1 Complexities of Tone

In the theoretical study of linguistics, tone was not given much attention until rather late in modern times. The IPA chart, for example, included tone and pitch marks under 'additional notes' for the first time only in 1932. Then, as is still the case now, tones were described in terms of melodic profiles such as high, low, rising, falling, and so on. At the simplest level of understanding, these different melodic profiles corresponded to contrasts in meanings, exemplified in Tables 1 and 2.

In some languages, such as Wàpã and Igala, the melody applies across syllables (see Tables 3 and 4).

While such descriptions are phonetically faithful, the phonological complexities that are presented by tones can only be apprehended through a much more sophisticated model. This Element shall endeavour to demonstrate these complexities.

Humanity had, of course, been aware of pitch modulations in spoken languages before the twentieth century. With regard to ancient scholarship, the earliest mentions were Indian, Greek, and Chinese. In India, Pāṇini's *Aṣṭādhyāyī* (circa fourth century BCE) expounded on the Vedic accents *udātta* 'raised, high pitch', *anudātta* 'not raised, low pitch', and *svarita* 'sounded, falling pitch'. In classical Greek, the acute, grave, macron, and circumflex accent marks presumably indicated melodic shapes, as they relate to such musical terms as *tonos* 'tone, lit. to stretch (the strings of musical instruments)', *okseia* 'sharp, high', and *bareia* 'low'. From a facsimile of the thirteenth-century copy of the Periplus of Pseudo-Scylax, dating back to the mid-fourth century BCE, we know that Dionysus Thrax (second–first century BCE) described these accents in musical terms. However, given that classical Greek and Sanskrit are not known to involve tone contrasts at the level of words, it remains uncertain whether these are diacritics for musical chanting. The Romans probably inherited the Greek symbols in the form of the neume, a type of diacritic mark found in Latin liturgical texts to denote proper intoning.

Modern interpretations of historical texts suggest that the earliest known Chinese scholarship that makes references to linguistic tone comes from the fifth century CE. Also, traditional Chinese scholarship would include the field of *xiaoxue* 'small learning'; this covers *wenzixue* 'orthographic studies', which deals with Chinese characters, *shengyunxue* 'phonology', and *xunguxue* 'exegesis', all of which are philological in nature and in the service of apprehending archaic documents.

Yip (2002: 1) suggests that by some estimates, 60–70 per cent of the world's languages may be tonal. For Yip, a tone language is one in which 'the pitch of a word can change the meaning of the word'. The examples in Tables 1–4, of Thai, Yoruba, Wàpã, and Igala, have illustrated this lexical function of tone. In addition,

Table 1 Tones in Thai (Hudak 2009)

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'

Table 2 Tones in Yoruba
(Awobuluyi 1978: 148; Akinlabi and Liberman 2000: 33; Peng 2013: 346)

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'

Table 3 Tones in Wàpã
(Welmers 1973: 116)

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

Table 4 Tones in Igala (Welmers 1973: 116)

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'

Table 5 Verb conjugation by tone in Iau (Bateman 1990: 35–6; Hyman and Leben 2020)

Verb	**Gloss**	**Inflectional meaning**
bá	‘came’	Totality of actional punctual
bā	‘has come’	Resultative durative
bá̋	‘might come’	Totality of action incompletive
ba᷅	‘came to get’	Resultative punctual
bâ	‘came to endpoint’	Telic punctual
ba᷇	‘still not at endpoint’	Telic incompletive
ba᷆	‘come (process)’	Totality of action durative
bâ¯	‘sticking, attached to’	Telic durative

Table 6 Tones for case in Somali (from Yip 2002: 140, see also Hyman 1981, Banti 1988)

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

there are languages where differences in tone can signal grammatical function. An excellent example comes from Iau, a Lakes Plain language from West Papua, Indonesia (Bateman 1990: 35–6, cited in Hyman and Leben 2020), where tone signals verbal inflection (Table 5).

In Somali, tone signals differences in grammatical case.

In Table 6, the vowels carrying the accent mark (e.g., á, ó, í) are articulated with a high tone. As can be seen, words in the nominative case do not carry any marking (i.e., they 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 vowel. Tones can also indicate grammatical person. In Chimwiini, this is done by the position of the high tone, for example *jile: ṉa**má*** ‘you sg. ate meat’ contrasts with *jile: **ṉá**ma* ‘s/he ate meat’ (Kisseberth and Abasheikh 2011). For a succinct and comprehensive treatment of grammatical tone, see Hyman and Leben (2020).

Other than signalling lexical contrasts and grammatical markings, tone also appears to signal prosodic differences. In languages where prosody is described in terms of stress, the main acoustic correlate of stress is often fundamental frequency (F0). Thus, in English *uni<u>ver</u>sity*, the stressed syllable *-ver-* is normally articulated with a higher F0 than the other syllables. In perception experiments, syllables with higher F0 are more likely to be perceived as stressed than other parameters such as duration or intensity (Bolinger 1958, Fry 1958, Beckman 1986). In fact, Goldsmith (1978) suggested that stress and intonation may be treated with tone features (see also Gussenhoven 2004: 49–70). Depending on one's theoretical assumptions, languages like Japanese use placement of the high tone to distinguish meaning, which may seem an uncanny resemblance to stress placement in English (consider *reFUSE* 'to decline' with *REfuse* 'trash'). Perhaps for this reason, Japanese has often been described as pitch accented (McCawley 1978 may be the first to offer a theoretical treatment; de Lacy 2002 points out that pitch accents in many languages interact with stress).

In Table 7, the L and H stand for High and Low respectively. In Japanese, L is not part of the equation – any syllable not given H is L by default. Except for the initial syllable, all syllables preceding H would also be articulated as H. Thus, in the example in Table 7, H is assigned to the first syllable in *háshi-ga* 'chopstick', to the second in *hashí-ga* 'bridge', and the third in *hashí-gá* 'edge'. *Ga* is a nominative case marker. The properties of the Japanese-type pitch-accent system are also seen in languages that have not been conventionally recognized as tonal languages, such as Hong Kong English. There is some evidence that syllables in Hong Kong English are lexically specified for tones, similar to Japanese (Wee 2016a). This property may have been employed as a literary device in Hong Kong's English poetry (Wee 2016b).

In addition to all these, one must be aware that tones are not stable. The same syllable-morpheme may have a different tone depending on context (Shih 1986, Chen 2000, Zhang 2014, Wee 2019, among others). For example, in Hainanese, low-tone syllables become high when followed by another syllable (of any tone). In the same context, syllables with a falling tone become low (see Table 8).

Tone instability comes in many forms, and should be treated differently. In cases where tones alternate in collocation with others, it is often called tone

Table 7 Japanese pitch accent

Word	Melodic profile	Gloss
háshi-ga	H-L-L	'chopsticks'
hashí-ga	L-H-L	'bridge'
hashí-gá	L-H-H	'edge'

Table 8 Tone instability in Hainanese
(Yun 1987: 14–23, Wee 2020)

In isolation	When followed by another syllable
toiʔ 'wash' | low	toiʔ biʔ 'wash the rice' | | high low
	toiʔ ɦue 'wash the flowers' | | high high
mak 'eye(s)' | falling	mak tai 'in front of eyes/presently' | | low mid
ɦiʔ 'ear' | falling	ɦiʔ ʔuat 'ear pick' | | low falling

sandhi. There are also cases of tone deletion (e.g., Standard Chinese, Chao 1948, 1968), and of movement (e.g., in Chizigula, see Section 4.1).

This section has outlined that complexities of tone come from a number of factors. Firstly, they interface with morphosyntactic entities. At the word level, tone yields lexical contrasts. They can also be applied to other aspects of grammar to signal grammatical case and intonation. Secondly, tones may signal prosodic differences. This is an area that has recently caught the attention of many tonologists. Thirdly, tone is unstable and can change forms and positions. Tones can be inserted, deleted, and spread (Goldsmith 1976, Gussenhoven 2004, Wee 2019, Hyman and Leben 2020). Subsequent sections will deal with these topics in turn all of which lead us to the need for rather complex structures in the representation of tone.

2 The Physical Dimensions of Tone

The reader will notice that in any discussion of tone, labels like 'high', 'low', 'rising', 'falling', 'dipping', 'peaking', and so on, are really just spatial metaphors to describe the perceived melodic shapes. The same kind of metaphor is used for musical melodies, corresponding to the psychoacoustic concept of 'pitch'. By connecting linguistic tone to pitch, one looks in the direction of physical properties such as fundamental frequency (F0) and the vocal folds which produce F0 in speech. In reality, the physics of tone will involve intensity, duration, and other acoustic properties as well, in relation to the human articulatory systems. Through an overview of the physical dimensions of tone, this section hopes to reveal some of the complexities that are less often studied.

2.1 Pitch Perception

It is worth remembering that before recording technologies were conveniently available, fieldwork investigations relied on pitch perception for the description of tone, drawing analogies with music. For example, Low (1828) laid out Thai tones according to a system based on a treble clef, and Pallegoix (1854) marked each Thai tone as a literal melody in Western musical notation. The most notable of such works might be Chao's (1930) system of tone letters, where the numbers 1–5 produce a scale ranging from very low to very high. Any tone contour can thus be captured by a series of such tone values, say [51] is a full falling tone and [35] is a mid rising tone. Level tones can be represented with two identical numbers, for example, [55] would be a very high flat tone. Chao explicitly acknowledged inspiration from music, pointing out that one could imagine [1] to [5] as analogous to the musical interval known as the perfect fifth (e.g., C to G, or D to A). However, Chao added that a linguistic tone may not align itself so squarely with any particular music interval, and recommended that [1] simply represents the lowest pitch level in normal speech and [5] the highest. Chao's foresight is borne out by later studies, which show that the interval spans do indeed vary widely across speakers in different languages, as seen in Table 9.

As technology has become more affordable, field investigations can now be readily substantiated by recordings and tone descriptions informed by detailed acoustic measurements (such as Zhang and Liu 2011, 2016, among others) (more in Section 2.3). At the same time, analysts usually use a combination of approximate pitch perception, logarithmic-adjusted and normalized F0 profile measurements, and phonological contrast in their description of tones, instead of relying exclusively on acoustic details. In any case, field data published from longer ago would rely more on the ear of the linguist than on instrumental measurement and analysis. To the extent that the descriptions lend themselves to phonological understanding, ears and machinery should both be brought to bear on the subject. After all, the perceived tone is not necessarily a direct mapping of F0, and the perception of pitch can be affected by a listener's native phonological system.

2.2 Physiology

On account of the importance of F0, tonal production relies largely on the control of the vocal folds. The vocal folds are connected at one end to the thyroid cartilage and at the other to the arytenoid cartilages. Adduction (closing)

Cambridge University Press & Assessment
978-1-009-07806-1 — Complexity in the Phonology of Tone
Lian-Hee Wee , Mingxing Li
Excerpt
More Information

Table 9 Pitch ranges of speakers in twenty Chinese languages (Wee 2017)

Language	Gender	F0 range of lowest pitch (Hz)	F0 range of highest pitch (Hz)	High-Low quotient (median values)
Túnxī 屯溪 (Hui)	F	135–152	245	1.71
Méixiàn 梅縣 (Hakka)	F	144–149	267–278	1.86
Nánníng 南寧 (Yue)	F	120	180–268	1.87
Taipei 台北 (S. Min)	F	92–173	253	1.91
Fúzhōu 福州 (N. Min)	F	116–142	252	1.95
Hǎikǒu 海口 (S. Min)	M	78–92	160–176	1.98
Wēnzhōu 溫州 (Wu)	M	149–152	294–303	1.98
Hong Kong 香港 (Yue)	F	149	282–325	2.04
Táoyuán 桃園 (Mandarin)	M	75–89	163–173	2.05
Jiàn’ōu 建甌 (N. Min)	F	135–145	301–302	2.15
Sūzhōu 蘇州 (Wu)	F	129–161	289–350	2.20
Shàntóu 汕頭 (S. Min)	F	143–150	316–379	2.37
Shèxiàn 歙縣 (Hui)	F	134–150	338	2.38
Xiāngtán 湘潭 (Xiang)	F	145–167	371	2.38
Xiàmén 廈門 (S. Min)	M	75	181	2.41

Table 9 (cont.)

Language	Gender	F0 range of lowest pitch (Hz)	F0 range of highest pitch (Hz)	High-Low quotient (median values)
Hángzhōu 杭州 (Wu)	M	76–131	256	2.47
Shànghǎi 上海 (Wu)	M	75–79	193	2.51
Nánchāng 南昌 (Gan)	M	75–82	187–213	2.55
Guǎngzhōu 廣州 (Yue)	M	75–80	200–218	2.70
Chángshā 長沙 (Xiang)	M	75–113	290	3.09

Music interval	Corresponding High-Low quotient*	Musical interval	Corresponding High-Low quotient
Major second	1.122462	Minor second	1.059463
Major third	1.259921	Minor third	1.189207
Perfect fifth	1.498307	Perfect fourth	1.334839
Major sixth	1.681792	Minor sixth	1.587401
Major seventh	1.887748	Minor seventh	1.781797
Octave	2		

* Quotients for music intervals based on twelve-tone equal temperament, which uses $2^{n/12}$, an irrational number, as a multiplier.

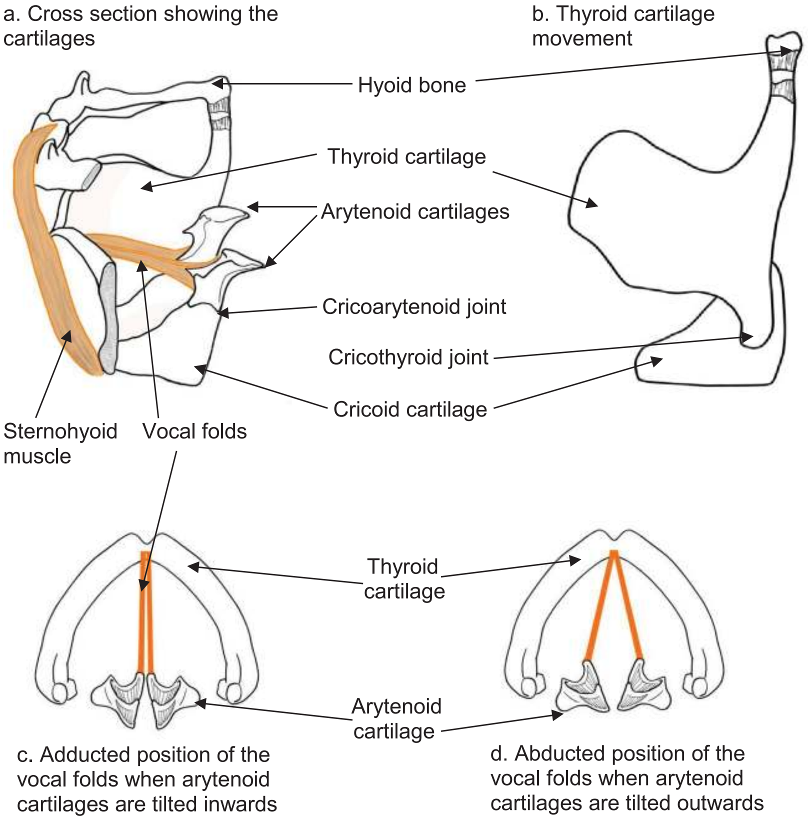

Figure 1 The larynx (see also Erickson 1976, Ohala 1978, Erickson *et al.* 1995)
Images by Winnie H. Y. Cheung

and abduction (opening) are controlled by the tilting of the arytenoid cartilages inwards or outwards (see Figure 1).

In the adducted position, the vocal folds allow pressure to build up as air is pushed out of the lungs. Vibration results through repeated bouts of air release through the vocal folds to produce F0. Through control of the arytenoid cartilages, the folds may be made tighter to create higher frequencies or looser to create lower frequencies. However, for very low frequencies, the sternohyoid muscle (Figure 1a) will have to pull the thyroid cartilage downwards to lengthen the vocal folds (Ohala and Hirose 1970, Ohala 1972, although see Niimi, Horiguchi, and Kobayashi 1991 for how the sternohyoid can be used to raise F0 in singing as well, albeit not in speech). In articulatory terms, this means that mid tones are easiest to produce, followed by high tones, then low tones. This concept will become very relevant when one looks at how many levels of tone-height contrast are found in different languages (see Section 3).

Table 10 Tone inventories of Jinxing Bai exhibiting phonation

Normally phonated	Specially phonated
[55]	[66] tensed
[33]	[44] tensed
[31]	[42] glottalized
[35] (loanwords only)	[21] tense/creaky, fricated

The fundamental frequency may not be the only parameter relevant for tone (Andruski and Ratliff 2000, Kuo, Rosen, and Faulkner 2008, Brunelle 2009, Garellek *et al*. 2013). Vietnamese, for example, requires glottalization to contrast between two tones that are otherwise quite similar in terms of their perceived pitch: the *sắc* 'sharp' tone and the *ngã* 'tumbling' tone are both described as mid rising, but the latter has an intervening glottal (Nguyen and Edmondson 1997: 8 and Pham 2003: 57). Similarly, Hudak (2008) observed that certain tones are glottalized across Tai languages. Wang's (2012: 41–5) fieldwork uncovered four different dialects of Bai in which voice quality was an important aspect of their tonal inventory. One of these is reproduced in Table 10.

The specific voicing qualities of these Bai cases require further study. In any case, issues of voice quality in relation to tone can be tricky. Even with appeal to minimal pairs, one may still find it hard to ascertain whether specifics of phonation should be attributed to differences in the vocalic segments or to tone (Andruski and Ratliff 2000, Garellek *et al*. 2013). This difficulty is particularly evident for the connection between segmental properties and tone in terms of tonogenesis (Thurgood 2007, Brunelle and Kirby 2016, Gehrmann and Dockum 2021).

2.3 Acoustics

Given the early emphasis on pitch in the study of tone, the most obvious acoustic correlate would be F0. It is on this basis that much research focuses exclusively on F0 patterns in tone description, but one should note that human perception and processing of linguistic tone may be due to a composite of different acoustic cues (Andruski and Ratliff 2000, Kuo *et al*. 2008, Brunelle 2009, Garellek *et al*. 2013, Kuang 2013).

A case for F0 can be seen in the rather neat match between traditional descriptions of tone contours and the F0 tracks in Standard Chinese, which is said to have four lexical tones (Table 11).

As per the system in Chao (1930), T(one) 1 is a high flat tone [55], T2 a rising tone [35], T3 a dipping tone [214] and T4 a falling tone [51]. Figure 2 offers the time-normalized (using Xu 2013) F0 profiles of each tone averaged from six