Proceedings of the 41st Annual Regional Meeting of the Chicago Linguistics Society (2008), 41(2), 129-144.

On the dual relativity of tone

Laura C. Dilley
MIT and The Ohio State University

Introduction

A longstanding debate in tonal phonology concerns how to characterize the relative nature of tone. Two types of tonal relativity are observed. First, tones are relative to a speaker's pitch range. Thus a canonical lexical High tone is higher in a given speaker's pitch range than a canonical lexical Low tone in the same position. Second, tones are relative to one another in sequence. Thus a High tone is canonically higher than an adjacent Low tone. The necessity of defining both types of relativity in the grammar is exemplified by the fact that tones can drift down over the course of a phrase, such that a High tone near the end of the phrase may be lower than a Low tone near the beginning of the phrase. In spite of this drift, alternating High and Low tones canonically maintain their heights relative to one another.[1] This suggests that tones are defined both with respect to the speaker's pitch range, as well as with respect to one another in sequence. The question of how best to capture this dual relativity in a formal linguistic theory has posed a challenge for many years.

Here, I argue that a phonology which defines both types of tonal relativity is required for cross-linguistic descriptive and explanatory adequacy. To illustrate this, I will begin by examining autosegmental theory (Williams 1971/1976, Leben 1973, Goldsmith 1976), which implicitly defined tones only with respect to the speaker's pitch range. I then show how this leads to significant problems in describing the tonal systems of some languages. Next, I show that the mathematical theory behind musical melody provides an attested framework for capturing the dual relativity of tone, and that concepts and formalisms from music can be applied to language. Finally, I show how this framework captures some basic facts for two tonally distinctive languages, Yoruba and English.

The Conception of Tone in Autosegmental Theory

Autosegmental theory entailed a great many claims regarding the nature of the representation of tone in language which can and should be separated and examined independently on their own merits. Perhaps the most well-known claim of autosegmental theory is that tones and segments are represented on distinct, temporally coordinated levels of representation, or "tiers". Separating tones and segments in this way permitted an account for a number of lexical tone phenomena which had no comprehensive explanation at the time (Odden 1995).

---

[1] A few languages exhibit total downstep, whereby a H in a LH sequence is lowered to the level of the preceding L (cf. Stewart 1993). This phenomenon is not the focus of the present paper.

Proceedings of the 41st Annual Regional Meeting of the Chicago Linguistics Society (2008), 41(2), 129-144.

Another claim of autosegmental theory was that tones are not only separate from segments, but that tones *are* segments.[2] This led to strong, implicit claims about the nature of tones which can be traced to the notion of segments in traditional generative phonology. In generative phonology, segments consisted inherently of an opposition (de Saussure 1966). Thus, /i/ was defined through its opposition to /a/, as in *seed* vs. *sawed*. Similarly, a High tone was defined as higher than a Low tone would be in the same position. Moreover, segments were viewed as encapsulated bundles of features which could be strung together like beads on a string. Tones, too, were conceived of as having these characteristics.

Defining tones in this way effectively meant that only one manner of relativity for tones was captured in the phonology: tones were defined solely as relative to the speaker's pitch range. In contrast, the relativity of tones one to another in sequence was relegated to the domain of phonetics proper. The implications of this phonological treatment for the overall linguistic model were not a focus of early autosegmental endeavors. Over the past 25 years, a number of influential proposals have been put forward regarding the nature of the corresponding phonetic component. In the following, I argue that under all comprehensive phonetic proposals within the autosegmental framework, the phonetic component has too much power, leading to problems with descriptive adequacy. Moreover, I show that problems of *explanatory* adequacy arise from the claim that tones are defined in the phonology as relative only to the speaker's pitch range.

The Problem of "Free Scaling" Languages

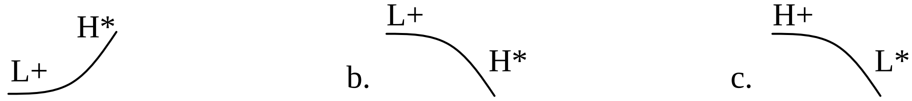
The crucial test cases for the strong view of tonal relativity put forward under autosegmental theory were languages which exhibit "free scaling" of tones, such as English, French, German, Turkish, etc. In such languages, tones may be scaled either high or low in the speaker's pitch range, due to the influences of various phonetic and paralinguistic factors. Because these languages afford more degrees of freedom in the placement of tones than in many lexical tone systems, they potentially present a greater challenge for descriptive and explanatory adequacy, given autosegmental theory's claims about the phonology.

The central challenge for models of the phonetic component accompanying autosegmental theory concerned how to constrain the relative heights of adjacent tones. Phonetic factors have generally been assumed to be in control of two dimensions of variability: the heights of tones in the pitch range, and the relative heights and intervals between tone pairs. To see why permitting too much freedom in the phonetic component can lead to problems, consider the kinds of contours that might result from scaling a LH sequence, such as a L+H* accent, if relative tone height is not sufficiently constrained. The pattern in (1)a shows the

[2] The notion that tones are segments was the basis for the term "autosegment".

expected shape for L+H*, in which L+ is lower than H*. On the other hand, the contour in (1)b shows an unexpected shape for L+H*, in which L+ is *higher* than H*. The pattern in (1)b presents a critical ambiguity with that of (1)c, which represents the expected F0 shape for a HL sequence, such as a H+L* pitch accent.

(1) a. b. c.



This example illustrates that when phonetic rules do not sufficiently constrain the relative heights of H and L tones, two kinds of problems result. The first is *overgeneration* of possible F0 contours. This is exemplified in the fact that L+H* might give rise to patterns in both (1)a and (1)b. The second is *indeterminacy* of phonological representations. This is shown by the fact that a falling pattern could correspond underlyingly to either H+L* or L+H*. The crucial restrictions that must be in place in order to prevent these problems are that L+ must not rise above H*, nor must H* fall below L+.[3]

How well have proposals regarding the phonetic component that accompanies the phonology of autosegmental theory avoided problems with overgeneration and indeterminacy? It turns out that all models which have proposed a comprehensive account of factors scaling H and L tones (Pierrehumbert 1980, Liberman and Pierrehumbert 1984, Pierrehumbert and Beckman 1988) give rise to these problems. (See Dilley 2005, Ch. 4, for a discussion and proofs.)

It is worth considering whether existing phonetic models could be modified so as to appropriately restrict relative tone height, thereby avoiding the problems described above. The answer is yes, although no such proposals have been put forward. However, a critical issue which would remain unaddressed by simply restricting parameters in the aforementioned phonetic models is that all these models deal only with one side of the phonetics-phonology coin: the mapping from *phonology* to *phonetics*. None of these models provides even a starting point toward an account of how listeners perceptually recover tones from speech. Moreover, given the multiplicity of rules and parameters that are required to support the strong phonetic component assumed under autosegmental theory, there is no clear way that a listener could “reverse engineer” the signal to recover the underlying representation. To complicate matters, most of the parameters of these models are assumed to be inaudible and to lack an acoustic basis.

An alternative approach to patching up these phonetic models would be to define a tonal representation in which relative tone height is part of the phonology. Such a move readily addresses the problems with overgeneration and indeterminacy

[3] More generally, preventing these problems requires that the relative heights of *every pair* of adjacent tones be constrained at some level of the grammar.

Proceedings of the 41st Annual Regional Meeting of the Chicago Linguistics Society (2008), 41(2), 129-144.

discussed earlier by defining L as canonically lower than adjacent H in the phonological representation. This has several advantages. First, it eliminates the need for a bloated and overly powerful phonetic component. Moreover, it leads to a proposal for how listeners perceptually recover phonological information about tone from the phonetics. Instead of assuming that listeners "throw away" relative tone height, we can instead assume that relative tone height is directly relevant to the phonology. Evidence that relative tone height is used to decode the phonology comes from a study by Wong and Diehl (2003), who showed that Cantonese listeners relied more on relative tone height in recovering lexical tones than on position in a speaker's pitch range, supporting a phonological account.

There are additional reasons for assuming that relative tone height is part of the phonological representation. First, in lexical tone languages, we find contrastive use of relative tone height. For example, in Igbo *ámá* ('distinguishing mark') contrasts with *ám!á* ('street') (Williamson 1972, V. Manfredi, pers. comm..), where '!' indicates the site of pitch lowering; the distributional facts about tone in this language fail to support the analysis of the downstepped version as a separate lexical tone category. Another line of evidence comes from an experiment by Dilley (2005). In this experiment, the relative levels of two English syllables were varied along a continuum in an imitation task. In response, listeners produced categorical responses, a finding which also supports a phonological analysis for relative tone height. Finally, the relative heights of tones in sequence comprise the most salient aspect of the representation for other kinds of tonal sequences, such as musical melodies (Dowling and Fujitani 1971, Dowling and Harwood 1986).

Collectively, these observations suggest that tones should be characterized in the phonology as both relative to the speaker's pitch range and to one another. How should this dual relativity of tone formalized? One proposal for incorporating relativity of tones into the phonology comes from Snider (1999). This proposal defines two types of binary features: tone features (H, L) and register features (h, l). A key component of this proposal is a continuously-valued parameter called the tone-register ratio, which controls the relative heights of tones. It turns out that for tones with opposite feature values of tone and register (e.g., a H tone in low register), relative tone height is underdetermined by the tone-register ratio. As a result, this proposal encounters problems of overgeneration and indeterminacy similar to those discussed earlier. Moreover, it is not clear how this approach would be extended to a language such as English. For these reasons, a different approach was pursued in capturing the dual relativity of tone.[4]

---

[4] Another phonological proposal for representing tones as relative to one another in the phonology was Clark (1978). However, Clark's proposal entailed that the phonology only represented one manner of relativity, namely, between tone bearing units in sequence. While both Clark's and Snider's proposals showed insight by defining tones as relative to one another, it is not clear how either approach could be amended to address all the problems considered here.

Proceedings of the 41st Annual Regional Meeting of the Chicago Linguistics Society (2008), 41(2), 129-144.

Dual Relativity in Musical Melody

Music theory offers an attested formal framework and mathematics for describing the dual relativity of tone which can readily be adapted to account for tonal phenomena in language. Prior to presenting a linguistic proposal, it will be useful to give some background in music theory; this in turn will be facilitated by first defining some musical terminology.

Musical notes are named with letters of the alphabet. The lowest white key on the piano is an A. The next six ascending white keys from left to right are named B through G, then the cycle repeats starting with A, on up the keyboard. The interval distance between every eight notes, such as the distance from one A to the next higher A, is termed an *octave*. Notes are given numerical subscripts to indicate the octave in which they occur. Thus A_1 is the lowest note on the keyboard, while C_8 is the highest. Finally, the black keys are named in relation to white keys on the keyboard. Thus $A^{\#}$ (A-sharp) and A^{b} (A-flat) correspond to the black keys just to the right and to the left of each A on the keyboard, respectively. Now, consider that the first part of the melody "Mary Had a Little Lamb" may be played with any of the following note sequences, and many others:

(2)	a.	Key of C	E_4	D_4	C_4	D_4	E_4	E_4	E_4
	b.	Key of F	A_4	G_4	F_4	G_4	A_4	A_4	A_4
	c.	Key of D	$F^{\#}_4$	E_4	D_4	E_4	$F^{\#}_4$	$F^{\#}_4$	$F^{\#}_4$
	d.	Key of G^b	B^b_5	A^b_5	G^b_5	A^b_5	B^b_5	B^b_5	B^b_5

Given that any of these sequences of notes is easily recognized as the familiar nursery rhyme tune, the obvious question is: what are the properties of these note sequences which convey the melodic representation to the listener? It is certainly not the individual fundamental frequencies. If N_i is the frequency of the i^{th} musical note, the frequencies of the notes in (2) are given in (3), to the nearest 1 Hertz:

			N_1	N_2	N_3	N_4	N_5	N_6	N_7
(3)	a.	Key of C	330	294	262	294	330	330	330
	b.	Key of F	440	392	349	392	440	440	440
	c.	Key of D	370	330	294	330	370	370	370
	d.	Key of G^b	932	831	740	831	932	932	932

It turns out that each sequence of notes reflects the *same ratio relationships* among the frequencies. There are two kinds of ratio relationships attested in the above sequences. First, all note sequences in (3) involve the same sequence of ratios between successive note *pairs* N_{i+1}/N_i shown in (4). This is known as a *sequence of intervals*.

(4)

N_2/N_1	N_3/N_2	N_4/N_3	N_5/N_4	N_6/N_5	N_7/N_6
0.89	0.89	1.12	1.12	1.0	1.0

Second, all note sequences in (3) involve the same sequence of ratios between the frequency of each note and the frequency of the referent note of the key. The referent note of the key is called the *tonic*; for example, C is the tonic note in the key of C.[5] Dividing each of the frequencies N_i by the frequency of the tonic note N_t gives the sequence of ratios in (5). This is known as a *sequence of scale notes*.[6]

(5)

N_1/N_t	N_2/N_t	N_3/N_t	N_4/N_t	N_5/N_t	N_6/N_t	N_7/N_t
1.26	1.12	1.0	1.12	1.26	1.26	1.26

These examples illustrate the first important concept from music theory, namely, that a given note can be defined as relative both to another note, as well as to the tonic note of the key.

A second important concept is that the representation of musical melody is "layered", with different layers entailing other layers. For example, the sequence of scale notes in (5) readily gives rise to the sequence of intervals in (4). To see why, note for example that the first ratio in (4) for the tone sequence N_1 N_2 is N_2/N_1 = 0.89. However, this is also the result of dividing the second ratio in (5), N_2/N_t, by the first ratio, N_1/N_t, so that:

(6) $$\frac{N_2/N_t}{N_1/N_t} = \frac{N_2}{N_1} = \frac{1.12}{1.26} = 0.89$$

Note that the sequence of intervals in (4) can be extracted from the sequence of scale notes in (5), but it is not possible to generate (5) from (4) without knowing which note is the tonic.

Another way in which melodic representations are layered is that the interval sequence entails a level of representation called the *up-down sequence*, which describes whether the next note is higher, lower, or the same relative to the current note.[7] This corresponds to ratios N_{n+1}/N_n which are greater than 1, less

---

[5] The notion of a centrally most important referent note in scale systems appears to be a musical universal. In non-Western scale systems, the centrally most important note is called the *mode*.

[6] Major and minor scales in Western tonal music are defined by seven notes numbered I-VII. For the major diatonic scale, the frequencies of these notes with respect to the tonic are approximately 1.0, 1.12, 1.26, 1.33, 1.50, 1.68, and 1.89, respectively.

[7] In the music literature, the up-down sequence is termed *contour*, a term which I have avoided due to its confusability with F0 contour.

than 1, or equal to 1, respectively. Thus the interval sequence in (4) reduces to the up-down sequence in (7):

(7)

N_2/N_1	N_3/N_2	N_4/N_3	N_5/N_4	N_6/N_5	N_7/N_6
<1	<1	>1	>1	$=1$	$=1$

Note that while an up-down sequence can be generated by an interval sequence, it is not possible to generate the latter from the former.

Finally, given the sequence of absolute frequencies in a melody, the scale notes, intervals and up-down pattern can be recovered.[8] However, the actual frequencies have little importance for the representation of musical melody (Burns 1999).

All of this leads to the question of whether a melody can be more or less specified for these levels of representation. Looking across musical cultures and genres, the answer is clearly *yes*. Some musical melodies are characterized only by an up-down pattern, such as Australian aboriginal music (U. Will, pers. comm.). Other melodies or parts of melodies are specified for both interval and up-down pattern but not scale, including Western atonal music and folk music forms in the Middle East, North Africa, and India (N. ben Zvi, pers. comm.).[9] Melodies can also be specified for scale, interval, and up-down pattern. The majority of Western music falls into this category. Finally, for highly familiar melodies, such as TV show theme songs which are always heard in the same key, absolute frequency can also be represented, together with scale, interval, and up-down sequence (Schellenberg and Trehub 2003).

Next, a third concept from music is that the frequencies defining musical notes are related through *reciprocal ratios*. For example, stepping down from D_4 (294 Hz) to C_4 (262 Hz) corresponds to a ratio of (262/294) = 0.89, while stepping up from C_4 to D_4 corresponds to a ratio of (294/262) = 1.12.

A fourth and final concept from music theory is that the notes in musical melodies are associated with positions in rhythmic structures such as metrical grids (Lerdahl and Jackendoff 1983). Given the hierarchical structure of metrical grids,

[8] Actually, recovering the scale notes requires that the intervals conform to specific frequency ratios. In Western tonal music, the intervals are generally restricted to the specific ratios associated with the major diatonic scale. Even musically untrained listeners have unconsciously internalized these ratio relations, a fact which permits them to perceptually recover the tonic note hence the entire scale (Krumhansl and Kessler 1982).

[9] Musical forms known as makams and ragas fall into this category. These forms are frequently ambiguous with respect to mode, since the cues to mode type come at discrete points throughout a musical piece. The result is that the mode is often locally undefined over long passages (N. ben Zvi, pers. comm.)

Proceedings of the 41st Annual Regional Meeting of the Chicago Linguistics Society (2008), 41(2), 129-144.

it turns out that ratio relations may exist between notes which are nonadjacent in a sequence, so that these relations form the “backbone” of certain melodies. For example, in the work of J. S. Bach and other Baroque composers, high and low notes commonly alternate, such that nonadjacent H tones or nonadjacent L tones “carry the melody” while the remaining notes are “backgrounded”.

It turns out that the ratio relations between nonadjacent notes which “carry the melody” can be mathematically calculated from the sequence of ratio relations shared by *adjacent* notes. To illustrate this, suppose we assume that N_2 and N_5 in (3) are associated with metrically prominent positions. Then the ratio relation between these two notes can be calculated by the sequence of adjacent relations among N_2, N_3, N_4, and N_5, as follows:

(8) $$\frac{N_5}{N_2} = \frac{N_5}{N_4} \cdot \frac{N_4}{N_3} \cdot \frac{N_3}{N_2}$$

This concludes the present review of music theoretic concepts. Before illustrating how these principles can be applied to language, it is necessary to address a possible objection to this approach. A common criticism is that musical systems are more rigidly structured than linguistic tonal systems, so that comparisons between the two are not valid. Two points can be made in response. First, Western musical systems are less rigid than they may first appear. Thus, singers frequently produce F0 values that are sharper (higher) or flatter (lower) than the objectively correct frequency, yet listeners perceive the singing as being in tune (Sundberg 1987). This is probably due in part to categorical perception for musical intervals, which obtains even for musically untrained listeners (e.g., Siegel and Siegel 1977; Howard, Rosen, and Broad 1992). Recent findings of categorical perception for linguistic tonal categories (Hallé, Chang, and Best 2004) lend further support to analogies between music and language. Second, the relatively more rigid forms of Western tonal music are not universal. For example, the scale system of many musical cultures in the South Pacific, such as Javanese gamelan music, is characterized by scale notes which have a wide range of frequency ratios relative to those used in Western music (e.g., Perlman and Krumhansl 1996). These observations suggest that developing a framework which pursues connections between music and language may indeed prove fruitful.

A Framework for the Dual Relativity of Tone

In the framework proposed in Dilley (2005), concepts from music are used to formalize the dual relativity of tone in language. Tonal categories are conceived of as abstractions of frequency ratios formed by a tone and a referent, analogous to ratio relations among musical note pairs. These linguistic ratio abstractions are referred to as *tone intervals*. Thus a tone interval $I_{r,n}$ relates a tone T_n to some generalized tonal referent r according to the following basic relation:

(9) $I_{r,n} = T_n/r$

The dual relativity of tone is captured by the fact that two types of referent may be defined, giving rise to two types of tone interval. On the one hand, the referent may be another tone in the sequence, yielding a *syntagmatic tone interval*. For a temporally-ordered sequence of tones T_1 T_2, tone T_1 is the referent and the corresponding syntagmatic tone interval is given in (10).

(10) $I_{1,2} = T_2/T_1$

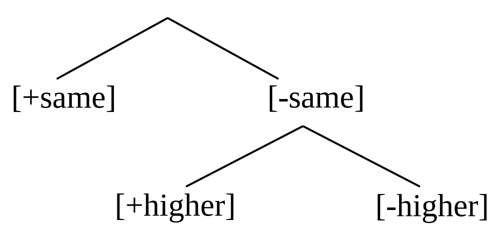
On the other hand, the referent can be a "tonic" tone level, analogous to the tonic of a given scale in music, yielding a *paradigmatic tone interval*. Thus, for a tone T_n and tonic α, the corresponding paradigmatic tone interval is given in (11).

(11) $I_{\alpha,n} = T_n/\alpha$

Syntagmatic tone intervals are thus equivalent to the notions of interval and up-down pattern in music, while paradigmatic tone intervals are equivalent to the notion of scale in music.[10,11]

It is important to emphasize that the representation is based on ratio relations between tones and their referents, which are formalized as tone intervals, rather than on the properties of tones themselves. Consistent with this approach, each tone interval (rather than tone) is assigned two binary-valued features, [+/-same] and [+/- higher]. These features are arranged into the geometry shown in (12):

(12)



[10] The analogy between musical relativity with respect to a scale note and linguistic relativity with respect to a tonic defined for a speaker's pitch range is supported by the fact that vocal performers often select a key for their performance which is different from the one that is written in order to make the performance "fit" more comfortably within the vocalist's range.

[11] The presence or absence of paradigmatic tone intervals is assumed to be a property which distinguishes languages from one another. Languages which probably represent structures as in (11) include many languages of Africa and East and Southeast Asia. Examples of languages which probably do not represent structures as in (11) include English, Swedish, Japanese, Turkish, etc.

Proceedings of the 41st Annual Regional Meeting of the Chicago Linguistics Society (2008), 41(2), 129-144.

If a tone interval is specified as [+same], the corresponding tone will be at the same level as the referent in the abstract tonal space: $I_{r,T} = 1$. If a tone interval is specified as [-same, +higher], then the corresponding tone will be higher than the referent: $I_{r,T} > 1$. Finally, if a tone interval is specified as [-same, -higher], then the tone will be lower than the referent: $I_{r,T} < 1$. These features are assumed to define language-universal relations between tones and referents. The ratio values of tone intervals can be further restricted in a language-*specific* way to correspond to specific ratios or ratio ranges, in order to capture specific sizes of changes up or down. This is formalized by taking the intersection of $I_{r,T} > 1$ or $I_{r,T} < 1$ with some other range of tone interval values. For example, suppose T_2 steps down lower than T_1 by a small amount equal to the distance of a minor third, which corresponds to a ratio of 0.84. Then stepping down by a minor third corresponds to the intersection of $I_{1,2} < 1$ with $I_{1,2}= 0.84$, which gives $I_{1,2}= 0.84$.

Tones are assumed to associate with respect to a metrical grid, as in music. A variety of evidence suggests that tone placement is sensitive to metrical properties (e.g., Inkelas and Zec 1988, Manfredi 1993, Bickmore 1995, de Lacy 2002). In this way, the present account provides the first comprehensive proposal for capturing interactions between tone and metrical structure within a unified framework. Interactions between tones and grids are captured through different intrinsic preferences of tones for associating with particular metrical grid positions. Starred tones, *T**, associate with grid columns of height 2 or higher, while unstarred tones, *T*, associate with grid columns of height 1. In addition, unstarred tones may be specified for a directionality property which requires them to associate with a metrical position to the left or the right of a starred tone. In this way, unstarred and starred tones are predicted to occupy adjacent metrical positions of contrasting metrical strength. Bitonal accents in English and other languages are thus viewed as a *sequence* of independent tones whose association is coordinated with respect to the metrical structure and with respect to one another. This proposal thus provides a phonological account for a significant body of outstanding phonetic evidence that the two tones in bitonal accents function as independent yet coordinated tonal entities (e.g., Arvaniti, Ladd, and Mennen 1998; Ladd *et al.* 1999; Dilley, Ladd, and Schepman 2005).

Once tones have been situated with respect to the grid, pairs of tones at every grid level are joined into some syntagmatic tone interval which defines the relative height and possibly the interval distance between the tones in the tone pair. Tone intervals resulting from all levels of the grid are represented on a structure called a tone interval matrix. See Dilley (2005) for the details of this analysis.

Examples from Two Languages

I will briefly sketch how this framework characterizes some basic facts from two languages: Yoruba and English. Yoruba has three tones: H, M, and L. Based on

Proceedings of the 41st Annual Regional Meeting of the Chicago Linguistics Society (2008), 41(2), 129-144.

data of Laniran and Clements (2003), it appears that Yoruba tones can be represented in terms of the paradigmatic tone intervals given in (13):[12,13]

(13) a. For all $H \in H_i$: $I_{\alpha,i} = 1.25 \pm 0.04$
b. For all $M \in M_j$: $I_{\alpha,j} = 1.13 \pm 0.03$
c. For all $L \in L_k$: $I_{\alpha,k} = 1.0$

Recall that when musical notes are defined with respect to a common referent, the interval shared by a sequentially-ordered musical note pair can be calculated by dividing scale ratios by one another, as in (6). If Yoruba tones can be formalized in a manner analogous to a musical scale, as claimed here, then we should be able to calculate the interval between two non-identical tones in sequence from the ratios in (13). Then the approximate predicted values for the interval formed by a sequence $T_1\ T_2$ (= T_2/T_1) for all pairs of lexical tones is shown in (14):

(14)

$H_1\ M_2$	$M_1\ H_2$	$M_1\ L_2$	$L_1\ M_2$	$H_1\ L_2$	$L_1\ H_2$
0.90±0.04	1.11±0.04	1.13±0.03	0.88±0.03	0.80±0.03	1.25±0.04

These values are in good agreement with Laniran and Clements' data, supporting the proposal that the Yoruba tone system can be treated as formally analogous to a scale system in music. One important exception to the predictions above concerns the interval from L to H in alternating HLH sequences, which is closer to an average of 1.15 rather than 1.25. This means that the interval from L to H is smaller than what is predicted based on the interval from H to L.[14] In fact, this is the key to an account of downstep, i.e., the lowering of successive H tones in HLH contexts. We can assume there is a rule which replaces the default value of the interval predicted by paradigmatic tone intervals for L and H with a different value, which is given in (15):[15,16]

(15) *Downstep Rule*. In $H_1\ L_2\ H_3$: $I_{2,3} = 1.15 \pm 0.03$

---

[12] These values represent the average mean and average standard deviation of ratios of H to L F0 values and M to L F0 values in corresponding positions in all-H, all-M, or all-L sentences for the two speakers: TJ, a male, and FA, a female.

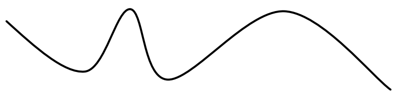
[13] I assume that one lexical tone category is always specified to be at the tonic level, which is defined as the level for which $I_{\alpha,n} = 1.0$. Note that the tonic level, α, may decline over the course of a phrase, just as a choir can drift flat over time.

[14] In particular, the interval from L to H is smaller than the reciprocal of the interval from H to L.

[15] If successive H tones in the sequence are associated with metrically prominent positions, then the Downstep Rule is formally equivalent to the rule that for grid level $j > 1$, for $H_n\ H_{n+1}$, $I_{n,n+1} = 0.92 \pm 0.03$.

[16] Yoruba also apparently has a rule for raising the initial H tone in downstepping sequences of the form HLHLH… Data from speaker TJ suggests that this rule sets the initial H tone, H_1, to an interval value of approximately 1.38 ± 0.02 with respect to the tonic.

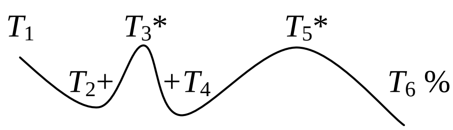
I will now briefly consider some aspects of English intonational phonology. Unlike Yoruba tone, I propose that English intonation is comprised entirely of syntagmatic tone intervals. To illustrate how this proposal works, consider the utterance *An American linguist!* spoken with an animated, declarative accentual pattern, as shown in (16).

(16) 

An Am- er- i- can lin- guist!

In this framework, intonation patterns are constructed from a sequence of tones which are joined into syntagmatic tone intervals. These tones and tone intervals are supplied by an inventory of established tunes, together with language-specific and language-general rules for generating novel tunes.[17] Starred and unstarred tones then associate with specific positions in a metrical grid.

The first step in analyzing the pattern in (16) is to determine the locations of tones with respect to the metrical structure. The contour in (16) has high peaks on *An*, *er-* and *ling-*, as well as low valleys on *Am-*, *i-*and *-guist*. High and low turning points correspond to positions of tones in this framework.[18] The tones on *An*, *Am-*, *i-* and *-guist* are unstarred, since these syllables are metrically weak; moreover, tones on *er-* and *lin-* are starred tones (i.e., pitch accents), since these syllables are metrically strong. This is shown in (17), with tones numbered for reference.[19] These tones are associated with the metrical grid shown in (18).

(17) 

An Am- er- i- can lin- guist!

(18)

```
                            x
             x              x
x     x      x     x    x   x     x
T1    T2+    T3*  +T4       T5*   T6 %
An    Am-    er-   i-  can  lin-  guist!
```

---

[17] For example, English disallows the phrase-final sequence *H* L H L%*, while Swedish and Italian permit this sequence (Ladd 1996, Dilley 2005).

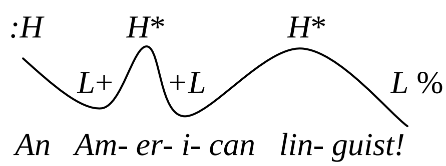
[18] All perceptually significant turning points in an F0 curve correspond to tones according to Dilley (2005). One reason for this is that all transitions between tones are assumed to be monotonic.

[19] I assume that the first and last tones in an intonational phrase associate with the first and last positions in the corresponding metrical grid, rather than with respect to a starred tone. For this reason, the "+" is left off phrase-initial and phrase-final unstarred tones.

Based on the relative heights of pairs of tones in (17), we can infer that the sequence of corresponding syntagmatic tone intervals is given in (19):[20]

(19)
$$\begin{bmatrix} & & & I_{3,5} = 1 & \\ I_{1,2} < 1 & I_{2,3} > 1 & I_{3,4} < 1 & I_{4,5} > 1 & I_{5,6} < 1 \end{bmatrix}$$

Finally, it is possible to collapse the tones and tone intervals into a unitary notation similar to that used previously in autosegmental descriptions. Suppose we replace each tone which is higher than or lower than the tone to its *left* with *H* or *L*, respectively, retaining other symbols such as "*" and "+". Moreover, suppose we replace the initial tone with a symbol :*H* to reflect its relative height with respect to the tone to the *right*. The resulting representation is given in (20).[21]

(20)


Summary and Conclusions

Achieving cross-linguistic descriptive and explanatory adequacy for tonal facts requires defining a phonology which captures the dual relativity of tones with respect to the speaker's pitch range and with respect to one another. A framework for capturing this dual relativity which builds on music theoretic concepts is tone interval theory (Dilley 2005). This theory proposes that the phonological primitives are abstractions of frequency ratios termed tone intervals. The dual relativity of tone is captured in the fact that tones may be defined relative to other tones in sequence or relative to a tonic level based on the speaker's own pitch range. This framework addresses several outstanding problems in the literature. First, it eliminates the problems with overgeneration and indeterminacy associated with earlier proposals. Moreover, it provides a unified framework for describing interactions between tonal and metrical structures. In addition, it provides a phonological account of a number of outstanding phonetic facts regarding the

---

[20] In the case of a very low fall at the end of the phrase, I assume that a large pitch interval is implemented phonetically: $I_{5,6} << 1$.

[21] It is not clear how this contour would be described under the autosegmental-theory inspired ToBI system for intonation transcription (Pierrehumbert 1980, Beckman and Ayers-Elam 1997). By contrast, the tone interval framework accounts for this pattern by characterizing unstarred tones as independent entities which show coordinated alignment with respect to starred tones, resulting in "tritonal" configurations. Similar configurations have been noted by Grice (1995), who proposed tritonal pitch accents for English.

behavior of so-called bitonal accents. Finally, it provides the beginnings of an explanation for how the phonological representation may be perceptually recovered from phonetic input.

References

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics, 26*, 3-25.

Beckman, M., & Ayers-Elam, G. (1997). Guidelines for ToBI labeling, version 3.0: The Ohio State University.

Bickmore, L. S. (1995). Tone and stress in Lamba. *Phonology, 12*, 307-341.

Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed., pp. 215-264). San Diego: Academic Press.

Clark, M. M. (1978). *A dynamic treatment of tone, with special attention to the tonal system of Igbo.* Ph.D. dissertation, University of Massachusetts at Amherst.

de Lacy, P. (2002). The interaction of tone and stress in Optimality Theory. *Phonology, 19*(1), 1-32.

de Saussure, F. (1966). *Course in General Linguistics*. New York: McGraw-Hill Book Company.

Dilley, L. C. (2005) *The phonetics and phonology of tonal systems.* Ph.D. Thesis, MIT.

Dilley, L., Ladd, D. R., & Schepman, A. (2005). Alignment of L and H in bitonal pitch accents: Testing two hypotheses. *Journal of Phonetics*, *33*(1), 115-119.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49*(2), 524-531.

Dowling, W. J., & Harwood, D. L. (1986). *Music Cognition*. Orlando, FL: Academic Press, Harcourt Brace Jovanovich Publishers.

Goldsmith, J. (1976). *Autosegmental phonology.* Ph.D. dissertation, MIT, Cambridge, MA.

Grice, M. (1995). Leading tones and downstep in English. *Phonology, 12*, 183-233.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395-421.

Howard, D., Rosen, S., & Broad, V. (1992). Major/minor triad identification and discrimination by musically trained and untrained listeners. *Music Perception, 10*(2), 205-220.

Inkelas, S., & Zec, D. (1988). Serbo-Croatian pitch accent: the interaction of tone, stress, and intonation. *Language, 64*, 227-248.

Krumhansl, C. L. and Kessler, E. J. (1982) Tracking the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*, 334-368.

Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant “segmental anchoring” of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America*, *106*(3), 1543-1554.

Laniran, Y. O., & Clements, G. N. (2003). Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics, 31*, 203-250.

Leben, W. R. (1973). *Suprasegmental phonology*. Ph.D. dissertation, MIT, Cambridge, MA.

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oerhle (Eds.), *Language Sound Structure* (pp. 157-233). Cambridge, MA: MIT Press.

Manfredi, V. (1993). Spreading and downstep: prosodic government in tone languages. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 133-184). Berlin, New York: Mouton de Gruyter.

Odden, D. (1995). Tone: African languages. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 444-475): Blackwell.

Perlman, M., & Krumhansl, C. L. (1996). An Experimental Study of Internal Interval Standards in Javanese and Western Musicians. *Music Perception, 14*(2), 95-116.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation.* Ph.D. dissertation, MIT, Cambridge, MA.

Pierrehumbert, J., & Beckman, M. (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Schellenberg, E. G. and Trehub, S. E. (2003) Accurate pitch memory is widespread. *Psychological Science*, *14*, 262-266.

Siegel, J. A., & Siegel, W. (1977). Categorical perception of tonal intervals: Musicians can't tell *sharp* from *flat*. *Perception & Psychophysics, 21*(5), 399-407.

Snider, K. (1999). *The Geometry and Features of Tone.* Dallas: Summer Institute of Linguistics and the University of Texas at Arlington Publications in Linguistics 133.

Stewart, J. M. (1993) Dschang and Ebrié as Akan-type total downstep languages. In H. van der Hulst and K. Snider (eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 185-244): Berlin: Mouton de Gruyter.

Sundberg, J. (1987) *The Science of the Singing Voice.* Northern Illinois University Press.

Williams, E. S. (1971/6). Underlying tone in Margi and Igbo. *Linguistic Inquiry, 7*, 463-484. [Ms. written in 1971].

Williamson, K. (1972). *Igbo-English Dictionary*. Benin City: Ethiope Publishing Corp.

Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter- and intra-talker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research, 46*, 413-421.