
</tr>
<tr>
<td>z</td>
<td>[ژ]</td>
</tr>
<tr>
<td>j</td>
<td>[ژ]</td>
</tr>
<tr>
<td>zb</td>
<td>[ژ]</td>
</tr>
<tr>
<td>g</td>
<td>[ژ]</td>
</tr>
<tr>
<td>p</td>
<td>[ph]</td>
</tr>
<tr>
<td>t</td>
<td>[th]</td>
</tr>
<tr>
<td>c</td>
<td>[tsh]</td>
</tr>
<tr>
<td>q</td>
<td>[tцh]</td>
</tr>
<tr>
<td>cb</td>
<td>[tцh]</td>
</tr>
<tr>
<td>k</td>
<td>[kh]</td>
</tr>
<tr>
<td>m</td>
<td>[m]</td>
</tr>
<tr>
<td>f</td>
<td>[f]</td>
</tr>
<tr>
<td>s</td>
<td>[s]</td>
</tr>
<tr>
<td>x</td>
<td>[ç]</td>
</tr>
<tr>
<td>sh</td>
<td>[ç]</td>
</tr>
<tr>
<td>h</td>
<td>[x]</td>
</tr>
<tr>
<td>n</td>
<td>[n]</td>
</tr>
<tr>
<td>l</td>
<td>[l]</td>
</tr>
<tr>
<td>y</td>
<td>[j] &amp; [ç]</td>
</tr>
<tr>
<td>n-g[n]</td>
<td></td>
</tr>
<tr>
<td>w</td>
<td>[w]</td>
</tr>
</tbody>
</table>

fig 7.1. Simplified vowel inventory, for teaching purposes.
7.1.2. For [(j)V, (j)VV, (j)VN], a similar criterion is followed, but without renouncing realism: thus, a diphthong like [σʊ] will not be simplistically reduced to ‘[σu]’, even though [ʊ] is not used individually elsewhere. In other words, simplicity must be combined with realism (fig 7.1).

After a well meditated consideration, we have found that fourteen vocoids are ideal for a very convincing approximation to neutral Chinese pronunciation: [i, i, i; e, e; A, a, a; ə, ə; u, u; o]; to these, we add [ɻ, ɻ, ɻ] (for n[ʊ] and the various occurrences of érhùà). We keep [i, i, i], [e, e] and [A, a, a] distinct, as well as [ə, ə] and [u, u], instead of simply [i, e, a, x, u], because certain combinations (especially those with [-n, -n]) are actually easier to reproduce, if the inventory available is just a bit more complex than what would be phonemically indispensable.

\[
\begin{array}{llllllll}
| A | a | \text{[a]} & a | \text{[æ]} & uai | \text{[waɛ]} & ao | \text{[aʊ]} & iao | \text{[jaʊ]} \\
\hline
| (w)an | \text{[waŋ]} & ian | \text{[jɛn]} & ian | \text{[jɛn]} \\
| (i)ang | \text{[jən]} & uang | \text{[waŋ]} \\
| E | e | \text{[ɛ]} & ie | \text{[je]} & üe | \text{[jɛ]} & ei | \text{[ei]} \\
| en | \text{[ɛn]} & er | \text{[jɛn]} & eng | \text{[jɛn]} & ueng | \text{[jɛn]} \\
| I | i | \text{[i], [jɪ]} & i | \text{[u], [ɻ]} & ui | \text{[wɛi]} \\
| in | \text{[iŋ]} & ing | \text{[iŋ]} \\
| O | o | \text{[o]} & ou | \text{[oʊ]} & uo | \text{[wʊ]} \\
| ong | \text{[oŋ]} & iong | \text{[joŋ]} \\
| U | u | \text{[u̯]} & ü | \text{[u̯]} & iu | \text{[jʊ]} \\
| un | \text{[ʊn]} & ün | \text{[u̯n]} \\
\end{array}
\]

It is not hard to notice that our selection tries to incorporate as many vocoids as possible that are in common with the two major accents of English: modern British English, and the so-called ‘General American’ English. Of course, both native and non-native English speakers will use only one of the two accents (or even a different
one, from Australian, or Canadian, to Irish and Scottish, and so forth, not to mention the huge number of regional us accents other than the ‘General’ one). However, what matters is trying to keep certain sounds distinct.

For instance, neutral British speakers will certainly be able to produce a perfect [ɛ], but may find that their [æ] is a more familiar replacement for [ɛ]: let it be so (many Chinese themselves, after all, say [je:n] instead of [je:n]). Similarly, while [ʒ] is common in British English (eg fur [fʒ]), most Americans could not be able to distinguish it from their regular [ə]: no panic, let them use [ə], at first; listening to audio files and practicing kinesthesia, with a constant ear feedback, will enable those really interested in acquiring [ʒ]—and, at a more advanced stage, also [ə]—to succeed. The same story goes, within certain limits, for those who speak languages admitting even fewer distinctions: for instance, neutral Italian and Spanish have only [a], but not [ə], nor [æ]. Articulating [a] everywhere is fine, and phonemically correct; yet it is preferable for the transcription itself to keep using three symbols anyway, as there will be some students who, being more sensitive than others to the quality of vowel sounds, want to know where [ə] and [æ] should be used, in case they manage to acquire them. In other words, this ‘simplified’ transcription is the ideal intermediary between the essentiality of pīnyīn, and the full realism of the neutral ‘normalized’ pronunciation.

Consequently, what we propose here is not the ‘international’ pronunciation suggested in § 2.8.3, on the basis of that originally presented in our Pronuncia cinese per italianti, which was intended—primarily—for someone who simply wants to be understood, irrespective of how foreign his or her accent may sound. Here, instead, we are dealing with something more sophisticated, and more suitable for those who learn Chinese... the ‘professional’ way, so to speak.

7.1.3. As far as ěrhùà is concerned, we have devised a reasonable scheme, derived from that proposed in § 2.7, with just minor adjustments. Then, we have:

A. [a]: huà-huàr [ʰwáwær]; [j]: chuāngliánr [çhwaŋjíær]; [ɾ]: yuánquàn [ɲwàŋçńær]; [w]: hâowán [kɔwànr].

101
B. [v̩]: bêikér [bêi.khɛr] (ŋ[ŋɛr'khɛr]);
 [m̩]: shîjun [s̩i.khɛm] (ŋ[s̩i'khɛm]);
 [w̩]: qûqûr [tʃh̩-tʃhû] (ŋ[tʃh̩-tʃhû]);
 [w̩]: kôuwâr [khou.ʁe] (ŋ[khûw'ʁe]).

C. [o̩]: tûdûr [tu.i.ʁ] (ŋ[tûi.ʁ]);
 [j̩]: zhîiûr [tʃɪ.i.ʁ] (ŋ[tʃɪ'ûi.ʁ]);
 [o̩]: nàimâdor [n̪a.m̩.aʊ] (ŋ[n̪a'maʊ]);
 [j̩]: màimiâdor [m̪a.m̩.aʊ] (ŋ[m̪a'maʊ]);
 [w̩]: shûnûr [s̩am.ʁwɔ] (ŋ[s̩am'ʁwɔ]).

D. [o̩]: piâncâng [phjɛn.tʃhɑ̌], (ŋ[phjɛn.tʃhɑ̌]);
 [j̩]: dûjiâng [dɔju.tʃjɒ̌] (ŋ[dɔju'ʃjɒ̌]);
 [w̩]: tiâncâng [θhjɛn.tʃhɔ̌] (ŋ[θhjɛn.tʃhɔ̌]).

E. [o̩]: bânêng [hɑ̌n̪w̩] (ŋ[hɑ̌n̪��]);
 [j̩]: qûmûng [tʃh̩.m̩.j̩] (ŋ[tʃh̩.m̩.j̩]);
 [w̩]: xiâwëng [tʃjɔ̌.w̩] (ŋ[tʃjɔ̌.w̩]).

F. [o̩]: tôukûng [tʃou.kh̩u̯] (ŋ[tʃou.kh̩u̯]);
 [j̩]: kûqiòng [tʃu̯.tʃhj̩] (ŋ[tʃu̯.tʃhj̩]).

Needless to say, the learners can choose to make a step forward, and acquire the full inventory of native érhùâ instead: actually, in the course of using ‘simple’ [q], many speakers spontaneously will end up articulating [ŋ], due to the effect of coarticulation and articulatory economy, which obviously give preference to what makes pronunciation easier and more natural.

7.2. When the present ‘simplified’ phonetic transcription is used in ordinary handbooks, and in extended texts, no phonemic transcription is needed, as pìnyín certainly suffices to give a rough idea of how the sound system of Mandarin Chi-
nese works. Instead, tones and intonational units shall always be presented in full notation, as any simplification could easily jeopardize mutual intelligibility. Here is an example of connected speech, transcribed according to the simplified pronunciation described above:

7. Phonetic transcriptions in current handbooks

7.3. As we have just demonstrated, introducing phonetics in a common Chinese course is not that complicated, nor confusing, provided the teacher knows how to present the subject in a professional and interesting way. If phonetic symbols are inserted, here and there, just for the sake of it, without a consistent method, they will be certainly ignored, if not frowned upon.

The Chinese teachers must know (at least) the basics of their students' mother tongue, including its phonological and phonetic system, and should be able to distinguish all the relevant vowels and contoids. Sadly, most teachers from China have a certain command of English, but not a scientific knowledge of its pronunciation; rarely do they speak any other foreign language, before going to the country where it is spoken, and working there for a while.

When we started learning Chinese at Rome University, back in 1998, one of the first 'oddities', from the point of view of an Italian native speaker, was to determine the exact value of \( p, t, k \), as opposed to \( b, d, g \). Reading the three couples according to their Italian pronunciation \(-[p, b], [t, d], [k, g]-\) they sounded identical to our
Chinese teacher. He could not help insisting on the ‘aspiration’ factor, for \( p, t, k \)
\(/ph, th, kh/, totally ignoring the fact that \([Ch]\) do not belong to neutral Italian, un-
less one is using emphasis, and \textit{randomly} articulates a \textit{slightly} ‘aspirated’ con
toid, which nevertheless \textit{has no phonemic relevance}. It was a really frustrating situation,
with both parts stuck in a deadlock, and unable to tell what was wrong in the first
place. One can easily figure out what invariably happens, when foreign students try
to ‘refine’ their tones and, even, dare to learn Chinese intonation, and their teacher
has nothing more to offer, but going on and on repeating the sentences contained
in the textbook, ‘hoping for the best’. All this, of course, because no reliable and
scientifically coherent notational system for tones or intonation is available.

7.3.1. The general lack of interest in the potential of phonetics, now apparently
prevailing among most teachers and the authors of language courses, does not
prove that this has always been the case. In fact, as far as Chinese is concerned, it
has to be remembered that one of the first and most prestigious textbooks ever pub-
lished (besides Zhào Yuánrèn’s \textit{Mandarin Primer}) is the well known \textit{Chinese}, by the
Linguaphone Institute in London, whose first edition dates back to almost a cen-
tury ago. This brilliant product is \textit{entirely} transcribed in IPA (also because \textit{pinyin}
was yet to come, in the Thirties), including quite an extensive tonetic notation,
with thick symbols for the tones in \([\$]\), and thin ones for the tones in \([\$]\).

It is worth noticing that the phonetic notation adopted in \textit{Chinese} is surprisingly
very similar to the ‘simplified’ one presented above, notwithstanding some peculiar
choices, mainly due to typographical limitations (and the general ‘sobriety’ of IPA
in those –otherwise glorious– pioneering days). Here is a complete list of the \([C]\) in
the two systems:

\begin{verbatim}
can[b, ph]  off[b, ph]  
can[d, th]  off[d, th]  
can[g, kh]  off[g, kh]  
can[x]  off[x]  
can[s]  off[s]  
can[z]  
\end{verbatim}

\begin{verbatim}
can[dz, tsh]  off[dz, tsh]  
can[dz, tsh]  off[j, ch]  
can[dz, tsh]  off[dz, tjh]  
can[z]  off[\textit{r}]*  
\end{verbatim}

*described as a constrictive: \textit{can}[z]
[V] are treated in a less similar way, yet the two systems remain largely coherent and –so to speak– *mutually intelligible* (fig 7.2). Please observe the following list:

\[
\begin{align*}
\text{can} & [\text{je}, \text{qe}] & \text{off} & [\text{i}\text{e}, \text{y}\text{o}] & \text{can} & [\text{j}\text{a}] & \text{off} & [\text{V}\text{a}] \\
\text{can} & [\text{x}] & \text{off} & [\text{x}]; \text{after } [\text{g}, \text{kh}, \text{x}]: [\text{a}] \\
\text{can} & [\text{w}o\text{\textg}] & \text{off} & [\text{u}\text{o}]; \text{after } [\text{ph}, \text{b}, \text{m}, \text{f}]: [\text{e}] \\
\text{can} & [\text{w}] & \text{off} & [\text{u}] & \text{can} & [\text{z}] & \text{off} & [\text{i}] \\
\text{can} & [(\text{w})\text{e}\text{\textg}, (\text{w})\text{\textg}e] & \text{off} & [(\text{u})\text{e}\text{\textg}, (\text{u})\text{\textg}] \\
\text{can} & [\text{a}\text{\textg}, \text{j}\text{\textg}, (\j)\text{\textg}\text{u}] & \text{off} & [(i)\text{\textg}; \text{o}; \text{u}\text{i}] \\
\text{can} & [\text{i}\text{n}, \text{y}\text{n}, \text{u}\text{n}] & \text{off} & [\text{i}\text{n}, \text{y}\text{n}; \text{i}\text{n}] \\
\text{can} & [\text{j}\text{n}, \text{q}\text{e}\text{n}] & \text{off} & [\text{i}\text{\textg}\text{n}, \text{y}\text{n}] \\
\text{can} & [(\text{w})\text{\textg}a\text{n}, (\j)\text{\textg}a\text{n}, \text{w}\text{\textg}a\text{n}] & \text{off} & [(\text{u})\text{\textg}a\text{n}, (i)\text{\textg}; \text{u}\text{\textg}] \\
\text{can} & [(\text{w})\text{\textg}\text{n}] & \text{off} & [\text{\textg}a\text{n}, \text{u}; \text{n}] \\
\text{can} & [(\j)\text{\textg}\text{n}] & \text{off} & [(i)\text{\textg}] 
\end{align*}
\]

fig 7.2. Vowel diagrams as presented in *Chinese*, by the Linguaphone Institute.

7.3.2. Clearly, this analysis was produced when no highly technological equipments were available to phoneticians. They simply had their ears, their brains, some quite rudimental phonographs, and a limited number of Chinese informants at hand. They employed such ‘antediluvian’ devices as the *artificial palate*, modeled out of fragile resin, or even papier-mâché; the palate had to be covered in chalk powder, or cocoa powder, then inserted in the mouth, and immediately removed, after the relevant phone had been uttered, to determine its articulatory details.

Vocoids were even harder to analyze: a thin metallic chain, often with a small lead ball in the middle, had to be kept on the tongue surface, refraining oneself from spitting it out, or swallowing it; at the same time, the phonetician had to pronounce a vocoid, or a diphthong, as clearly and naturally as a normal human being—in such a tight spot—can do, while an X-ray apparatus would take multiple photo-
graphs of the phonatory act (with a much higher exposure to radiations than in modern X-ray procedures). Nonetheless, those pioneers would do their best, not with the purpose of a mere display of scholarly expertise, but to provide their readers with something that could be really useful to learn languages better.

The idea that IPA should be limited to serious articles, essays and books ‘for insiders only’ (which no-one else would ever read) belongs to more recent generations of phoneticians. On the contrary, the International Phonetic Alphabet was created, primarily, as a practical tool for teaching languages to everyone. A radiant example of that is precisely what we find in the first Linguaphone courses (on 78 rpm records, and in their second editions, on 45 rpm records); unfortunately, the succeeding editions –on audiocassettes, and CD’s– reduced the phonetic sections to one plain and cold page, or completely abolished any IPA symbols..

7.3.3. Given such premises, it is quite hard to understand why the Journal of the Phonetic Association, as recently as in 2003, sanctioned an official description of Mandarin Chinese, by Lee, W-S. & Zee, E., Illustrations of the IPA – Standard Chinese (Beijing), which is embarrassingly inferior to what was available decades before. Just have a look at the ‘official’ vowel diagrams (fig 7.3).

fig 7.3. Vowel diagrams as presented in Journal of the Phonetic Association (2003).

Apart from the questionable collocation of [x] in the first diagram, the other two trapezia mix up true diphthongs, like ‘[ei, ai, au]’ (can[ei, æə, ɑʊ]), and what should be treated not as diphthongs, or even ‘tripthongs’, but as [JIV] sequences, eg ‘[ye, uei, iau]’ (can[ɤe, ʊə, جا]). Regarding ‘[uai, iau]’, we cannot help criticizing the lower ‘angle’, which should indicate ‘[a]’, but –clearly– rather lies in the area of off[ə] (for ‘[uai]’, the angle might be rendered as off[ɑ], too, but nothing lower than that; not even off[ɐ], which would appear less absurd).
The officially sanctioned symbols for Mandarin [C] are not ‘wrong’ in se, but are certainly far from being the ideal choice. For one thing, postalveolar /tsʰ, ts, s, z/ appear as ‘[tʃʰ, tʃ, s, z]’, maybe to avoid that someone would interpret ‘[tsʰ, ts, s, z]’ as the corresponding apicopalatal contoids, used—for instances—in many Dravidian languages (can[tʃʰ, tʃ, s, z]). We shall not repeat here what we have said about the official treatment of /u/ [u, u]. Besides, the authors do not even mention the possibility of some kind of consonantal gradation, nor do they draw any clear line between phonemics and phonetics. Tones and taxotones are treated vaguely and clumsily, while intonation is... not discussed at all. If the members of the International Phonetic Association really accept those four pages as the official description of one of the most important—and most studied—languages in the world, we shall ‘let silence reign’, instead of speaking frankly, and hurting somebody’s feelings.

7.3.4. There is, nonetheless, something that must be said, loud and clear. What is most annoying in many official analyzes is not the use of offIPA in se, but rather its misuse. Undeniably, offIPA is structurally simpler than canIPA: from one point of view, this is a practical advantage, but also a functional limitation. In other words, when one is dealing with just a few dozens of symbols, and is not forced to be precise, a parlous mistake is less likely to occur (hence our harsh criticism of the slovenly treatment reserved to ‘[uai, iau]’ in the JIPA diagrams); canIPA symbols, on the contrary, number in the hundreds, and that obviously requires a good deal of care and perfectionism. However—we note—this greater effort eventually pays off, whenever a transcription must be accurate and detailed, but should not lose its readability and—why not?—even its ‘typographical appeal’.

Since the introduction of IPA to date, AD 2011, many generic analyzes have been realized for most languages. The only reason why someone may be interested in writing another essay on the pronunciation of a language is trying to do something better than it has been so far; possibly, with a better system. Invariably, canIPA works more efficiently than its official ‘competitor’, when the complexity of certain languages (including Chinese) poses so demanding challenges, that excessive simplicity would result in a waste of time. A more flexible system is, of course, harder to use properly, and we do not deny the risks that this aspect involves.

At the same time, we would happily expect the same conscientiousness on the part of those who remain loyal to the old school. If a transcription aims at being
simple, let it be simple... but not sloppy; consequently, if the generic symbols adopted may have numerous values, an author is compelled to pay the utmost attention to what we call **vocograms** and **orograms**: if these are drawn correctly, an otherwise ambiguous transcription regains reliability and, ultimately, its scientific relevance. Whether this was the case with **JIPA’s Standard Chinese**, we will let our readers judge by themselves.

7.3.5. More important than that, we are interested in proving that **offIPA** can be used with appreciable results, provided everything is done to compensate for its transcriptional limitations. First of all, no transcription should ever become an ugly conglomeration of diacritics. Sure, there are some **canIPA** symbols with *three* diacritics, usually representing peculiar articulations for co-articulations, as in [ɨː̃]: nasalized, laterally contracted, and laryngealized. But, in most cases, unitary symbols are available, and no diacritic is necessary at all.

When using **offIPA**, instead, diacritics are inevitable almost everywhere: the point is deciding whether all of them should be printed—or handwritten—at all times, or should they be presented only once, in conjunction with the relevant illustrations. In fig 7.4, we give a practical example: only official symbols are used, apart from **can[़़़़़़़़़़़़़]**, whose value is intuitively self-evident.

**fig 7.4.** Official vowel diagrams of Chinese ‘normalized’ pronunciation (with **offIPA** symbols).

![Diagram](image-url)
7. Phonetic transcriptions in current handbooks

The diacritics help understand which symbol corresponds to which dot (or arrow); but, once the exact value of a certain vocoid has been understood, thanks to the diagrams, the diacritics could be gradually abandoned, until only primary symbols are used. That would work acceptably well, if we stick to a teaching-oriented kind of transcription. But, as we know, the real pronunciation of neutral Chinese implies frequent cases of laryngealization: [Y, η,  centerX, ℹ]; nasalization is quite common, too, eg [Ȳ, Ȳ]. Vowels may be partially devoiced, [Y], but also totally devoiced, [Ȳ]. We might have to indicate more than one peculiarity at the same time, as in [⺈], and –adding partial labialization– [⺈]. And if we were to annotate all the subtleties involving Chinese [C], some symbols could feature even more diacritics!

Now, it is clear that doing the same in offIPA would not be viable, especially in continuous transcriptions. The only option would be that of skipping all basic diacritics, whenever more specific ones have to be inserted. Frankly, that does not seem very satisfying, nor entirely scientific; rather, it proves –once again– that nobody should expect to ride a bike and travel faster than sound: that is what supersonic jets have been invented for! However, a bike still can be a good bike, if we do not try to use it as a jet. The bottom line is: let us keep offIPA true to itself.

7.3.6. In this paragraph, we present an example of how offIPA can express a good deal of phonetic information, after having duly considered its typographical restraints and the purposes for which it was originally invented: learning the pronunciation of foreign languages. As such, we will not attempt to transcribe neutral Chinese, with all its taxophones and other peculiarities, but the slightly simplified pronunciation described in § 7.1.1-2. Where necessary, the exact value assigned to an official symbol will be specified between ‘[]’, by means of diacritics. Let us start from the consonantal inventory:

<table>
<thead>
<tr>
<th>b</th>
<th>p</th>
<th>f</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ɓ]</td>
<td>[ph]</td>
<td>[f]</td>
<td>[m]</td>
</tr>
<tr>
<td>d</td>
<td>t</td>
<td>n</td>
<td>l</td>
</tr>
<tr>
<td>[ɗ]</td>
<td>[th]</td>
<td>[n]</td>
<td>[l]</td>
</tr>
<tr>
<td>z</td>
<td>c</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>[ʐ]</td>
<td>[ʈʃ]</td>
<td>[s]</td>
<td></td>
</tr>
<tr>
<td>j</td>
<td>q</td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>[ʐ]</td>
<td>[ʈʃ]</td>
<td>[ɕ]</td>
<td>[j]</td>
</tr>
<tr>
<td>zh</td>
<td>ch</td>
<td>sh</td>
<td>r</td>
</tr>
<tr>
<td>[ʐ]</td>
<td>[ʈʃ]</td>
<td>[ʃ]</td>
<td>[ʈʂ]</td>
</tr>
<tr>
<td>g</td>
<td>k</td>
<td>h</td>
<td>ng</td>
</tr>
<tr>
<td>[ŋ]</td>
<td>[kh]</td>
<td>[χ]</td>
<td>[n]</td>
</tr>
<tr>
<td>[ŋ]</td>
<td>[w]</td>
<td>[ŋ]</td>
<td>[ŋ]</td>
</tr>
</tbody>
</table>

109
Of course, all the relevant articulatory illustrations (orograms, labiograms and laryngograms) must be provided, as in fig 3.2-7. As far as the actual symbols are concerned, \( \text{off}[C] \) may correspond to \( \text{can}[C, \_\_C] \); this can be positively exploited, after all, by presenting \( \text{off}[C] \) as ‘synthetic’ symbols, capable of expressing a certain free oscillation. Similarly, even if \( \text{off}[h] \) is primarily given as a constrictive, or ‘fricative’, in \( \text{off}[Ch] \) it should be articulated, rather, as an approximant: \( \text{off}[[h\-]] \) (\( \text{can}[h] \)).

For \( /\text{ts}(h), \_\_s, \_\_\_s/ \), the corresponding ‘retroflex’ \( \text{off}[d\_z, ts, s, s] \) are fine, but the relevant orograms should help the students distinguish Mandarin postalveolars from true ‘retroflex’ contoids (\( \text{canIPA: apicopalatals} \)). The notation \( ‘[[z\-]]’ \) dispels any doubt about the exact articulation of \( [s] \).

Since our font contains unitary symbols for all stopstrictives, we obviously use them: \( [ts, d\_s; t\_c, d\_z; ts, d\_z] \). Most Unicode typefaces still have \( ‘[ts, d\_s; t\_c, d\_z; ts, d\_z]’ \), but the general trend, nowadays, is expressing \([KS]\) as \([KS]\) everywhere, as if they were sequences of a stop plus a constrictive, instead of single articulations. Font editing is now a relatively easy and inexpensive task, and some minor enrichments to the typefaces adopted in a textbook should not really pose any difficulty. However, if there is no choice but following the official recommendations, up to the last, the correct renderings should be \( [ts, d\_z; t\_c, d\_z; ts, d\_z] \), avoiding sad things like \( ‘[t\_F, ts]’ \).

Regarding the vocalic inventory, there is little to say, more than what has been clearly explained in § 7.1.2. Please refer to the official trapezia in fig 7.5, and compare them with our \( \text{canIPA} \) vocograms in fig 7.1.
A
a [a] [ä]
ai [ae] [æ] uai [wae] [œ]|œ] ao [ao] [oœ] iao [jao] [œœ]
(u)an [w(an) [ã] ian [jɛn] [ɛ] iu an [ɛn] [ɛ]
(i)ang [j(ian) [ã] uang [w(an) [ã]
E
e [ɛ] [œ] ie [iɛ] [œ] iœ [œ] [œ] ei [ei] [œ]
en [ɛn] er [ɛ] [œ] eng [ɛn] [œ] ueng [w(ɛn) [œ]
I
i [i] i [œ] [œ] ui [œ] [œ] ui [œ] [œ] ui [œ] [œ]
in [ɛn] [œ] ing [ɛn] [œ] ing [ɛn] [œ]
O
o [ɔ] [œ] ou [œ] [œ] uo [wɔ] [œ] uo [wɔ] [œ] o
ong [ɔn] [œ] iong [œ] [œ] iong [œ] [œ]
U
u [u] u [œ] [œ] iu [œ] [œ] iu [œ] [œ]
un [wɔn] iun [w(œn) [œ]

For /çu/, as said, we use [ç], ie off[œ], and the same contoid can be employed in the articulation of érhùã: [Vã]. For transcribing more genuine articulations, those with [œ], the official inventory offers no unambiguous solution. Clearly, we cannot employ off[œ] to render our cant[œ], a postalveolarized prevelar semi-approximant (with lateral contraction, see § 2.3); the only official alternative is off[œ]. In canIPA, the same symbol is –more conveniently– employed for the most typical realization of American English r, a postalveolarized prevelar slightly rounded approximant (with lateral contraction, as well), which we could render as ‘[œœ]’ in offIPA: [œ] is quite advanced, [œœ], and the [œ] diacritic –generically described as implying ‘rhoticity’– apparently is the only passable solution to mark lateral contraction. (Again, here, [œ] is a simple postalveolar approximant: [œ], or [œ].)

However, since off[œ] is usually employed as an alveolar approximant, a notation like ‘[Vœ]’, or ‘[Vœ]’, could lead to legitimate criticism. As a practical solution, ‘[œœ]’ might work well; for can[œ], then, we would have off[œ, œ, œ, œ] &c. Anyway, offIPA always leaves room for ‘personal preferences’! Therefore, when using a very small type size, we suggest ‘[Vœ]’ as less hard to decipher than the tiny ‘rhoticity hooklet’.

111
8. Peking.

*The Peking accent vs the Peking dialect*

8.1. We have carefully considered whether to include analyses of regional pronunciations in a thesis that specifically focuses on *neutral* Chinese pronunciation. That would be the logical continuation of this work, and—in fact— we have already started collecting material for a thorough examination of the most important regional accents, from both the People’s Republic of China and other parts of the vast Chinese-speaking world.

To this end, we have prepared a detailed pronunciation questionnaire (enclosed in a separate appendix, at the end of this thesis). Based on the recordings of several informants and numerous other audio files, we are collecting a rich sound database, which—we hope— will be useful also to other researchers interested in the subject.

For the purposes of this dissertation, nonetheless, we think it is advisable not to venture into some descriptions that may certainly be interesting, but would also require a lot of space, to be presented in full (certainly much more than what would be advisable in a monographic essay, like the present one). Otherwise, we should limit ourselves to scattered considerations, without actually providing accurate transcriptions nor illustrations. Quite evidently, such a reading could be rather disappointing.

However, Peking certainly deserves an exception, given the close relationship between the Common Language and the dialect spoken in the capital. But it is important to point out that we will deal with the *Pekingese accent* of Mandarin Chinese, not the *Peking dialect* proper, as the latter is an independent language, with its phonological, lexical and grammatical peculiarities. Surely, there is no sharp boundary between the *accent* and the *dialect* of Peking, but rather a *continuum* of more or less marked variants, from a kind of almost neutral pronunciation, with very few regionalisms, down to a sort of perfect *diglossia*, where the speaker unconsciously oscillates, even within the same utterance, depending on the context and the linguistic register.
8.2. Peking is a huge city with about twenty million inhabitants, many of whom were not born in the capital. The vast number of immigrants from other parts of China, who clearly do not speak Pekingese natively, as well as the influence of mass media and the increasingly higher literacy rate, have deeply changed the linguistic situation in the area. Many dialects coexist with different degrees of proficiency in speaking pǔtōnghuà. Moving from one part of the city to another, nowadays, is a frequent phenomenon, as social and job mobility has increased dramatically in recent years. Under such circumstances, pronunciation will evolve more quickly, making it harder to define a univocal pattern for describing what a ‘Peking accent’ sounds today.

The old-fashioned pronunciation, consecrated by xiàngshēngr crosstalk masters as Hóu Bāolín and Guō Qîrú,¹ is no longer the only one used by the locals; indeed, what many call ‘Pekingese’ now is a largely stereotypical relic, confined to local comedy and a small number of (very) aged speakers. From a phonemic and phonetic point of view, traditional Pekingese largely coincides with Mandarin. The parapthonic settings, instead, are remarkably different: for example, Hóu Bāolín’s Pekingese-accented pǔtōnghuà sounded high-pitched and crisp, with clearly uttered tones and a rapid flow; this is precisely how proper xiàngshēngr delivery style is supposed to be. By contrast, Hóu’s stooge, Guō Qîrú, would intentionally speak with a less refined –yet understandable– pronunciation, which was decidedly more on the side of the dialect proper, while interacting with Hóu, to convey the theatrical idea of a naïve character.

Video-sharing websites, like YouTube and its Chinese alter ego, Yóukù, provide quite a rich selection of xiàngshēngr performed by Hóu and Guō, which is highly advisable to carefully listen to, in comparison with neutral voices from contemporary Chinese courses audio files. Also the recordings attached to the old Linguaphone Institute Chinese Course offer abundant examples of Peking-accented pronunciation.

Therefore, and contrary to popular belief, the ‘traditional’ pronunciation described above has little to do with the ‘lazy mutter’ most often –but unjustly– associated with speakers from the bútòngr alleys.² This is infamously known for its anarchical use of érhùà (see §2), and the frequent reduction –if not ‘annihilation’!– of constrictive and stop-strictive contoids. Foreigners, as well as many Chinese themselves, would typically find it nearly unintelligible, but also fascinating, and would
even try to imitate it, in order to ‘sound local’. We cannot help discouraging such a
practice: true locals will hardly appreciate a bad imitation of something that they
consider exclusively their own, while people from other parts of China could won-
der why Westerners should waste their time learning how to mispronounce the
Common Language.

8.3. Upon suggestions from our informants, we managed to determine three ex-
amples of what can be honestly targeted as a ‘modern’ Peking accent (again, not
dialect). First of all, Mr Jiäng Wén (1963-), a well known movie star and director;
then, Ms Yang Lán (1968-), a successful entrepreneur and talk show hostess; finally,
Mr Zhäng Bïn (1969-), a sport anchorman from CCTV5.

We chose to select them for three good reasons: (1) all of them are about fifty
years old, that is the most representative age range for an accent (with ‘mature’
people, although not old; then, all things considered, they are still ‘young’ speakers,
but free from temporary phonic habits so typical of real young people); (2) their
pronunciation is recognisably Pekingese, but still within a flawless spoken
pǔtōnghuà, being the classical example of a ‘mediatic pronunciation’ of the capital
city; (3) all of them are nationally well-known, therefore one can easily find plenti-
ful audio and video material on the Web, ranging from plain interviews to the
presentation of nationwide programs.

Anyone interested in gaining an insight into the differences between a (‘clean’
Pekingese) mediatic accent and a proper Pekingese dialect (too marked to be ac-
cepted as pǔtōnghuà) can easily succeed. It is sufficient to compare the pronuncia-
tion of the three above-mentioned people with that of Mr Jiäng Kün (1950), a mas-
ter of xiängshengr, who, on the contrary, habitually speaks using the real Pekingese
dialect (even if it is no longer that of Hóu Bâolín).

Fundamental traits

8.4. Impressionistically, contemporary Peking-accented Mandarin sounds rather
annoyingly ‘buzzing’ in comparison with its neutral counterpart. However, that
does not imply an overall tenser articulation. On the contrary, our recordings,
coupled with articulatory kinaesthesia, suggest a laxer articulation of both vocoids
and contoids, mainly as a result of the same reduction and attenuation phenomena
described in § 1.5.1-3, 3.4.2-3, and 3.7: primarily, /ts(h), s/ [z] (& [j]) and /ts(h), s/ [ç],
in unstressed syllables with ‘zero’ tonemes, not in an intoneme. This is not sur-
prising, and actually happens in more colloquial varieties of English, too, as anyone will easily notice, by comparing the language spoken daily among relatives, friends and colleagues, with that adopted in dubbing foreign movies and documentaries.

The typical buzz accompanying Pekingese Mandarin is rather a consequence of the behaviour of /😜ʊŋ/ and ěrhua. Depending on the speaker and the elocution speed, [ʊŋ] may feature a noticeably stronger lateral contraction than in neutral pronunciation; frequently, it is realised as a postalveolar intense approximant [ɻ], possibly ‘coloured’ by the addition of a certain degree of lateral contraction (representable –if really needed– with ‘[ɻɻ]’); finally, a more marked variant is [ʊŋ] (with a laxer and less advisable alternative, [ʊŋ])

The massive use of ěrhua is perhaps the most stereotypical characteristic of Peking-accented Mandarin. However, so far we have not found two single speakers from Peking who would totally agree on which words should take the ‘extra’ /-ʁ/ (ie one not strictly requested in neutral pronunciation) and which ones should not. If a clear rule existed, we might treat the behaviour of ěrhua in Pekingese pronunciation not simply as a morpho-lexical phenomenon; but, until solid evidence of a certain ‘logic’ is found, we prefer not to reproduce here a boring list of words, as those invariably presented in self-styled ‘serious’ essay. As far as our field of study is concerned, the mere frequency of ěrhua is of little importance, compared to its true phonetic nature, which shows oscillations similar to that just described for /ʃʊŋ/. Therefore (beside neutral or nearly neutral [ŋ]), [ŋ, ŋ] are also possible (including their nasalized counterparts).

Modern Pekingese also features a distinctive realisation of /w/ as [ʊ], a labiodental approximant (or [v], a semi-constrictive), except in /wɔ/: wànwéiwâng ‘World Wide Web’ p[/wɔŋ’wən’wɛwɪŋ], -wɔŋ- (with further variants for /ei/; see below), but tâ wên wǒ ‘he/she asks me’ p[/θaiwɔŋ,wɔŋ(ŋ)], -wɔŋ(ŋ)] (obviously, other stress patterns are possible). Sometimes, but not systematically, partial labialization can occur: [ʊ, V]; yet, [ʊ] can be safely employed in a ‘broad Pekingese’ kind of normalized transcription.

Such variants are gradually gaining ground among youngsters, while being far less frequent, if not virtually absent, among most elders. In their early attempts to learn Chinese, foreign students whose native language (including English) treats /w/ and /v/ as distinct phonemes are easily puzzled by that ‘v-sound that should not be there’. Some of them will actually use their own [v] instead, in search of a more
genuine pronunciation, only to be mildly reprimanded by their teacher, who may reinterpret a fully constrictive [v] as a badly pronounced /l/. Therefore, unless one is able to distinguish [v] from [v] (as many Dutch speakers are, for example), /w/ [w] is the best and only choice for teaching purposes.

Another trend—popular among young Peking speakers— involves articulating /ti, σu/ as narrower diphthongs: [ee, σo] (fig 8.1). In quicker speech and in preintonemes, even monotimbral realizations are possible, [ee, σo] (particularly in /o$/).

As far as stress and intonation are concerned, there are no significant differences with respect to neutral Chinese. However, a relatively large number of common words ending with a marked toneme, /($)$/, feature a ‘zero’ toneme in Peking- accented Chinese: pángxié ‘crab’ [pʰan'⁵xe], wéiqùn ‘apron’ [waɾtʃiŋ, 'weා], nán- gua ‘pumpkin’ [nəŋ'⁵wa]; these coexist with their ‘normal’ realizations, as presented in most dictionaries: pángxiè [pʰan'⁵xe], wéiqùn [waɾtʃiŋ, 'weා], nánguā [nəŋ'⁵wa]. Similarly, post-verbal monosyllabic pronouns wǒ, nǐ, tā are more systematically reduced to /o$/ than in neutral pronunciation (unless emphasis is added): wǒ wèn nǐ ‘I ask you’ [wɔŋŋn, nǐ], gēn tā zōu [tʂan'-ta, tʂoŋŋŋ, -tʂoŋŋŋ].

fig 8.1. Diphthong reductions in Pekingese accent, especially among younger generations.
Notes:

1 Xiàngshèng is a traditional Pekingese comic performance, featuring jokes and funny dialogues arranged in a crosstalk pattern. It requires an exceptional mastering of the language—traditionally, the Peking dialect, rich in puns and popular jargon—and a good deal of dry wit.

Hóu Bâolìn (1917-1993) is still remembered as an outstanding xiàngshèng performer. He was most revered for his creativity and ability to play with language, but also for his high-profile approach to xiàngshèng: he actively promoted the use of pûtônghuà in performing modern crosstalk, both to spread the Common Language throughout China (thanks to radio and—later—tv broadcasting), and to make a quintessentially Pekingese art become popular across the country and beyond.

Always assisted by his lifetime friend and colleague Guô Qîrú (1900-1969), he even dedicated some famous xiàngshèng to the relationship between dialects and what he believed should be the ‘national’ language of all Chinese. Such famous pieces of crosstalk as Pûtônghuà yû fângyán [Common Language and Dialects], and especially Bêijinghuà [The Peking Dialect], remain as valuable testimonies of how the delicate matter of language evolved in China in the past decades, besides being a precious collection of authentic Pekingese pronunciation.

2 The hùtòng are ancient narrow streets, or alleys, most commonly associated with the oldest and most traditional parts of Peking. In that city, hùtòng are formed by several lines of sìbêyùânr, the traditional Pekingese square courtyard residences; many neighbourhoods have been formed by joining one sìbêyùânr to another to form a hùtòng, and then joining one hùtòng to another. The word hùtòng is also used to refer to such neighbourhoods.

Since the mid-20th century, the number of hùtòng has dropped dramatically in Peking, as many of them have been demolished to make way for new roads and buildings. More recently, some hùtòng were designated as protected areas, in an attempt to preserve this aspect of Chinese cultural history.

Usefulness of computer-aided analysis applied to phonetics and tonetics

9.1. As far as our field of study is concerned, and coherently with the ‘natural’ approach that we chose to adopt, no particular equipment is absolutely necessary, apart from a good recording apparatus (whether a classic tape-recorder, or a more modern digital device) and a high quality headset. Our ears and brain will do the rest. That does not mean that speech analysing softwares –such as the easily available *Praat*— are useless to ‘natural’ phoneticians; just the opposite is true, but the user must be aware of the technical limitations that all similar computer-aided research protocols –so far– have been unable to overcome.

The bottom line is that instrumental data will bear fruit only when supported by, and submitted to, the phonetician’s own scientific reasoning, which –in turn– should be mediated by the attentive listening of well selected audio samples. It must be remembered that machines will record and analyse *everything*, without previously filtering out what a human being would spontaneously discard, for being fortuitous and therefore not representative, such as background noises, or voice distortions due to contingent factors. Yet, such a remarkable degree of (unnecessary) comprehensiveness, paradoxically, fails to provide a full account of what a diligent phonetician really needs.

As a blatant demonstration of our point, we invite anyone to download the latest version of *Praat* from Boersma and Weenink’s webpage. We have chosen *Praat* precisely because it is readily available and is free of charge, in addition to being constantly updated both by the authors and by a large community of users worldwide. Once installed, load the software and select an appropriate audio file, for example a .wav clip (note that .mp3’s are not supported) from any Chinese course at hand. A wide range of choices are available to the user, including the possibility of ‘drawing’ vectorial images for successive editing (the three illustrations enclosed hereby precisely originate as *Praat*-generated .eps vectors).²
9.2. With little effort, it is easy to determine where the vocal chords are vibrat-
ing and how intensely –within certain limits– making it possible to tell fully voiced
phones from fully voiceless ones. But, as soon as one tries to analyse mixed phona-
tion contoids, or partially devoiced vocoids (among others), what appears on the
screen is often very elusive and ambiguous.

In a similar way, constrictive articulations are not hard to find, as they produce a
unique, densely ‘indented’ contour. However, it is practically impossible to deter-
mine the exact place of articulation of a given contoid, apart from very loose dis-
tinctions, such as ‘front’ (like [s]) vs ‘back’ (like [x]). Nor is it possible to univocally
distinguish ‘slit’ and ‘grooved’ contoids, or certain constrictives and their corre-
sponding approximants, which even untrained persons may be able to recognise on
the spot. Interestingly, Praat appears to be less reliable precisely where the human
ear does its job with more ease!

The only reasonable way to detect the characteristics of a given contoid through
Praat and other ‘Praat-esque’ softwares, would be to compare the tracings and dia-
grams extractable from the informant’s recording with those from ‘specimen’ audio
files, ie audio samples of (generally single) phones whose articulation is certain and
‘certified’. Typically, a ‘specimen’ should feature the voice of a neutral speaker,
preferably a professionally trained one, like a dubbing voice actor; or, the phoneti-
cian’s own voice, provided the articulation is performed after careful phono-
kinaesthesia.

Even in that case, however, instrumental phoneticians cannot confidently assert
‘this phone is...’, but only ‘based on what I see on the screen, this phone looks more
or less similar to...’. At the end of the day, not an exceptional performance, unless
the ultimate evidence is provided: an X-ray of the phonatory organs while the
phone is being uttered, or a sequence of photographs –if not an X-ray film!– for dy-
namic articulations, eq to distinguish [iː] and [jː].

Now, it goes without saying that such a purely instrumental approach would be
‘mathematically correct’, for being detached from the human element –which is
subjective and relatively fallible– but nothing proves that it could also be resolutive,
in terms of scientific reliability; or, of any practical use, for example to those who
develop an interest in phonetics to improve their language learning skills. Even
more importantly, the instrumental approach is simply not viable, when a phonetic
survey is to be conducted on a large basis of informants, and an expensive and
cumbersome X-ray machine... happens not to be near at hand! Not to mention the non-secondary fact that the amount of radiation to be administered to each informant would be more than enough to cause severe consequences.

9.3. With regard to vocoids, *Praat* clearly shows their *quantity* (: duration), obviously in terms of seconds, or fractions of a second, according to a mental category readily understandable by human beings: time. But, as far as vocalic *quality* is concerned, the software can simply report a number of ‘representative frequency values’, or *formants*, expressed in hertz (: cycles per second), which —on the contrary— humans cannot measure naturally.\(^3\) It is the coexistence of the right formants *as a whole* that allows the hearer to disambiguate vowels; or, in the case of language teachers, to tell whether their students have learned how to pronounce the vowel sounds which do not exist in their mother tongue(s), or continue to employ their native phones as a substitute. In this respect, mere frequency values are of little use.

From the point of view of a ‘natural’ phonetician who wants to prepare the *canIPA* vocogram, or the *offIPA* trapezium, of a given language (or just of a certain speaker), computer-measured formants may be worth the effort of being collected and analysed only if the same microphone and the same recording device are invariably used, without altering any of the settings, and always under the same environmental circumstances (humidity, pressure, sound insulation degree \(\&c\)). Considering how variable the situation can be, when a phonetician records different informants, or when they produce an audio file by themselves, and then send it to the researcher, anyone can easily imagine why we are far less enthusiast than ‘acoustic’ phoneticians are about dealing with those ‘cryptical’ Hz values.

It must be remembered that the same vocoid, *e.g.* [a], can be produced by different formant combinations, depending on the speaker’s sex, age and physical characteristics, too; this happens even within the same speaker, depending on whether one is whispering, or roaring out loud in a stentorian way, since also voice intensity plays a certain role. Finally, of course, vocalic formants are –slightly– influenced by surrounding contoids, too, which makes the scenario even more intricate. Quite enough to look at the ‘prodigies’ of computer-aided phonetics with a bit of prudent scepticism.

Therefore, before recording a pronunciation test, every informant should provide some basic data to ‘calibrate’ the software, so that it is possible to treat the formant
values relevant to a single utterance not as an absolute, mathematically measurable
item unrelated one another, but as something comparable with a general ‘grid’,
where both a five-year-old boy’s [a] and that pronounced by his grandmother can
be collocated according to a universal criterion.

Such preliminary data should include, at least, a perfect rendition of the four ex-
treme vocoids [i, æ, ə, ñ], and the corresponding rounded ones, [y, ɶ, ɶ, u], whose
formant values will constitute the fundamental ranges which all other vocoids shall
be compared to. Unfortunately, apart from being a tedious and time-consuming
process, that would require the uncommon blessing of finding a non-phonetician
informant who perfectly knows what to do and how, to articulate unfamiliar vo-
coids, including the theoretical ones not attested in any living language, e.g [m, ɶ].
The inevitable question is, should we really bother going through so much trouble
to meet a machine’s needs, when our ears and brains can do the job for us just fine?
As long as voice-analysing softwares lack the ability to automatically accommodate
to different voices without the tiresome pre-setting procedures described above, the
‘natural’ approach will remain the winning choice.

9.4. In our view, a software like Praat really offers phoneticians a helping hand
only in studying tonality, which is very intuitively represented by a series of irregu-
lar segments, or –when these happen to be tightly connected– by a solid line. The
pitch contours relevant to complete sentences, if considered as a whole, can be a
good hint to the intonational patterns of non-tonal languages; while the segments
corresponding to single syllables reveal how taxotones behave in tonal languages.

In Chinese, the two layers of intonation and tonemicity are tightly intertwined,
which makes it difficult to interpret Praat-generated tracings, if one does not know
what to look for in the first place, and how. But, no matter how careful and rig-
ourous one can be, yet there is little hope to truly understand Chinese intonation
merely by means of instrumental measurements; even more so, for those who actu-
ally have to learn it, to improve their pronunciation of the language. To provide a
clear example of our point, we have extracted the tonal contour of six short sen-
tences submitted to a reliable informant from Peking, and recorded in a nearly
sound-proof environment (of course, the recordings can be made available to any-
one who may be interested in double-checking what we report hereby).

The original tracings produced by Praat were somehow too ‘coarse’ to be used as
such, as in some parts no tonal contour appeared at all (mainly due to an excessive
laryngealization), while in others the same contour was clearly incoherent with what could be heard from the earphones. Consequently, fig 9.1.2-3 must be considered as the result of a patient and meticulous editing process, whose length and error-proneness in se should be exasperating enough to warn wannabe instrumental phoneticians against any form of ‘easy optimism’.

All four fundamental intonation units, /, ; , ?/, are duly represented. We intentionally selected sentences where the first toneme abounds, given that the slightly deformations produced by intonation are far easier to observe, when the toneme’s theoretical profile is a horizontal one. Please, be noted that tonality is indicated by the solid thick line, while the dotted line represents intensity (which must be considered, due to its influence on tonality, at least at the perceptive level).

The grey areas suggest the corresponding intonational bands; the criterion we followed is the ‘grapho-arithmetic mean’ of the actual height of the various taxotones, compared to the height that we would expect to find in /;/, when the tonal band is clearly expanded, and the four tonemes are realised with their ‘fullest’ taxotones, ie those typically used at the beginning of all Chinese courses.

9.5. The six sentences are presented hereby in pinyin romanization first, then in cmIPA transcription, and are followed by their English translation:

*Mā dànxīn jià bān ma?
/¿ ma `tan`sin`esja`paN`ma?/
[¿ ma `tan`ciN`deja`paN`ma.] ‘Is Mom worried about working overtime?’

*Mā hēn dānxīn jià-bān.
/¿ ma `h`Nn`tan`sin`esja`paN./
[ ma `haN `tan`ciN`deja`paN.] ‘Mom is very worried about working overtime.’

*Tā shì nì māma ma?
/¿ thaN`suN ,n`maNmaNma?/
[tha`suN ,n`maNmaNma意义上.] ‘Is she your mother?’

*Tā bù shì wǒ māma?
/`tu`puN`suN ,woNmaNmaNma.
[tha `puN`suN ,woNmaNma意义上.] ‘She is not my mother.’
Ni kàn nà-bèn shú?
/¿nîkhan ˌna.ˈpənˀˈʂu/.
[¿nîkhaŋ ‘na.ˈbənˀˈʂu]
‘Which book do you read?’

Ni ài kàn nà-bèn shú jiù kàn nà-bèn shú.
/niˈraːkhan ˌna.ˈpənˀˈʂu/ ˌtsjouˈkhan ˌna.ˈpənˀˈʂu./
[niˈraːkhaŋ ‘na.ˈbənˀˈʂu] ˈtsjoukhaŋ ‘na.ˈbənˀˈʂu’
‘Read whatever book you like.’

fig 9.1.

fig 9.2.

fig 9.3.
Notes:

1. *Praat* (also the Dutch word for ‘talk’) is a free software program for the analysis of speech in phonetics. It has been designed and continuously developed by Paul Boersma and David Weenink of the University of Amsterdam. It can run on a wide range of operating systems, including various Unix versions, Mac OSX and Microsoft Windows (95, 98, NT4, ME, 2000, XP, Vista). The program also supports speech synthesis, including articulatory synthesis.

2. In this case, we used *Praat* 5.2.03.

3. For a detailed explanation of how formants are generated, and to produce the distinct ‘timbre’ of the various vocoids, please refer to a reliable book or, with extreme caution, to a similar website.
10. Conclusions

We will not try to summarise in a few lines the considerable amount of data presented in this work, as every aspect of Mandarin Chinese pronunciation has been described in details, but we also did our best to avoid any unnecessary rigmarole. The purpose of these conclusions is highlighting the main reasons why—in our view—this analysis is worth reading.

Phonology in general

Though pinyin is, all in all, a fair representation of the essential components of pǔtōnghuà pronunciation, yet it cannot be taken as a true phonemic transcription. It is surprising, then, that the matter is not addressed effectively by most of the analyses that we have consulted. Many linguists feel content with a simplified kind of phonetic transcription, whose IPA symbols are so generic that—apparently—a separate phonemic transcription would be almost superfluous, also because the -emic and the -etic layers are often mixed up.

That is why we regarded it important to draw a clear line between them, and devised a thoroughly meditated phonotonemic scheme. In it, we made it clear that the ‘aspirated consonants’ are simply /Ch/ sequences, and there is no real need for an extra ‘/ʨ, tɕh, ɕ/’ phonemic set, as /ʦ, tsh, s/ can safely include the taxophones represented by j, q, x in pinyin. Instead, the ‘mythological apical vowels’ have been re-conducted to their genuinely vocalic nature, by adopting ‘/u/’, irrespective of the numerous taxophonic variants that may occur. Where possible, a realistic symbol has been employed, instead of a generic one (as the IPA practice recommends): thus, /e, ɛ, ɻ/, instead of something like ‘/ei, ai, au, ou/’.

Vowels and consonants

We chose to include eight phonemes in the vocalic system of Mandarin Chinese, as that makes it easier and more natural to manage the numerous taxophones required in neutral pronunciation. Diphthongs have been analysed with full realism, rejecting traditional renditions like ‘/ei, ai, au, ou/’ even at the phonemic level.
We are particularly proud of §2, where the ěrhùà phenomenon is treated in
great depth; once again, avoiding superficiality and abstractness. Instead, a set of
laterally contracted vocoids has been introduced, finally explaining why ěrhùà does
not sound exactly like the ‘American r’, as many phoneticians still seem to think.

Our treatment of Mandarin contoids accepts no compromise, and clearly indi-
cates—for the very first time, as far as we know—three stress-related consonantal de-
grees for stops, stop-strictives and constrictives. This three-level system can imply a
partial or total voicing of the contoid; and, for /Ch/, a difference in ‘aspiration’
strength. Other linguists have had a similar intuition, too, but sadly never imple-
mented it in a truly systematic way, as we have done in this work.

**Tonemes, tones and intonation**

Instead of focusing only on consonants and vowels, we devised an effective nota-
tional system, to indicate tonemes and intonemes as well. At the phonotonetic
level, our transcription is even capable of indicating stress, stress-related taxotones
and important intonation details at the same time, without jeopardising readability
and typographical elegance. It may look complex, of course, but that is simply be-
cause what it represents is complex. However, it is not complicated, nor visually
chaotic, since each symbol of our inventory univocally corresponds to a single ele-
ment that is indispensable for a neutral pronunciation, and has been carefully de-
dsigned not to interfere with other symbols.

**Suggested pronunciation**

Most people learning Chinese are not phoneticians, and do not actually need—or
want—to acquire a genuinely neutral pronunciation. The average student would be
quite happy with a reasonably simplified scheme, but not as generic as that generally
presented in Chinese handbooks. For this specific purpose, we have provided a sug-
gested pronunciation which, while renouncing some phonetic subtleties, yet features
no peculiarities that an average Chinese native speaker would consider regional, or
even typical of a foreign accent.

As a final statement, we hope that this work can provide a useful contribution for
a better understanding of the pronunciation of Mandarin Chinese, whose impor-
tance as a global language will consistently increase in the near future.
Bibliography


Canepari, L. (1983) *Phonetic Notation·La notazione fonetica*. Venezia: Cafoscarina (with 2 enclosed audiocassettes; almost canIPA).


Delattre, P. et alii (1951) *vwajel sëtetik a ðø formāt e vwajel kardinal*, in ‘Le Maître Phonétique’ (IPA).


