1 Introduction

Among the sounds of languages, consonants and vowels need no explanation to the lay person, but tones are another matter entirely. Tell someone you are writing a book about 'Tone', and they look blank, and yet by some estimates as much as 60–70 per cent of the world's languages are tonal. Begin to explain that you are interested in languages that use the pitch of the voice to convey meaning, and more often than not you will be interrupted with a remark such as 'Oh, that must be really interesting: those emotions and nuances and subtleties are so important when we're speaking!' Politely explain that actually you are interested in languages that use pitch to distinguish one word from another, not just to convey subtleties, and most people will assume that such languages are rare, and probably spoken only by isolated communities in less developed countries – until you point out that Mandarin Chinese (885,000,000 speakers), Yoruba (20,000,000) and Swedish (9,000,000) are all tonal.

Perhaps because of these misapprehensions (particularly prevalent in Western cultures), even among linguists tone is sometimes seen as a specialized topic that the general linguist can largely ignore. Undergraduate courses often pay it only cursory attention, and even graduate courses may devote no more than a class or two to the topic. The goal of this book is to fill that gap. It assumes a basic knowledge of phonological theory, but no prior acquaintance with the phonology of tone.

1.1 What is a tone language?

In all languages vowel height and consonantal place of articulation are central to conveying the meanings of words, and so we do not usually categorize languages as being 'vowel-height languages' or 'place-of-articulation languages'. Tone is different in that only a subset of languages (albeit a rather large subset!) make use of it in this way. For a linguist, then, tone has a very specific meaning. A language is a 'tone language' if the pitch of the word can change the meaning of the word. Not just its nuances, but its core meaning. In Cantonese,

for example, the syllable [yau] (which we might spell 'yow' to rhyme with 'how' in English), can be said with one of six different pitches, and has six different meanings:

(1) *[yau] in Cantonese*

high level	'worry'
high rising	'paint (noun)'
mid level	'thin'
low level	'again'
very low level	'oil'
low rising	'have'

In longer words, it matters *where* the tones go. For example, in Dagaare, a Gur language spoken in Ghana, a bisyllabic word can be first low then high, or the reverse, and the meaning changes completely; the acute accents show high tone, and the grave accents show low tone.

(2)

LH	yʊ̀ò́rí	'penis'
HL	yʊ́ó̀rì	'name'

In other languages, the only thing that matters is that the lexical tone of a word appear somewhere in that word, but its exact location may change depending on the morphology of the complex word, and the surrounding phonological context. In Chizigula, a Bantu language spoken in Tanzania, some words have all syllables low-toned, like the various forms of the verb /damany/ 'to do', whereas others have one or more syllables with a hightone, as in the syllables marked with acute accents in the forms of the verb /lombéz/ 'to request':

(3)

Toneless verbs:		*H-tone verbs:*	
ku-damany-a	'to do'	ku-lombéz-a	'to request'
ku-damany-iz-a	'to do for'	ku-lombez-éz-a	'to request for'
ku-damany-iz-an-a	'to do for each other'	ku-lombez-ez-án-a	'to request for each other'

The high tones are part of the lexical entry of certain verb roots, like /lombéz/ 'request', but they show up on the penultimate syllable of the complex verb form, and not necessarily on the verb root itself. Nonetheless, the tone is always there somewhere, and distinguishes high tone verbs from toneless verbs like /damany/ 'do'. This book is about languages like Cantonese, Dagaare and Chizigula, which are called 'tone languages', or more precisely 'lexical tone languages'.

It is not entirely straightforward to decide when a language is a tone language and when it is not. Many languages have occasional uses of pitch to change meaning. In American English, if one says 'Uh-huh' with high pitch on the first syllable and low pitch on the second, it means 'No'. If one says it with low on the first syllable and high on the second, it means 'Yes'. The only other difference between the

two words is whether the second syllable begins with a glottal stop in [ʔáʔà] 'No' or an [h] in [ʔà há] 'Yes', so these words are close to a minimal pair distinguished only by tone. Nonetheless, we would not want to call American English a tone language, because in the overwhelming majority of cases pitch does not change the core meaning of a word, so that 'butter' means 'butter' whether it has a high-low or a low-high pattern. It is true that at the level of the sentence, or, more precisely, utterance, pitch can denote such things as statements, questions, orders, lists, and so on, but we reserve the word 'intonation' for this use of pitch, and it seems to be found in all languages, whether or not they have lexical tone, as we shall see in chapter nine. Using pitch like this 'to convey "postlexical" or sentence-level pragmatic meanings in a linguistically structured way' (Ladd 1997) is not enough to earn a language membership in the class of tone languages.

A subtler question is how we distinguish between what are called stress languages and tone languages. In English, the words 'guitar' and 'glitter' are pronounced with different pitches. In normal statement intonation, 'guitar' has high falling pitch on its last syllable, but 'glitter' starts the fall on the first syllable. Should we then conclude that these words have high falling tones on different syllables in the lexicon? The answer is no, because it turns out that the actual pitch of these syllables depends entirely on the intonation pattern of the utterance into which they are put. Suppose we say the following two dialogues:

(4) A. Tom's just bought himself a guitar.
B. A guitar? I thought he played the drums.

(5) A. I thought I'd sprinkle glitter on her birthday cake.
B. Glitter? You can't eat glitter.

If the second speaker in each case is incredulous about the first speaker's statement, she can say the words 'guitar' and 'glitter' with a quite different pitch pattern. 'Guitar' will have a very low then rising pitch on the last syllable, and 'glitter' will have a very low pitch on the first syllable, rising into the second syllable. There is no truly high pitch anywhere in either word in this context. So pitch does not stay in any way constant for these words. Instead, what is held constant is that in each word one of the two syllables is more prominent than the other, and attracts the intonational pitch, whether it is the statement's high fall, or the incredulous response's extra low-rise. In 'guitar', this is always on the second syllable, whereas in 'glitter' it is always on the first. English then is what is termed a stress language, not a tone language. Stress languages have one other common property, not illustrated by our sample words so far. The stressed syllable does not usually have to be identified in the lexicon, but is generally picked by a counting algorithm that starts from one end of the word,

and selects, for example, the second-to-last syllable, or the first syllable, as the stressed one. Other factors, such as syllable size and morphological structure, may also affect stress placement, but in the typical stress language it is not lexically marked.

This simple typology of tone languages versus stress languages is blurred by the existence of a large group of languages called accentual languages. Such languages, which include, for example, Japanese, Serbo-Croatian, and some types of Dutch, have lexical tones, but what makes them special is that these languages have only a small number of contrasting tones (usually only one or two), and these are sparsely distributed or even absent on some words and usually belong to specific syllables, from which they are inseparable. There is no absolute division between accent languages and tone languages, just a continuum from 'accent' to 'tone' as the number and denseness of tones increase, and they become freer to move around. I shall follow many previous authors in taking the position that the so-called accentual languages are just a subclass of tone languages, and adopt a definition of a tone language from Hyman (in press) that is designed to include the accentual languages under its umbrella:

(6) *Definition of a tone language:*
'A language with tone is one in which an indication of pitch enters into the lexical realization of at least some morphemes.'

Although accentual languages as a subtype of tone language fall under the purview of this book, most of my examples will be drawn from those languages that everyone calls tonal, and the term accentual will still be used from time to time for convenience. For further discussion see chapters six and especially nine.

Before we look at actual tone languages, there are some important background issues that we need to discuss. First, it is essential that we understand something of the phonetics of tone, the basic mechanisms that underlie our ability to produce different pitches. Second, we need to discuss where the work of the phonology ends and that of the phonetics begins. Third, we need to think about the place of the tonal phonology in the larger grammar, including how phonology, syntax, and semantics communicate so that tonal information originating in any of these components is integrated into the larger whole. In this introduction I will give a brief overview of each of these issues, but they will arise again at various points in the book. The discussion is necessarily technical at times, and assumes a solid prior background in general linguistic theory. Some readers may prefer to skip one or more of these sections for now and return to them later. In that case the reader can proceed to chapter two, which jumps right in to the subject matter of this book, beginning with an overview of the range of tonal contrasts found in languages.

1.2 How is tone produced?

This book is a book on phonology, not phonetics, but it is still important to have some idea of how tones might be produced and perceived. An understanding of the phonetics of tone sheds light on the relationship between tone and other aspects of the phonology, such as voicing in obstruents, and also helps our understanding of the tonal phonology itself, for example in understanding some phonological processes as the phonologization of phonetic processes. In this chapter I will discuss mainly the production side of the picture, and leave perception for chapter ten, where it leads naturally into first-language acquisition. I start with a discussion of the larynx, and how pitch differences are produced. I then move on to discuss some consequences of the physiological constraints on the realization of pitch – peak delay and declination; these are important here because they have been phonologized in many languages.

There are three terms that need to be distinguished in any discussion of tone: fundamental frequency (F_0), pitch and tone. In this order, the terms move from a purely phonetic term, F_0, to a truly linguistic one, tone. F_0 is an acoustic term referring to the signal itself: how many pulses per second does the signal contain, where, in the case of the speech signal, each pulse is produced by a single vibration of the vocal folds. The frequency of these pulses is measured in Hertz (Hz) where one Hertz is one cycle per second. The next term, pitch, is a perceptual term. What is the hearer's perception of this signal: is it heard as high in pitch or low in pitch, the same pitch as the previous portion of the signal, or different? The mere existence of F_0 differences may not be enough to result in the perception of pitch differences. The F_0 changes could be too small, or be the result of segmental or other factors for which the hearer unconsciously compensates. Pitch can be a property of speech or non-speech signals. For example, music varies in pitch constantly, and we talk of a high-pitched scream, bird-call, or squeal of tyres. Tone, on the other hand, is a linguistic term. It refers to a phonological category that distinguishes two words or utterances, and is thus only a term relevant for language, and only for languages in which pitch plays some sort of linguistic role.

1.2.1 The larynx

The perception of tone is dependent in whole or in part on pitch perception, and thence on fundamental frequency, or F_0. For distinct tones to be perceived, the signal must contain F_0 fluctuations, and these must in turn be large enough to be perceptible as pitch differences. The fundamental frequency of a sound, which we perceive as pitch, is primarily determined by the frequency of vibration of the vocal folds inside the larynx. The following explanation of the

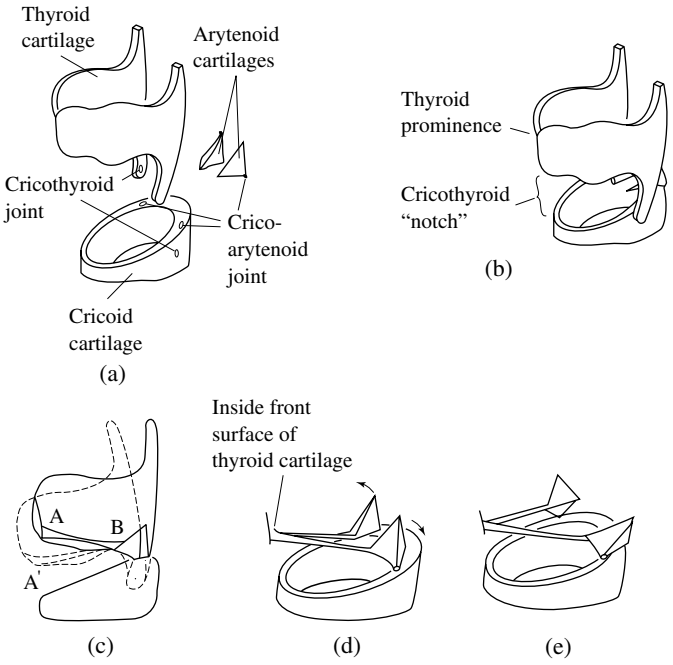

Figure 1.1 *The larynx, from Ohala 1978 (a) An exploded schematic representation of laryngeal cartilages and their movements. (b) Cartilages as they are normally joined. (c) Manner of rotation of thyroid and cricoid cartilages which cause vocal cords, AB, to increase in length, A′B. (d) Adducted position of the vocal cords when arytenoid cartilages are tilted inward. (e) Abducted position of the vocal cords when arytenoid cartilages are tilted outwards.*

laryngeal mechanisms responsible for regulation of pitch is taken mainly from Ohala 1978 and Hirose 1997.

The larynx is composed of two rings of cartilage, the cricoid cartilage, and the thyroid cartilage; the latter is an open ring, sitting on top of the former. There are also two smaller pieces of cartilage, called the arytenoid cartilages, sitting on top of the rear rim of the cricoid cartilage. Figure 1.1, particularly (a–b), should help to visualize the anatomy. The vocal folds (often wrongly called the vocal 'cords') are two bands of muscle, the vocalis muscle, that join the thyroid cartilage and the two arytenoid cartilages. They can be seen clearly in Figure 1.1 (d). The space between them is the glottal opening (the glottis) that allows air to pass from the lungs into the mouth. Rotation of the arytenoid cartilages brings the vocal folds closer together or further apart, thus opening or closing the glottis. This can be seen in Figure 1.1 (d–e).

We are now ready to understand why the vocal folds vibrate at all. First, the vocal folds are brought rather close together by the adductor muscles. Air is forced

through the narrow glottal opening from the lungs, and Bernoulli's Law exerts a sucking effect that draws the vocal folds closed. Pressure from the lungs then builds up behind the closure, and eventually bursts through, releasing a puff of air and reducing the sub-glottal pressure again. The cycle re-starts. Each burst of air is one cycle of vocal fold vibration, and this may happen from a low of around eighty times per second in normal male speech to a high of around 400 times per second for a female voice. Note that, because the vibration is caused by the pressure drop across the glottis, it will only take place if the pressure in the lungs and the oral pressure are different. If there is complete oral closure, as in a stop consonant, the oral pressure may not be sufficiently different from the sub-glottal pressure for vibration to take place, whereas during sonorants air flows out of the mouth, keeping the oral pressure low and the pressure drop high. This creates ideal conditions for vibration, and the ensuing voicing is known as spontaneous voicing.

In a stop consonant, keeping the voicing going requires particular conditions. If the vocal folds are stiff, they will only vibrate if there is a large pressure difference across the glottis. As a result stop consonants produced with stiff vocal folds are voiceless. Since the vocal folds are stiff, the following vowel is produced with raised pitch. If the vocal folds are slackened, they vibrate more readily, and thus it is possible to keep voicing going. Because the vocal folds are slack, the following vowel has lower pitch (Halle and Stevens 1971). A striking example in which this effect has become phonological is found in Songjiang, a Wu dialect of Chinese. The numbers are a way of showing pitch. 5 means highest pitch, 1 means lowest pitch, and so on. Where there are two digits they refer to the pitch at the start and the end of the syllable respectively

(7) *Songjiang tones:*

ti	53	'low'	di	31	'lift'
ti	44	'bottom'	di	22	'younger brother'
ti	35	'emperor'	di	13	'field'

What you can see is that the words in the right-hand column, which begin with a voiced obstruent, have lowered versions of the pitches of the words in the left-hand column, which begin with a voiceless obstruent. This connection between voiceless obstruents and high pitch, and voiced obstruents and low pitch, is widely attested in natural languages, and in many cases it is possible to trace the origins of tonal contrasts back to a prior contrast in voicing on obstruents, in a process known as tonogenesis.

In vowels and sonorant consonants the rate of vibration of the vocal folds is controlled by a number of factors. Rotation of the thyroid and cricoid cartilages with respect to each other causes changes in the length of the vocal folds. By these means, the vocal folds can be deformed in several ways, and as a result they may or may not vibrate, and the frequency of vibration may be controlled.

For those readers interested in a little more detail, we know that pitch differences come from adjusting the mass and stiffness of the vocal folds (Hirose 1997). The crico-thyroid muscle contracts, and this elongates the vocal folds, decreasing their effective mass and increasing their stiffness. This increases the frequency of vibration, and raises pitch. In tone languages, it can be shown very clearly that it is the activity of the crico-thyroid muscle that is primarily responsible for raising pitch. An increase in the activity level of this muscle precedes each pitch peak by a few milliseconds. Pitch lowering has slightly more complex causes. The activity of the crico-thyroid muscle is reduced, while the thyro-arytenoid muscle contracts, thickening the vocal folds and increasing their effective mass.

Apart from internal changes to the larynx, there are some other articulatory mechanisms that have been implicated in pitch control. The main one is larynx lowering. There is some reason to think that lowering the larynx may play quite an important role in lowering pitch, presumably because it stretches and thins the vocal folds somewhat (see Ohala 1978 for discussion). One way or another, then, vocal fold vibrations are the primary source of pitch differences, although other noise sources, such as the turbulent noise produced at the narrow constriction of the fricatives [s] and [ʃ], may also differ in pitch. Nonetheless, controlled pitch differences (as opposed to ones that are automatic concomitants of other aspects of articulation) are always produced at the larynx in speech.

This very brief and over-simplified explanation of the production of tone is sufficient for our purposes. For more details, the interested reader can consult Ladefoged 1975, Ohala 1978, Stevens 1997, Hirose 1997.

1.2.2 *Performance factors that affect pitch*

The physiology of speech production has further effects on the speech signal, and two of these effects deserve mention here. When the brain sends a signal to produce high tone, instructions go to the appropriate muscles. The muscles configure the vocal folds suitably, and the rate of vibration then increases, resulting in high pitch. All this takes a small but finite amount of time, and as a result the full flowering of high pitch is somewhat delayed. The delay is enough that the peak is typically at the end of the tone-bearing segment, or indeed often not reached until early in the following syllable. The term ‘peak delay’ is usually used for the latter case. This effect has been well documented in languages as diverse as Mandarin Chinese (Xu 1998, 1999b), Chichewa (Kim 1998, Myers 1999b) and Yoruba (Akinlabi and Liberman 1995). The schematic pitch trace in Figure 1.2 from Xu 1999b shows how three different tones on a medial syllable – high (H), falling (F) and rising (R) – are realized between two low tones. First, look at the heavy dashed line, which shows the realization of a high-toned syllable in between two lows. It can be seen that the high peak is not reached until the very end of the syllable. Now look at the solid line,

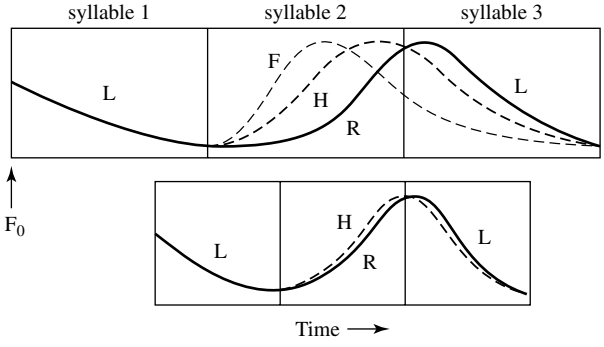

Figure 1.2 *Peak delay, from Xu 1999b. Schematic peak alignments in F, H, and R at normal speaking rate (upper panel), and in H and R at fast speaking rate (lower panel).*

which shows a rising tone between two lows. The culmination of the rise is delayed all the way into the following syllable. Finally, look at the light dashed line. The fall of the falling tone does not begin until more than half-way through the syllable. Although most of the data focusses on the delay of high peaks, it is probably true for any change in pitch movement, so that a low trough can also be delayed.

A second physiologically based phenomenon is declination, by which the pitch of an utterance falls as the utterance proceeds. This has been observed in tonal and non-tonal languages alike, but the mechanism is not fully agreed upon. One possibility is as follows. As an utterance proceeds, assuming the speaker has not paused for breath in the middle, the amount of air in the lungs decreases, and the sub-glottal pressure drops. As a result, the pressure difference across the larynx decreases, and the rate of vibration of the vocal folds slows, so the pitch lowers. This means that the same amount of muscular effort aimed at producing a high tone produces a lower-pitched version of this high tone later in the utterance than it does at the beginning. Of course, if additional effort is exerted, the pitch can be raised back up, but the overall trend is downwards. The problem with this plausible-sounding explanation for declination is that sub-glottal pressure has been measured, and it is clear that it drops very little during an utterance, and probably not enough to account for the size of the declination. See Ohala 1978 for some other possible mechanisms.

These two phenomena – peak delay and declination – are of interest here because they have been phonologized in many languages. For example, in Yoruba (Akinlabi and Liberman 2000b), peak delay has developed into a phonological process that turns a high-low sequence into a high-falling sequence by spreading the high tone. An acute accent shows high tone, a grave accent shows a low tone, and a circumflex shows a fall.

(8) rárà (HL) → rárâ (H HL) 'elegy'

More generally, tone spread or shift to the right is very common, but tone shift or spread to the left is much rarer. Our second phenomenon is extremely widespread, especially in Africa, where declination has apparently given rise to a phonological process called downdrift or downstep by which high tones are drastically lowered after low tones. See chapter six for details.

I end this section with a rather obvious point. Just like segmental contrasts, tonal contrasts can be affected by co-articulation effects (Peng 1997, Xu 1994). The laryngeal articulators, as has already been observed, have their own inertia, and it takes time for change to take place. Hearers seem well able to compensate for these effects, and continue to recognize the tones, but nonetheless caution must be observed in deciding whether some particular tonal effect is phonetic or phonological, and indeed the answer is not always clear. One relatively uncontroversial diagnostic is whether the effect in question is dependent on speech rate, and is variable in extent. If it is, it is usually classified as phonetic. If, on the other hand, it takes place at all speech rates, and is an all-or-nothing categorical affair, then it is usually classified as phonological.

1.3 The structure of the grammar: Phonetics and phonology

So far we have been discussing phonetics, but the main topic of this book is the *phonology* of tone. It is not always easy to know where phonology ends and phonetics begins, nor to understand the nature of the relationship between the two. In order to keep things clear, in this section I shall spell out what I am taking to be the division of labour between phonology and phonetics, and how they communicate with each other. In recent years there has been much discussion of these questions, but it would be beyond the scope of this book to go deeply into the issues. The interested reader is referred to any of the volumes in the Papers in Laboratory Phonology series, particularly the introduction to Beckman and Kingston 1990. In what follows I articulate issues that arise again later in the book. Some readers may find it hard to grasp their significance at this point, and may wish to (re-)read this section later.

I shall assume a rather traditional model with the following properties. Phonological representations are categorical, using either binary or unary features. It is the business of the phonology to generate an output out of these elements, in which most segments are specified for most features, but some may lack specifications for certain features. In particular, some syllables may lack tones at the end of the phonology. The phonetics then interprets this phonological output, making use of all phonological information: featural, structural, phrasal, and so on. This phonetic component ultimately produces instructions to the articulators; these instructions may or may not be binary, but in any case they result in a continuous signal in which every syllable is pronounced at some fundamental frequency. In