

Hanyu Pinyin Pronunciation Guide

Stephen M. Hou
Version: 11/19/2010

This guide is intended to teach native English speakers how to pronounce words written in Hanyu Pinyin, the official Romanization system for Standard Mandarin Chinese used in mainland China, Taiwan, and Singapore. Keep in mind that Romanization is not intended approximate English pronunciation – it is simply a mapping from Latin letters to the sounds of Mandarin. This is also true for other languages that use the Latin alphabet; for example, the letter “j” is pronounced differently in English, French, Spanish, and German.

Unless otherwise stated, the English example words are pronounced with a Standard American accent.

The Structure of Syllables

Instead of the inventory of sounds being fundamentally categorized into consonants and vowels, all the possible syllables in Standard Mandarin consist of at most three components: an initial, a final, and a tone. The initial, which is optional, is the first consonant group of the syllable. For example, in the syllable “ban”, “b-” is the initial; in the syllable “zhang”, “zh-” is the initial. The final is the remainder of the syllable. Seven initial consonants may stand alone as syllables without finals: zh-, ch-, sh-, r-, z-, c- and s-. In these cases, the letter “-i” is added as a “placeholder” to indicate an empty final. Finally, the tone is always present.

Some finals contain one of three *glides*: -i-, -u-, and -ü-. These are vowels in the middle of syllable that are followed by more vowels in the same syllable. It is important to note that such syllables are written as one syllable and thus are to be pronounced as one syllable. For example, English speakers are tempted to pronounce “liang” as the two-syllable combination “lee-ang”. However, the “-i-” sound is of very short duration and “glides” into the rest of the final.

When the glide sounds -i-, -u-, and -ü- have no initial (null initial), they are written as y-, w-, and yu-, respectively. For example, when initials are removed from “liang”, “tuan” and “lü”, they become “yang”, “wan” and “yu”, respectively.

Tones

Mandarin has four tones, plus a neutral (“fifth”) tone:

1st tone: Steady high (like singing).

2nd tone: Rises from mid-level to high (like “What?”).

3rd tone: Dips from mid-low to low; if at the end of a sentence or before a pause, it is followed by a rise.

4th tone: Sharp fall from high to low (like “Stop!”).

Tones are written as diacritical marks above a vowel in the syllable. For example, the four tones of the syllable “ma” are: (1) mā, (2) má, (3) mǎ, (4) mà. The neutral tone is written without any marks at all (ma). Alternatively, numbers can be written after the Roman letters: ma1, ma2, ma3, ma4, and ma5.

Mandarin has one general tone sandhi rule: When there are two 3rd tones in a row, the first one becomes 2nd tone.

Since no specific Chinese words are used in this guide, tones are not indicated in any syllable in the remainder of this document. The focus is on learning how to pronounce the initials and finals.

Initial Consonants

Zhuyin	Pinyin	Pronunciation
ㄅ	b-	Like b as in “boy”, but unvoiced. Like p as in “spin”.
ㄆ	p-	Like p as in “pin”, but with more aspiration.
ㄇ	m-	Like m as in “mom”.
ㄈ	f-	Like f as in “for”.
ㄉ	d-	Like d as in “door”, but unvoiced. Like t as in “stick”.
ㄊ	t-	Like t as in “Tom”, but with more aspiration.
ㄋ	n-	Like n as in “nun”.
ㄌ	l-	Like l as in “light”.
ㄍ	g-	Like g as in “gas”, but unvoiced.
ㄎ	k-	Like k as in “kite”, but with more aspiration. Like k as in “skin”.
ㄏ	h-	Like h as in “hall”. Northern speakers tend to have a rasp, like the ch in “Chanukah” and “Bach”.
ㄐ	j-	Similar to j as in “jeans”, but with wide lips (smile!), tongue behind lower front teeth, and unvoiced. Like the Korean ㅈ. Definitely <i>not</i> like the s in “pleasure” or “Asia”. “Beijing” is commonly mispronounced in the US media.
ㄑ	q-	Similar to ch as in “cheese”, but with wide lips (smile!), tongue behind lower front teeth, and with more aspiration.
ㄒ	x-	Similar to sh as in “sheep”, but with wide lips (smile!) and tongue behind lower front teeth. Like the Polish ś.
ㄓ	zh-	Retroflexed version of pinyin z- . Position the tongue and lips as if you were going to say “err...”, but try to say the English ds as in “beds” (and without voicing) instead. The resulting sound is the pinyin zh- . Sounds <i>vaguely</i> like the d in “drive”, but unvoiced. Like the Polish cz .
ㄔ	ch-	Retroflexed version of pinyin c- . Position the tongue and lips as if you were going to say “err...”, but try to say the English ts as in “bits” instead. The resulting sound is the pinyin ch- . This is different from the pinyin q- , which is actually more similar to the English ch than the pinyin ch- is.
ㄕ	sh-	Retroflexed version of s- . Position the tongue and lips as if you were going to say “err...”, but try to say the English s instead. The resulting sound is the pinyin sh- . This is different from the pinyin x- , which is actually more similar to the English sh than the pinyin sh- is. “Shanghai” is commonly mispronounced in the US media. Like the Polish sz and the Swedish and Norwegian rs .
ㄖ	r-	Position the tongue and mouth like you are going to say “err...”, but try to say the English l as in “light” instead. Unlike l , however, the tip of the tongue does not make contact with anything, but the sides of the tongue touch the roof of the mouth. The resulting sound is the pinyin r- . This is probably the most difficult sound in Standard Mandarin for native English speakers to pronounce correctly. Like the Polish ż and rz .
ㄗ	z-	Like ds as in “beds”, but unvoiced. Like the German z .
ㄘ	c-	Like ts as in “bits”, but with more aspiration. Like the Polish c .
ㄙ	s-	Like s as in “sand”.

More on Initial Consonants

The last ten consonants can be organized into a table:

	Palatals	Dental Sibilants	Retroflexes
Unaspirated Affricate	ㄐ / j-	ㄗ / z-	ㄓ / zh-
Aspirated Affricate	ㄑ / q-	ㄘ / c-	ㄔ / ch-
Unvoiced Fricative	ㄒ / x-	ㄝ / s-	ㄕ / sh-
Voiced Fricative			ㄖ / r-

Comparing the columns:

- **Palatals:** The tip of the tongue is dropped to a place behind the lower front teeth and the blade of the tongue is brought up to contact the palate (roof of mouth).
- **Dental sibilants:** The tip of the tongue is held against the back of the gum of the lower front teeth (alveolar ridge).
- **Retroflexes:** Position the tongue and mouth like you are going to say “err...”, but try to say the dental sibilants instead.

Comparing the rows:

- **Aspirated:** Force air through the mouth.
- **Voiced:** Use the vocal cords.
- **Affricate:** Sounds that begin as stops (closed airway) but end as fricatives.
- **Fricative:** Force air through narrow channel (like the s in “sit”).

Generally, the lips are wider (smile!) than they are when similar English sounds are pronounced.

Note: Many people outside Beijing (especially in Southern China and Taiwan) cannot pronounce the retroflex sounds correctly. They will frequently pronounce zh-, ch-, and sh- as z-, c-, and s-, respectively; r- tends to be pronounced in a variety of ways, including as pinyin l- and English z. For example, in Taiwan, the city of Shanghai is frequently pronounced as “Sanghai”.

Note: The use of the letter combinations q, x, zh, ch, sh, z, and c may seem quite unnatural for English speakers. After all, why not use “ts” and “ds” to instead of “c” and “z”, respectively? The Hanyu Pinyin system is widely praised for following two simple principles for initial consonants:

- Adding an “h” after z, c, and s makes them retroflexed.
- Each initial consonant is represented by exactly one letter, or by a combination of a letter with the retroflex indicator “h”.

Thus, the pinyin assignment of Latin letters to Mandarin initial consonants is quite optimal, given those constraints. Other Romanization systems, such as Wade-Giles system commonly used in the West before the 1980s and still used in Taiwan for personal and place names, use the same letter combinations to represent different sounds; the next letter determines which sound is meant. For example, in the Wade-Giles system, “ch-” (without an apostrophe) represents the pinyin j- when it is followed by -i- or -ü- (Wade-Giles “chiang” = Hanyu Pinyin “jiang”), and represents the pinyin zh- otherwise (Wade-Giles “chou” = Hanyu Pinyin “zhou”).

Note: Finally, Mandarin (and all other modern Chinese dialects) lacks consonant clusters that are frequent in English and other European languages, such as tr-, fr-, pl-, sn-, sm-, sp-, sc-, st-, etc. However, linguists believe such clusters were present in Old Chinese, the language of Confucius (ca. 500 BC).

Finals

Glideless Finals			
Zhuyin	Pinyin	Form with Null Initial	Pronunciation
N/A	-i	N/A	Buzzed continuation of the initials zh-, ch-, sh-, r-, z-, c-, and s-. For all other initials, see the “Group i Finals” table below.
ㄚ	-a	a	Like a as in “father”.
ㄛ	-o	o	Like “ awe ” or “ all ”.
ㄜ	-e	e	Like “ uh ” or the oo in “ look ”, but with wide lips (smile!).
ㄝ	-e	ê	Like e as in “wet”. Sound without glide and initial consonant only exists as interjection. Otherwise, it requires a glide.
ㄞ	-ai	ai	Like “ eye ”.
ㄟ	-ei	ei	Like ay as in “day”.
ㄠ	-ao	ao	Similar to ow as in “ now ”, but the starting vowel sounds more like the a in “father”.
ㄡ	-ou	ou	Like o as in “no”.
ㄢ	-an	an	Like the an in “ pan ” spoken with a British accent. In other words, the a is pronounced like the a in “ ax ”. It is <i>not</i> “ahn”. “ Mulan ” is commonly mispronounced in the US media.
ㄣ	-en	en	Like en as in “ ten ”.
ㄤ	-ang	ang	Like ang as in “ angst ”, or “ah-ng”. It is <i>not</i> like ang as in “sang”.
ㄥ	-eng	eng	Northerners tend to pronounce it like ung as in “ rung ”. Southerners and Taiwanese tend to pronounce it as pinyin -en, or as “ch-ng” (similar to pinyin -en, but ending with a nasal -ng sound instead of -n), and pronounce the syllable “feng” like “fong” (long o). It is <i>not</i> like ang as in “sang”.
ㄜㄝ	-er	er	Like er as in “ better ”. Northerners tend to pronounce it like the ar in “ car ” (as if they were pirates or something). Southerners and Taiwanese tend to pronounce it similar to pinyin -e (ㄝ).

Group i (y-) Finals			
Zhuyin	Pinyin	Form with Null Initial	Pronunciation
丨	-i	yi	Like ee as in “ see ” when appearing after b-, p-, m-, d-, t-, n-, l-, j-, q-, and x-. For all other initials, see the “Glideless Finals” table above.
丨ㄚ	-ia	ya	i + a. Like the German “ ja ”.
丨ㄛ	-io	yo	i + o. Sound only exists as interjection.
丨ㄜ	-ie	ye	i + ê. Like ye as in “ yet ”. It is <i>not</i> like “yay”. “Xie xie” (thank you) is frequently mispronounced by English speakers as “shay shay” (both initial and final are incorrect).
丨ㄞ	-iai	yai	i + ai. Like yi as in “ yikes ”. Uncommon.
丨ㄠ	-iao	yao	i + ao. Like “ yow ”, but the middle vowel sounds more like the a in “father”.
丨ㄡ	-iu	you	Contraction of i + ou. Like the English yo , as in “ yo , what’s up?”
丨ㄢ	-ian	yan	i + an. Like “ yen ”. It is <i>not</i> “yahn”.
丨ㄣ	-in	yin	i + n. Like een as in “ teen ”.
丨ㄤ	-iang	yang	i + ang. Like ang as in “ angst ”, but starts with a “y”.
丨ㄥ	-ing	ying	i + eng. Northerners tend to pronounce it “ee-uhng”. Southerners and Taiwanese tend to pronounce it as pinyin -in, or as “ee-ng” (similar to pinyin -in, but ending with a nasal -ng sound instead of -n).

Finals (cont.)

Group u (w-) Finals			
Zhuyin	Pinyin	Form with Null Initial	Pronunciation
ㄨ	-u	wu	Like oo as in “moose”. However, when the initial is j-, q-, or x-, see the table below for Group ü Finals.
ㄨㄚ	-ua	wa	u + a. Like wa as in “swan”.
ㄨㄛ	-uo	wo	u + o. Like “ wall ”.
ㄨㄞ	-uai	wai	u + ai. Like “ why ”.
ㄨㄝ	-ui	wei	Contraction of u + ei. Like “ way ”.
ㄨㄢ	-uan	wan	u + an. Like “ wax ”, but with “x” replaced by “n”. However, when the initial is j-, q-, or x-, see the table below for Group ü Finals.
ㄨㄣ	-un	wen	Contraction of u + en. Like “ when ”. However, when the initial is j-, q-, or x-, see the table below for Group ü Finals.
ㄨㄤ	-uang	wang	u + ang. The a is like the a as in “father”.
ㄨㄥ	-ong	weng	u + eng. Like “ song ” in British English (long “o”).

Group ü (yu-) Finals			
Zhuyin	Pinyin	Form with Null Initial	Pronunciation
ㄩ	-ü	yu	Position the lips as if to say “oo” but position the tongue as if to say “ee”. Like u as in French “ lune ” and the German ü . When the initial is j-, q-, or x-, we can simply write “u” without the umlaut dots because the syllables consisting of those consonants and the “u” sound without the dots are nonexistent in Standard Mandarin. Thus, “ju”, “qu”, and “xu” actually represent “jü”, “qü”, and “xü”, respectively.
ㄩㄝ	-ue	yue	ü + e. Both vowels are distinctly pronounced, but as one syllable. It is <i>not</i> “you” nor “you-ay”.
ㄩㄢ	-uan	yuan	ü + an. Keep in mind that “yuan” is pronounced as one syllable. It is <i>not</i> “yu-an”. However, when the initial is anything other than j-, q-, or x-, see the table above for Group u Finals.
ㄩㄣ	-un	yun	ü + n. However, when the initial is anything other than j-, q-, or x-, see the table above for Group u Finals.
ㄩㄥ	-iong	yong	ü + ong. In Mainland China, the -ü- here is pronounced more like a (pinyin) “-i-” (hence the spelling change).

Diphthongs

The diphthongs are much more fused in Chinese than in English. For example, the -ai final in the “hai” syllable of “Shanghai” is said with far less transition from the “a” to “i”, as compared to the similar sound in English, the “ye” in “bye” or the “ie” in “lie”. Thus, when an English-speaker says “Shanghai”, the “ai” sounds exaggerated to a native Mandarin speaker.

Non-Ambiguity of Finals

e: Two different sounds (ㄛ and ㄝ) are represented by the pinyin letter “e”. How do we know which sound is meant? (Recall: ㄛ is like “**uh**” or the **oo** in “**look**”, but with wide lips; ㄝ is like **e** as in “**wet**”).

- ㄛ: Only occurs immediately after (an) initial consonant(s) or by itself (*i.e.* no glide).
- ㄝ: Almost always requires a glide vowel immediately before it. The only exception is when this final appears by itself (and it does so only as an interjection), in which case it is written as “ê” to distinguish it from the syllable “e”, which is “ㄛ”.

i: When -i follows zh-, ch-, sh-, r-, z-, c-, or s-, it is simply a buzzed continuation of the initial consonant. For all other initial consonants, it is pronounced like **ee** in “**see**”. Pronouncing the -i as “ee” after the six aforementioned consonants produces syllables that do not exist in Standard Mandarin. For example, the syllable “see” does not exist, even though both “s” and “ee” sounds exist.

ü: We can usually leave out the umlaut dots over the -ü because most initials can be followed by either -u- or -ü-, but not both (see the Hanyu Pinyin Syllable Table), so it is unambiguous as to which sound is represented. The only syllables in Standard Mandarin for which both -u- and -ü- versions exist are:

- nū vs. nu
- lū vs. lu

As the umlaut dots are inconvenient or impossible to type on English keyboards, you may sometimes see “lū” written as “luu” or “lv” (the letter v is unused in Hanyu Pinyin). The syllable “nū” is similarly sometimes written as “nuu” or “nv”.

Contractions: Three finals are contractions when written in pinyin:

- -iu = i + ou
- -ui = u + ei
- -un = u + en

It is important to remember that these are only *written* shortcuts, not *spoken* ones. Learners of pinyin commonly mispronounce the -iu as “ee-oo” (like the **ew** in “**Jew**”), the -ui as “oo-ee” (like the French *oui*), and the -un as “uhn” (like the **un** in “**run**”). The contractions are unambiguous because the “ee-oo”, “oo-ee” and “uhn” vowel combinations do not exist in Standard Mandarin.

Consonants in the Finals

The only consonants that may come at the end of Mandarin syllables are “-n”, “-ng” and “-r” (and for “-r”, the only syllable possible is “er”). This contrasts with other Chinese dialects, like Cantonese and Taiwanese, which have plenty of syllables that end in “-k”, “-m”, “-p” and “-t”. Combine this with the earlier note about the lack of consonant clusters and you’ll conclude that the number of possible syllables in Mandarin Chinese is highly constrained (~400 not counting tones, ~1,300 counting tones), especially when compared to English (~8,000).