CHAPTER 1

Noise and Timing

Metrical stress theory (Liberman and Prince 1977) is concerned with three types of prominence: phenomenal, metrical, and structural. These are the same three types addressed in theories of musical rhythm (Lerdahl and Jackendoff 1983). Phenomenal prominence is based on acoustic contrasts; metrical prominence is based on temporal organization; and structural prominence is based on the grouping of linguistic or musical events. The three types are distinct, but they are intimately connected.

1.1 Contrasts

A *phenomenal prominence* is the most salient aspect of an acoustic contrast. In music, a phenomenal prominence, or *accent*, is most often a note that is louder than surrounding notes, longer than surrounding notes, or higher than surrounding notes. Musical accent can also result from other types of acoustic contrasts, however, such as the contrast between the timbres of different musical instruments. A phenomenal prominence in natural language, a *stress* or an *accent*, is similarly the most salient aspect of an acoustic contrast. The usual suspects are the same as they are in music – greater amplitude, longer duration, higher pitch – but stress also frequently arises from contrasts along other dimensions, such as aspiration, vowel quality, or voicing.

It is relatively uncontroversial that the acoustic contrasts used to create stress are not uniform from language to language.[1] Some languages, for example, rely primarily on contrasts in vowel length to create stress. Choctaw (Nicklas 1972, 1975) is one such language. Every even-numbered syllable from the left in Choctaw, except the final syllable, contains a long vowel. Long vowels are the salient acoustic phenomena that constitute stresses. The forms in (1) are combinations

[1] See Gordon and Roettger (2017) and Roettger and Gordon (2017) for a survey of studies on the acoustic correlates of stress in a wide range of languages. See also Lunden and Kalivoda (2015).

of /pisɑ/ 'to see,' /hɑbinɑ/ 'to receive a gift,' /či-/ 'you (object),' /-či/ 'causative,' and /-li/ 'I (subject).'

(1) a. čipi:sɑči:li 'you (obj.)-see-CAUS.-I (subj.)'
b. čihɑ:binɑ:čili 'you (obj.)-receive a gift-CAUS.-I (subj.)'

Other languages rely primarily on contrasts in pitch to create stress. In South Conchucos Quechua (Hintz 2006), the initial syllable and every even-numbered syllable from the right has a higher pitch. Higher pitch, indicated in (2) with an "H," is the salient acoustic phenomenon that constitutes stress.

(2) a. H H imɑkunɑ 'things' b. H H H tushukunɑqą 'dancers'

Other languages rely on other salient phenomena, or combinations of salient phenomena, to create stress.

While it is uncontroversial that the acoustic contrasts used to create stress vary from language to language, it is less often acknowledged that stress-creating contrasts are not necessarily uniform within a language or even for different positions within the same form. This makes it impossible to answer the question of what stress is from an acoustic perspective if one insists that the answer take the form of some invariant acoustic event. This is true for language generally, but it is also true for many, perhaps most, languages individually.

Consider the case of English. Hayes (1995:5–23) discusses a number of phenomena that are diagnostic of English stress. I will focus on just two of these, vowel quality and nuclear intonational tones, but interested readers should refer to Hayes's discussion for a broader, more thorough treatment.[2] The creation of stress with vowel quality is a particularly interesting situation. Contrasts in vowel quality are essentially contrasts in timbre, and its use represents a clear departure from the usual suspects: pitch, length, and loudness. The use of tone in English is also interesting. We have already seen an example of the use of tone in South Conchucos Quechua, but the English case is quite different from the South Conchucos Quechua case.

[2] Hayes's treatment also differs conceptually from the one presented here. In Hayes's conception, vowel quality and the attraction of nuclear intonation tones are diagnostics of stress rather than being stress themselves. Consulting Hayes's discussion will not only give readers a broader view of the empirical terrain but it will also acquaint them with an alternative conceptual perspective.

In English, the prominent full vowels [eɪ, ɛ, æ, ɑ, ʌ, ʊ, u] contrast with the much less prominent reduced vowel [ə].[3] The prominent full vowels always constitute stressed positions, and the reduced vowel always constitutes an unstressed position. It is common in English to see the prominent full vowels in a perfectly regular alternation with the reduced vowel, as in *Pascagoula*, *Susquehanna*, and *Achafalaya*.[4] It is also common to see them in a slightly less perfect alternation, as in *abracadabra*.

Abracadabra is actually a fairly unique case in English in that one single vowel quality creates its multiple stressed positions and a different single vowel quality creates its multiple unstressed positions. The prominent full vowel [æ] appears in the first and fourth syllables, the two stressed syllables. The less prominent reduced vowel [ə] appears in the second, third, and fifth syllables, the three unstressed syllables.

(3) *abracadabra* [ˌæ.brə.kʰəˈdæ.brə]

More typical is the situation where the multiple stresses are created by different prominent vowels. In *Pascagoula*, (4a), the [æ] in the first syllable and the [u] in the third alternate with [ə] in the second and fourth. [æ] creates one stress, and [u] creates the other. In *Susquehanna*, (4b), the [ʌ] in the first syllable and the [æ] in the third alternate with [ə] in the second and fourth. [ʌ] and [æ] both constitute stresses.

(4) a. *Pascagoula* [ˌpæs.kəˈgu.lə]
b. *Susquehanna* [ˌsʌs.kwəˈhæ.nə]
c. *Achafalaya* [əˌʧæ.fə.ˈlɑɪ.ə]

Finally, in *Achafalaya*, (4c), the [æ] in the second syllable and the diphthong [ɑɪ] in the fourth alternate with [ə] in the first, third, and fifth syllables. [æ] constitutes one stress, and [ɑɪ] constitutes the other.[5]

In addition to the prominent vowels [eɪ, ɛ, æ, ɑ, ʌ, ʊ, u] and the less prominent vowel [ə], there is a set of vowels of intermediate prominence: [ɪ, i, oʊ]. These vowels may constitute stress or not depending on the context.

[3] The classification of the vowels follows that of Hayes (1995). Hayes notes that his classification is based on his own speech and that there is likely significant variation among speakers and dialects.

[4] The reader may wonder why I am illustrating common English phenomena with fairly uncommon lexical items. I typically use borrowed toponyms for examples of individual English words because they do not have internal morphological boundaries that might influence the location of stress.

[5] I have recently run across two alternate pronunciations for *Achafalaya*: [ˌæ.ʧə.fəˈlɑɪ.ə] and [ˌæ.ʧəˈfæl.jə].

For example, when [i] is final with a prominent vowel in the preceding syllable, like the final [i] in *Chattahoochee*, (5a), [i] constitutes a stressless position. When [i] is flanked by [ə], however, as in *Monongahela*, (5b), [i] constitutes a stressed position.

(5) a. *Chattahoochee* [ˌʧæ.rəˈhu.ʧi]
b. *Monongahela* [məˌnɑŋ.gəˈhi.lə]

Two instances of [oʊ] flank the diphthong [ɑɪ] in *Ohio*, (6a). In initial position, [oʊ] constitutes a stressed position. In final position, with the prominent vowel [ɑɪ] in the preceding syllable, [oʊ] constitutes a stressless position. When [oʊ] is itself flanked by reduced vowels, however, like the [oʊ] in *Apalachicola*, (6b), [oʊ] constitutes a stressed position.

(6) a. *Ohio* [ˌoʊˈhɑɪ.oʊ]
b. *Apalachicola* [ˌæ.pəˌlæ.ʧəˈkʰoʊ.lə]

In the examples above we can see that the vowel quality contrasts that create stress in English are not always the same contrasts. It is not the case that there is a single vowel quality that constitutes a stressed position and a single vowel quality that constitutes a stressless position. There is actually a range of vowel qualities that always constitute stress, a single vowel quality that always constitutes stresslessness, and a small set of vowel qualities that can constitute either stress or stresslessness depending on the context.

The second phenomenon that constitutes stress in English is the nuclear tones of intonational tunes (Liberman 1975; Pierrehumbert 1980). Intonational tunes are tonal sequences that correspond to a particular meaning, the most well-known being the *declarative* tune and the *question* tune. Each tune contains a pair of edge tones that attach to the initial and final edges of the tune's domain. Each tune also includes one or more nuclear tones between the edge tones. The nuclear tones attach to particular syllables. Since the nuclear tones contrast in pitch with the preceding and following tones, we can interpret the nuclear tones as stress. The two tunes are constructed using three tones: H (high), M (mid), and L (low). The sequence of the declarative tune is M H* L. The edge tones are M and L, and the nuclear tone is H*. The sequence of the question tune is M L* H. The edge tones are M and H, and the nuclear tone is L*.

Though they are usually thought of as being characteristic of utterances containing longer phrases, intonational tunes are also associated with shorter utterances, even utterances of individual words. Consider the cases of *Nishnabotna* and *Tangipahoa*. In *Nishnabotna*, the [ɪ] in the first syllable and the [ɑ] in the third syllable constitute stressed positions.

When associated with the declarative tune, as in (7a), the nuclear H* tone attaches to the syllable containing [ɑ]. When associated with the question tune, as in (7b), the nuclear L* tone attaches to the same syllable.

			M	H* L		M	L* H
(7)	*Nishnabotna*	a.	[ˌnɪʃ.nəˈbɑt.nə]		b.	[ˌnɪʃ.nəˈbɑt.nə]	

In *Tangipahoa*, the [æ] in the first syllable and the [oʊ] in the fourth constitute stresses. When *Tangipahoa* is associated with the declarative tune, as in (8a), the nuclear tone H* attaches to the syllable containing [oʊ]. When *Tangipahoa* is associated with the question tune, as in (8b), the nuclear tone L* attaches to the same syllable.

			M	H* L		M	L* H
(8)	*Tangipahoa*	a.	[ˌtʰæn.ʤi.pəˈhoʊ.ə]		b.	[ˌtʰæn.ʤi.pəˈhoʊ.ə]	

In the examples in (7) and (8), we can see how stress based on vowel qualities combines with stress based on nuclear tones to create a three-way system of phenomenal prominence in English. Vowel quality contrasts distinguish positions with secondary stress from positions that are unstressed. Since the nuclear tone attaches to just one of the vowel quality stresses in a word, it has the effect of elevating it to the primary stress. The nuclear tone, then, distinguishes the position with primary stress from positions with secondary stress and positions that are unstressed.

Stress is not an acoustically unified phenomenon, then, in the sense of always being constituted by the same acoustic features. This is true from language to language, but it is also often true within a single language. Stress *is* an acoustically unified phenomenon; however, in the sense that it is always created by the most salient aspect of an acoustic contrast. This is a key part of our overall understanding of stress – and there will be more discussion of phenomenal prominence below and in subsequent chapters – but it does not really give us any insight into why languages have stress in the first place or why they distribute it in their forms in the particular ways that they do. To understand the presence of stress in natural language and to understand its distribution, we have to understand its function. To understand its function, we have to understand its relationship to metrical prominence and structural prominence.

1.2 Purposes

Though there does not seem to be a consensus of what stress actually contributes to natural language, there does seem to be a strong suspicion that it is not merely decorative. Two possible functions for stress

have been identified in the literature: *distinctiveness* and *delimitativity*.[6] Accentual systems can be *distinctive* in that they can be used to encode differences in meaning (Bloomfield 1933; Trager 1941; Hockett 1958; Trubetskoy 1969). If two words are otherwise identical, the difference in the position of their accents is the contrast that indicates that they have different meanings. Consider, for example, the pair of English words in (9). Although the meanings of the two words are related, the part of speech changes with the position of the primary stress. With primary stress on the initial syllable, *trusty* is an adjective meaning "trustworthy, reliable"; with primary stress on the final syllable, *trustee* is a noun meaning "a person responsible for administering property for the benefit of someone else."

(9) TRUSty *vs.* trusTEE

For a stress to be distinctive, its distribution cannot be predictable based on other properties of the word. Like other contrastive structures, the position of a distinctive stress cannot be derived but must be part of the underlying representation.

In contrast, predictability is essential to the second possible function of stress systems: *delimatativity*. Stress systems can be *delimitative* in that they can be used to mark the boundaries of accentual domains (Trubetskoy 1969). For example, Osumi (1995) describes stress as reliably initial in Tinrin. As (10) illustrates, when stress is encountered, it is always at the beginning of the word.

(10)
a.	ˈhuːe	'white'
b.	ˈhusɑːu	'sometimes'
c.	ˈʂɯvehɑɽu	'to like'

Poppe (1962) describes stress as reliably final in Uzbek. As (11) illustrates, when stress is encountered, it is always at the end of the word.

(11)
a.	ɑitˈdi	'he said'
b.	kitoˈbim	'my book'
c.	ɑŋlɑdiˈlɑr	'they understood'

[6] See Beckman (1986) for a thorough discussion of the various functions of stress and accent that have been proposed in the literature. A third possibility often cited as a function of stress systems is *culminativity* (Trubetskoy 1969). Most, perhaps all, accentual domains seem to be *culminative* in that each has a single (most prominent) instance of the relevant accentual type. Rather than being an independent function, as Beckman (1986) argues, culminativity seems to be a property that most, perhaps all, accentual systems have in common.

For stress to be delimitative, the location of the stress and the location of the boundary must be fixed relative to one another. The location of the boundary must be predictable given the position of stress, and the position of stress must be predictable given the location of the boundary. Like other non-contrastive structures, the position of a delimitative stress would be derived rather than stipulated in the underlying representation.

Two obstacles prevent these traditionally recognized roles – distinctiveness and delimitativity – from providing a foundation for a general account of stress. The first is that some stress systems appear to be neither distinctive nor delimitative. In the case of English, for example, the portion of the lexicon where primary stress is contrastive is fairly small, so it is difficult to see English as a true distinctive accentual system. Primary stress in other portions of the lexicon appears to be positioned by rule, but different rules govern stress in different portions. As a result, the position of the primary stress relative to the boundaries of accentual domains is somewhat variable, so it is also difficult to see English as a true delimitative system.

The second obstacle is that the distinctiveness and delimitativity are fundamentally incompatible. A stress can only be delimitative if its position is predictable, but it can only be distinctive if its position is not predictable. Because they are fundamentally incompatible, the distinctiveness and delimitativity functions cannot form the basis for an understanding of the role of stress in natural language generally.

Of course, it might simply be the case that stress is not actually a functionally unified phenomenon: there are distinctive stress systems, delimitative stress systems, and systems that are effectively neither, and any similarities between them are merely coincidence. This is not a view that I will consider here in any serious way. Instead, I will consider a third possible role for stress, *indicativeness*, that is similar to the role that accent plays in music. Under this view, the purpose of stress is to indicate a language's temporal organization – to indicate the representation of time against which the events of an utterance are structured. Whether a stress system is distinctive or delimitative (or effectively neither), it always indicates temporal organization. Before we can understand the indicativeness role of stress, however, we need to consider how language represents time.

1.3 Grids

The *metrical grid* is a device for representing temporal organization. It has proved to be particularly useful in two contexts: the analysis of stress patterns in natural language (Liberman 1975; Liberman and Prince 1977;

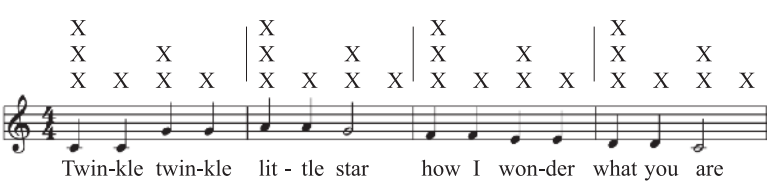
Prince 1983; Selkirk 1984) and the analysis of meter and rhythm in music (Lerdahl and Jackendoff 1983). On the page, a metrical grid is a collection of entries, each marked with an "X," representing the boundaries of temporal intervals. As (12) illustrates, the entries are organized into rows – typically referred to as *levels* – and columns. Entries on the same level represent boundaries following one another in time, separating intervals that occur in a sequence. Entries in the same column represent interval boundaries on different levels that occur at the same point in time.

(12)
X X X X
X X X X X X X X
X X X X X X X X X X X X X X X X

At a minimum, grid representations are representations of *temporal order*. They locate temporal intervals and boundaries relative to other temporal intervals and boundaries. The left-to-right distribution of columns on the page reflects the earlier-to-later distribution of the temporal boundaries that column entries represent. Similarly, the left-to-right distribution of the gaps between column entries reflects the earlier-to-later distribution of the temporal intervals that the gaps represent.

The grid can also capture durational relationships between temporal intervals. If we know that the gaps on one level of the grid represent isochronous intervals, then we know the *relative temporal duration* of the intervals represented by the gaps on higher levels. The grid in (12), for example, might be used to represent four measures of 4/4 time.

Twinkle, Twinkle, Little Star might be played in a way that is compatible with a range of time signatures, but 4/4 time is one of the most commonly specified. In Western musical notation, the upper numeral of the time signature indicates the number of beats in a measure, and the lower numeral indicates the note value for each beat. For 4/4 time, then, there are four beats per measure, and each beat has the length of a quarter note. The specification of a time signature is essentially an instruction to the performer to play or sing the piece in such a way that its accents are consistent with a particular metrical grid. By convention, when 4/4 is specified, the relevant grid has the structure in (13). Lines are added to the grid in (13) to indicate the division into measures.

(13)


In the grid in (13), each entry on level one represents the onset of a quarter note beat and each gap represents the duration of a quarter note. Since each gap on level two is coextensive with two gaps on level one, and since each gap on level one represents the duration of a quarter note, we know that each gap on level two represents the duration of a half note. Similarly, since each gap on level three is coextensive with two gaps on level two, and since each gap on level two represents the duration of a half note, we know that each gap on level three represents the duration of a whole note.

If we know the actual length of the intervals represented by the gaps on one level of the grid, then we would also know the *actual temporal duration* of the intervals represented by the gaps on higher levels. For example, if we decided that quarter notes should be played at the rate of 1 per 500 milliseconds (120 beats per minute), then we would know that each of the level-one gaps in (13) represents an interval of 500 milliseconds, each of the level-two gaps an interval of 1 second, and each of the level-three gaps an interval of 2 seconds.

While the representation of temporal order and temporal duration is an important issue in the analysis of music and natural language, use of the grid in both contexts has focused on the grid's ability to represent *temporal prominence*. The columns of the grid form a quantity-based hierarchy that can be interpreted as a hierarchy of prominence. Columns where a greater number of temporal boundaries coincide are more prominent than columns where a lesser number of boundaries coincide. Though other patterns are possible, the typical prominence relationships within a measure of 4/4 time are those illustrated in (12) and (13). Three temporal boundaries make up the first column of each measure; two temporal boundaries make up the third column; and one temporal boundary makes up each of the second and fourth columns. Since the first column contains more boundaries than the third, the first is more prominent than the third. In turn, since the third column contains more boundaries than the second and fourth columns, the third is more prominent than the second and fourth.

There are a few key points to keep in mind about metrical prominence. The first is that metrical prominence depends solely on the internal structure of the grid itself; it does not rely on anything external to the grid. The first column in (12) and (13), for example, is more prominent than the third column simply because more temporal boundaries coincide in the first than the third and for no other reason. Second, metrical prominence is a relationship between grid columns and not a relationship between anything that might correspond to grid columns, such as

the beginning of a note or the beginning of a syllable. Third, metrical prominence has no inherent acoustic properties, such as greater length, greater amplitude, or higher pitch. Finally, though metrical prominence is not identical to phenomenal prominence – accent or stress – phenomenal prominence can be coordinated with metrical prominence sufficiently for the former to indicate the latter.

1.4 Identification

If we happen to know the time signature of a piece of music, either because we have access to the sheet music or we are able to ask someone who already knows, then we know the piece's temporal framework. If we do not know the time signature, however, we can still usually determine a piece's temporal framework by listening to the piece and paying attention to the cues provided by phenomenal prominences. In the context of natural language, we typically have no other option than to rely on the cues provided by phenomenal prominences. English words, for example, do not come with time signatures. When spoken, however, they do come with intonational tunes, contrasts in vowel quality, and a number of other cues. Collectively, these cues are perceived as stress, and stress is the indicator of a temporal framework.

While all native speakers are able to use stress cues to identify metrical patterns in the context of language use, not all native speakers are equally good at using stress cues to consciously identify and describe metrical patterns. This should not be surprising. Using stress cues to identify a metrical pattern in the context of language use is a purely linguistic task, but using stress cues to consciously identify a metrical pattern and describe it is not a purely linguistic task. Hayes (1995:28–29) argues that metrical structure has a perceptual reality, however, even for speakers who are not able to identify metrical patterns consciously. He describes a demonstration something like the following.

A speaker is given a word and asked to pronounce it while tapping or clapping a certain number of times.[7] She will find it natural to produce some number of taps but not others. For example, if she is given the word *abracadabra*, she will find it natural to tap five, two, or one times, as in (14). Trying to tap four or three times, however, results in confusion.

[7] I should note that Hayes's tapping demonstration is also not a purely linguistic task and some people are better at it than others. For many people who cannot consciously identify metrical patterns based on stress cues alone, however, the tapping exercise can be quite helpful.