1 Tone Basics

Roadmap: Why do we believe that there is such a thing as TONE? What are the articulatory apparatuses relevant for tone production? How is tone notated or transcribed? What are the acoustic correlates of tone? What can tones do in human languages?

1.1 INVOKING TONE

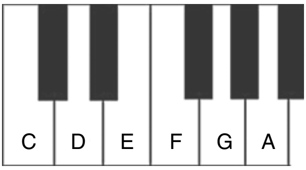
As a phonological construct, the notion of tone and its representation is motivated by puzzles: phenomena that do not square with expectations and hence require explanation (Uyechi & Wee 2009:xiii–xxiii). Thus, instead of starting out assuming that there is such a thing as tone in phonology and then going on to explaining what it is and how different people have understood it, it would be more objective to look at phenomena that might motivate such a postulation. For example, in Japanese, a Japonic language, *hashi-ga* can mean 'chopsticks', 'bridge', or 'edge'. There is nothing interesting here, since all languages have HOMOPHONES. For example, in English *run* [rʌn] can refer to an action of alternating one's legs forward quickly, a loose tread on clothing or a place where chickens are kept. In the case of English *run*, the word's meaning cannot be clarified without context. However, this is not the case with *hashi-ga* in Japanese. Every time a speaker says *hashi-ga*, all other speakers seem able to identify which meaning is intended without ambiguity! So, Japanese is not like English (not that we are Anglocentric), and there must be something more than just consonants and vowels that allows Japanese speakers to figure out which *hashi-ga* has been uttered.

This is a puzzle, and there can be different explanations for it; some of us might point to PITCH differences in each of the different *hashi-ga*. In describing Standard Tokyo Japanese, Haraguchi (1999) notes that if *ha* has a higher pitch than both *shi* and *ga*, then *hashi-ga* means 'chopsticks'. If *shi* has the higher pitch, then it means 'bridge' and if *shi* and *ga* are both higher than *ha*, then it means 'edge' With this observation, one could perhaps hypothesize that pitch is meaningfully contrastive in Japanese. If tone is the phonological correlate of what we perceive acoustically as pitch, then the study of pitch patterns in Japanese would be a study of phonological tone.

However, this is not the only possible hypothesis. One could equally hypothesize that the differences in Japanese are due to the locus of STRESS (i.e. the syllable chosen for emphasis or ACCENT), so that the pitch differences observed by Haraguchi are secondary effects. Perhaps, unlike stress in English, Japanese "stress" does not accompany any difference in vowel quality (in English *REcord* and *reCORD* have different vowels for *cord* depending on stress). Naturally, a creative linguist might offer other insights and end up with very different conclusions. Different hypotheses providing explanations of phenomena are competing theories that can be assessed for their various merits, ultimately enlightening us to the nature of the phenomena observed.

Let's try a different language spoken by, according to the website Ethnologue (citing the China 2000 census), approximately 850 million people. In Standard Chinese, a Han language, the syllable [wei] can mean 'magnificence', 'surround', 'tail' or 'feed'. It is not that [wei] is homophonic and has one of these four meanings. For speakers of Standard Chinese, [wei] 'magnificence' is homophonic with [wei] 'small'; and [wei] 'feed' is homophonic with [wei] 'position', but [wei] 'magnificence' and [wei] 'feed' are clearly distinct. There is therefore something here, other than the consonants and vowels, that distinguishes the syllables. When uttered normally (not whispering, for example), what is most distinctive about 'magnificence', 'surround', 'tail' or 'feed' seems again to be related to pitch. Unlike the Japanese example, it is harder to appeal to stress in different syllables in the case of Standard Chinese, as the case presented involves monosyllabic words. To be specific, the pitch profiles of each of the four [wei]s in Standard Chinese are different. This is often hard for non-speakers to perceive at first, but the pitch profiles can be understood in the same way as musical notes are concatenated to make a tune. As an approximation, consider musical notes that can be easily located on a piano keyboard, as in Figure 1.1.

A note G would be very high compared to a note C. Concatenating D and G in legato produces a tune that has a rising profile similar to the pitch profile for [wei] 'surround'. The sequence D⁀C⁀F yields a dipping pitch profile for [wei] 'tail' and G⁀C 'to feed'. Singing, using a slide whistle, sliding the finger board on the violin or playing



Standard Chinese [wei]

Meaning	Musical approximation	Tune
magnificence	G	High flat
to surround	D⁀G	Rising
tail	D⁀C⁀F	Dipping
to feed	G⁀C	Falling

Figure 1.1 Musical notes as an approximation of the Standard Chinese syllable [wei]

the trombone would produce good approximations of the pitch profiles.

From the examples of Japanese and Standard Chinese above, it should be clear that something in addition to consonants and vowels is needed. That something, it seems, is related to pitch. Because it is hard to call stress the pitch used in examples such as those seen in Standard Chinese, the designated appellation is tone.

1.2 PITCH AND VOCAL FOLD VIBRATION

Pitch is what we perceive when our minds register through our ears periodic changes in air pressure. The articulatory correlate of what is perceived as pitch in speech and singing is vocal fold vibration. (For more on acoustic correlates, see Section 1.5.)

Vocal folds are sometimes called vocal cords, but in reality, they are not cords but twin infoldings of mucous membrane stretched across the larynx. Adult males have thicker and longer folds (between 1.75 cm and 2.5 cm) than women (between 1.25 cm and 1.75 cm), which accounts for the difference in pitch ranges between men and women. The vocal folds are located inside the larynx, and they can be made to vibrate at different frequencies by movement of the cartilages

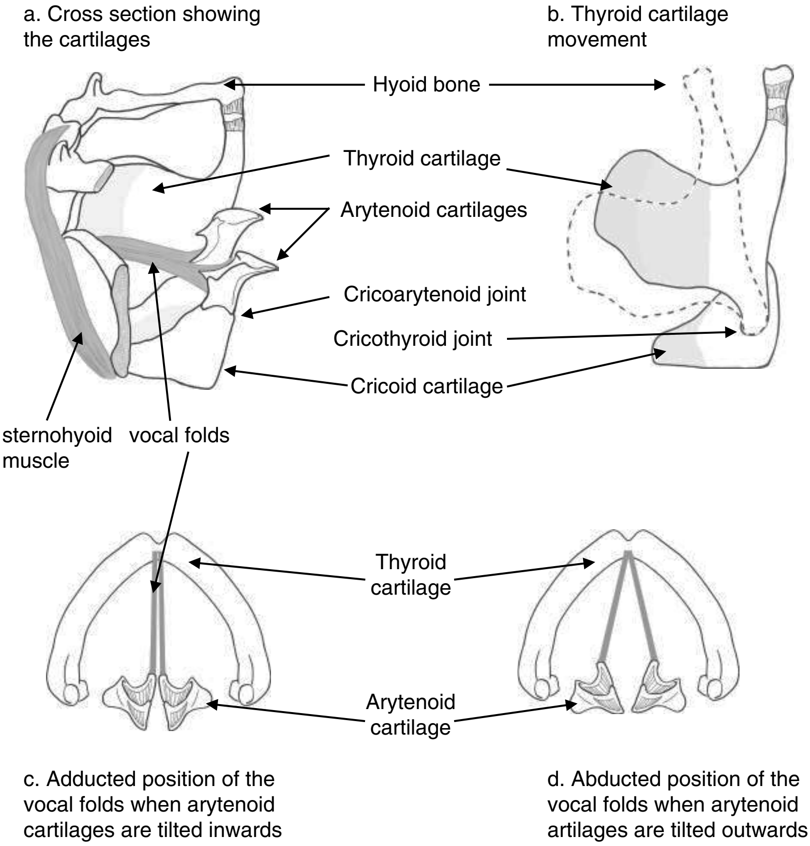


Figure 1.2 The larynx
Images by Winnie H.Y. Cheung. See also Ohala (1978).

(strung to muscles) that tighten, loosen or even lengthen the vocal folds, illustrated in Figure 1.2.

Figure 1.2a provides a cross section showing how the relevant cartilages are joined. The cricothyroid muscles turn the arytenoid cartilages so that the vocal folds are adducted or abducted. By tighter or looser adduction (Figures 1.2c, 1.2d), one can get rises and falls in pitch. However, if you try to lower your pitch further, you will soon feel a different muscle being pulled vertically. That is the sternohyoid muscle (Sagart, Halle, Boysson-Bardies & Arabia-Guidet 1986), which can draw the thyroid cartilage downward (Figure 1.2b), thereby lengthening the vocal folds.

Understanding how we manipulate the vocal folds allows us to make the inference that, in articulatory terms, mid flat tones are easiest to produce, and high tones require stronger vocal fold adduction, thus taking more effort. Low tones require even more effort, since the sternohyoid has to pull downward while the vocal folds maintain adduction. Contoured tones, such as falling and rising tones, would also require movement of the cartilages and are therefore harder to produce than level tones.

We can manipulate our vocal folds to produce different PHONATION, including the falsetto, which is what is done when creating the voice for Mickey Mouse. The vocal folds can also be configured to produce speech that sounds breathy, as when trying to sound sexually alluring in a Hollywood movie. Creaky noises and whispering are also the result of vocal fold manipulation (see Laver 1980).

1.3 NOTATION BASED ON PITCH

If pitch is the chosen line of approach toward understanding tone, a notation system for transcribing pitch and pitch profiles would be handy. Indeed, that is what most studies use for tone transcription, although it shall become clear later that fundamental frequency (F0) might be only one aspect of tone. For effective transcription, a description of pitch shape and height is necessary. The International Phonetic Alphabet (IPA) chart distinguishes between five different levels using accent marks, and it combines these marks to indicate contoured pitch profiles. This system is often used in tone studies of African languages. East Asian linguists prefer to use Chao's (1930) tone letter/number system, where the numbers 1 through 5 indicate five different pitch levels from lowest to highest. (The same numbers run in reverse for works describing South American tone languages, so be careful.) Where numbers are used for tone notation, this book always uses ascending numerical values to correspond to pitch height, akin to Chao's system. Uppercase letters are also sometimes used.

Table 1.1 offers some of the common notations. It does not exhaust the possibilities of combinations of the five pitch values to create different kinds of pitch profiles, but it should be adequate to demonstrate what one can do with any chosen transcription system.

To draw again on Standard Chinese, the four distinctions shown using the syllable [wei] are commonly given as [wei^{55}] 'magnificence', [wei^{25}] 'surround', [wei^{214}] 'tail' and [wei^{51}] 'feed'. Following Chao's

Table 1.1 *Notation systems for pitch profiles/tones*

	IPA	Numbers (1930)	Symbols (1930)	South American	Latin letters
Very high	a̋	5	˥	1	H
High	á	4	˦	2	
Mid	ā	3	˧	3	M
Low	à	2	˨	4	
Very low	ȁ	1	˩	5	L
Falling	â	51, 53, 31, . . .	˥˩, ˥˧, ˧˩, . . .	13, 35, 15, . . .	F
Rising	ǎ	13, 35, 15, . . .	˩˧, ˧˥, ˩˥, . . .	51, 53, 31, . . .	R
Dipping	ãa	313, 424, . . .	˧˩˧, ˦˨˦, . . .	353, 131, . . .	D, HLH, HMH, . . .
Peaking	a᷈	353, 131, . . .	˧˥˧, ˩˧˩, . . .	313, 424, . . .	MHM, LHL, LML
High fall	a̋à	53, 54, 42	˥˧, ˥˦, ˦˨	31, 21, 32	HM
Low fall	āȁ	31, 21, 32	˧˩, ˨˩, ˧˨	53, 54, 42	ML
High rise	āa̋	35, 45	˧˥, ˦˥	31, 21	MH
High peaking	āa̋ā	353, 454, 343, . . .	˧˥˧, ˦˥˦, ˧˦˧, . . .	313, 212, 323, . . .	MHM

Source for Numbers (1930) and Symbols (1930): Chao (1930)

(1930) number system, the superscript numbers describe the pitch profile or TONE VALUE of each of the four syllables. It is also possible to write them as [wéí, wèí, wèǐ, wéì], respectively. Note that, even for level tones, it is conventional for two numbers to be provided, so as to indicate the beginning and end points. Tone notation has not always been as straightforward as presented here. A glimpse of earlier efforts can be found in Section 1 of Chapter 3.

1.4 TONE AS CONTRASTIVE

Many languages use pitch profiles contrastively in the same way languages contrast different segments, such as consonants or vowels. If such pitch profiles are called tones, then just as one can provide an inventory of consonants and vowels for languages such as English, one can also provide tone inventories. Consider for example Thai (a Tai-Kadai language with 20 million speakers), Yoruba (a Defoid language with 19 million speakers), Wàpã (also spelled Wapan, a Junkun-Mbembe-Wurbo language with 100 thousand speakers) and Igala (a Yoruboid language with 800 thousand speakers). These are presented in Table 1.2.

Table 1.2 *Tones in Thai, Yoruba, Wàpã and Igala*

a. Tones in Thai

Tone \ Syllable	**[na]**	**[khaa]**	**[mai]**
High	mother's younger sibling	to do business in	wood
Mid	paddy field	to be lodged in	mile
Low	(a nickname)	a kind of aromatic root	new
Falling	face	servant	no/not
Rising	thick	leg	silk

(Hudak 2009)

b. Tones in Yoruba

Tone \ Syllable	**[ra]**	**[bu]**	**[bɔ]**
High	to disappear	to insult	to drop
Mid	to rob	to mildew	to worship
Low	to buy	to break off	to come

(Awobuluyi 1978:148; Akinlabi & Liberman 2000:33; Peng 2013:346)

c. Tones in Wàpã

ak	wi	
|	|	
Mid	High	'knife'
Mid	Mid	'millstone'
Mid	Low	'chicken'
Low	Low	'gourd'

d. Tones in Igala

a	wo	
|	|	
High	High	'guinea fowl'
High	Mid	'an increase'
High	Low	'hole (in a tree)'
Low	High	'a slap'
Low	Mid	'a comb'
Low	Low	'star'

(Welmers 1973:116)

Table 1.3 *Tones for grammatical category contrasts in Cantonese*

Tone \ Syllable	**[kan]** **'space'**	**[ji]** **'clothes'**	**[wɔ]** **'harmony'**
[55]	noun	noun	
[33]	verb	verb	verb
[21]			noun/adjective

Table 1.4 *Tones for case in Somali*

Nominative	**Vocative**	**Genitive**	**Absolutive**	
rag	-	rág	rág	'males'
orgi	órgi	orgí	órgi	'billy goat'
hooyooyin	hóoyooyin	hooyooyín	hooyoóyin	'mothers'
xaas	-	xaás	xáas	'family'

(from Yip 2002:140; see also Hyman 1981, Banti 1988)

In Thai, we can see that syllables can contrast for any of five different pitch profiles, giving an inventory of five tones. Yoruba has an inventory of three. Although exact values in the range of 1 to 5 are not provided for Thai or Yoruba, the labels do adequately reflect the tone contrasts. Wàpã and Igala offer another way of presenting tone contrasts, where pitch information is associated to each syllable in a word so that the profile for the entire word is discernible from the constituent syllables.

In the above examples, different tones yield different lexical words. However, differences in tone can also be grammatical. Tone may, in Cantonese, a Yue language, signal differences in grammatical category and, in Somali, differences in grammatical CASE. Cantonese is spoken by more than 72 million people, mostly in Hong Kong, Guangzhou and a substantial population of the ethnic Chinese diaspora around the world (especially in Chinatowns). Standard Cantonese has six different tones: [55], [35], [33], [21], [13] and [22] (Linguistic Society of Hong Kong 2002). Three of the tones exhibit noun–verb contrasts, as shown in Tables 1.3. and 1.4.

In Somali, a Cushitic language, tones may indicate grammatical case. In the words in Table 1.4, the vowels carrying the accent mark (e.g. á, ó, í) are articulated with a high MELODY. As can be seen, words in the NOMINATIVE case do not carry any marking (i.e. are unmarked) for high melodies. With VOCATIVES, the initial vowel would get a high melody, with GENITIVES the final vowel and with

ABSOLUTIVES the PENULTIMATE (i.e. second to last) vowel. In Yoruba, an H tone intervenes between subject and verb and docks itself to the subject, e.g. *ọ̄**mọ̄*** 'child' + *lọ̄* 'went' ➔ *ọ̄**mọ́*** *lọ̄* 'the child went' (Akinlabi & Liberman 2000). In Chimwiini, a Bantu language, the position of the high tone makes a distinction in indicating person, e.g. *jile:* ***n̪amá*** 'you [singular] ate meat' contrasts with *jile:* ***n̪áma*** 's/he ate meat' (Kisseberth & Abasheikh 2011). For a succinct and comprehensive discussion of grammatical tone, see Hyman and Leben (in press).

Since the languages above show that pitch profiles associated to syllables allow distinction of meanings and syntactic categories, it is reasonable to generalize the phenomenon as Statement 1.1.

Statement 1.1 Different pitch profiles can be distinctive in different ways across different languages.

Is Statement 1.1 surprising? That depends on what role one expects pitch to play in a language. The syllable [wei] in English means *way* (or *whey* or other homonyms) regardless of its pitch profile, but this is not so in Standard Chinese. However, in English, the differences in the pitch profiles may signal different syntactic constructions; a falling melody is normally associated with declarative sentences (e.g. *Way. This is how we do it* or *Whey. That's what Miss Muffet wants for breakfast*) while a rising melody is associated with interrogatives (e.g. *Way? This is how we do it?* or *Whey? You eat that?*). For the linguist who wishes to exclude English from the study of tone, the appeal is normally to the relevance of pitch for lexical distinction. Two commonly accepted definitions that do that are given in Statement 1.2.

Statement 1.2 Commonly accepted definitions of 'tone language'

a. A language is a tone language if the pitch of a word can change the meaning of the word (Yip 2002:1).
b. A tone language is one in which an indication of pitch enters into the lexical realization of at least some morphemes (Hyman 2001, echoing Welmers 1959, 1973: 80).

Both Statements 1.2a and 1.2b would exclude languages including English, Malay (a Malayo-Polynesian language) and German (West Germanic) from tone studies, relegating the effects of pitch in distinguishing interrogatives and declaratives to what is often called intonation. The definition in Statement 1.2a might further exclude the grammatical effects of pitch; recall case marking in Somali or lexical alternation in Cantonese, depending on how loosely 'meaning' is defined. For many of the preceding decades, the study of tone has been understood along the lines of Statement 1.2, but change in such a trend has certainly become observable a few years into the second millennium. One major reason is that independent studies on lexical tonal distinctions and intonation have begun to converge. The same tone features used for describing lexical tones can be used for describing intonation by distinguishing the different morphosyntactic domains and how they interface with phonology, hence there is no need to assume that tone and intonation are tonologically different. In any case, the definitions in Statement 1.2 appeal either to semantics or to morphology as being the key to distinguishing whether pitch is tonal, rather than to a phonological distinction between tone and non-tone (such as stress). (See Gussenhoven 2004:47 for a succinct discussion of the difficulties in maintaining the distinction between tone and intonation.)

Definitions such as Statement 1.2 are also challenged by phonetic studies that reveal tone and stress as employing similar acoustic correlates. The similarities between them will become clear when one probes into the acoustic correlates of tone in Section 1.5.

1.5 PROBING INTO THE ACOUSTIC CORRELATES OF TONE

In the preceding sections, tone was invoked by virtue of perceived pitch differences between syllables which, in articulatory terms, correspond to differences in vocal fold vibration. Vocal fold vibration produces periodic changes in air pressure as air is pushed out from the lungs. In acoustic phonetics, the periodic changes in air pressure that result from vocal fold vibration are called FUNDAMENTAL FREQUENCY (F0). Like all waves, F0 is measured in HERTZ (Hz, i.e. cycles per second, also abbreviated as cps). Human ears can perceive sounds within the frequency range from about 20 Hz up to about 20,000 Hz.

Since pitch is a subjective perception, it is hard to study directly and difficult to measure. In contrast, F0 is quantifiable and measurable, so using the acoustic correlate to inform us of articulatory activity in the vocal folds and auditory perception of pitch is particularly attractive. Recall for example the tones in Standard Chinese, repeated below with the transcribed tone values and a few other examples.

The list in Table 1.5 is a tiny sample of Standard Chinese syllables. In fact, when tones are taken into account, Standard Chinese has about 1,300 distinct syllables, each a possible word in its own right. To study the F0 properties of Standard Chinese tones, one can collect recordings