Prosodic Boundaries in Singing:
An exploration of the temporal effects of linguistic structure in western music

John Clayton

Jelena Krivokapić & Ian Quinn

Submitted to the faculty of the Department of Linguistics
in partial fulfillment of the requirements for the degree of
Bachelor of Arts

Yale University
May 1st, 2013

Abstract

The practice of singing dates back to the beginnings of human history and represents the convergence of two similar systems: language and music. Within these two systems, the fields of linguistic prosody and musical meter each use the temporal domain to express their underlying structure. At times, these expressional requirements differ, necessitating either compromise or dominance of one system over the other. For instance, in the case of misalignments between the metrical stress of music and of language, musical stress was found to be dominant (Palmer & Kelly, 1992).

Through an experiment carried out with participants singing a prepared text, this study examined the effect that prosodic boundaries have when misaligned with musical meter to examine whether prosodic boundaries will cause pre-boundary lengthening despite the temporal requirements of expressing musical meter. Results suggest that the necessity for final lengthening at a prosodic boundary is great enough to displace the subsequent beats temporally from their normal alignment due to musical meter.

INTRODUCTION

The study of language and the study of music have long been linked because of their obvious similarities. Both language and music represent methods of communication. Both language and music are expressed and perceived through the auditory or sonic domain, although the content and character of their meaning differ widely (Tan et al., 2010). It is in the process of

singing, however, that the two systems converge. This convergence of music and language has long caused authors to claim one system as the progenitor of the other, whether that be singing-to-speaking or vice-versa. In his recent and popular book, *The singing Neanderthals: The origins of music, language, mind, and body*, for example, Steven Mithen proposes that music played an instrumental role in the development of human culture and more specifically in the development of language (2007). Others, however, question whether one led to the other, citing a variety of evidence both in favor of and against the claims that language and music are processed together by the same mental faculties (Tan et al., 2010). Regardless of whether singing arises from a convergence of two separate mental faculties or one multi-talented mental faculty, many cognitive scientists, linguists, and musicologists have examined the overlap of music and language (Lerdahl & Jackendoff, 1983; Halle, 2004; Palmer & Hutchins, 2006; Palmer et al., 2001; Patel & Daniele, 2003).

In the examination of music and language, their many similarities frequently make the comparison and differentiation of the two systems difficult. Both have structural hierarchy which is expressed through cues primarily in the auditory domain. Indeed the categorization and naming of those cues may cause frustration due to the overlap of terms and concepts. In both systems, information may be conveyed based on pitch cues, loudness cues, sound quality cues, and temporal cues (Ladefoged, 2005; Tan et al., 2010). This project will focus solely on temporal cues, which integrate the information provided by the other cues and more importantly are defining characteristics of both musical meter and linguistic prosody. Linguistic prosody describes the hierarchical organization of syllables, stress, and boundaries higher than the segmental level (Ladd & Cutler, 1983; Liberman & Prince, 1977; Selkirk, 1984), and musical meter describes the hierarchical organization of beats, stress, and boundaries in musical phrases

(Tan et al., 2010; Lerdahl & Jackendoff, 1983). In particular, this paper will focus on the conflicting necessities of the tendency towards isochrony of musical meter and the pre-boundary lengthening of prosodic boundaries—two effects both measurable in the temporal domain.

Stress

The difficulty of defining musical meter beyond calling it the hierarchical organization of stress and beats quickly becomes apparent. Throughout our daily lives, we are inundated with auditory input that we interpret as being at regular time intervals, particularly in the field of music. The human brain excels at perceiving "rhythms" in even the most unrelated sounds. These rhythms are defined for the purpose of this paper as these perceived and prescribed temporal ratios between the onsets of acoustic events (Tan et al, 2010). This definition of rhythm contains multiple important aspects. First, the temporal relationship of the events need not be orderly or repetitive; thus, the remembered time intervals between the crashes of a falling tray of china represent a rhythm just as much as the regular clicks of a clock. Second, the particular length of any or all of the time intervals is not prescribed; what the rhythm prescribes is the ratio of lengths between the intervals. Thus, a regular metronome click occurring exactly every three seconds represents the same rhythm as a regular metronome click every one second because the length ratio between events is the same, while the overall speed of the events is different. And third, the defining units of a rhythm are merely the temporal intervals between onsets of acoustic events and not the length of the acoustic events themselves. A regular series of short snaps represent the same rhythm as a regular series of long-ringing gong strikes so long as the intervals between the initiations of each event hold the same ratio.

When a rhythm may be described hierarchically and cyclically as regular, it then becomes a musical "meter" characterized by predictable "beats" and "tempos" (Tan et al, 2010). The beat is defined as the most salient relative unit of duration within a meter and is usually the unit of time to which listeners will tap their feet or dance. As was stated before, rhythm represents the relative intervals between acoustic events, which necessarily means that if the speed of events increases or decreases while maintaining their same relative interval ratios, the rhythm must be the same. These differing rates of rhythmic production are called tempos, and are usually defined as the expected regular time between beats. Tempos are perceptually defined as being isochronous, having consistent time intervals between the onsets of the constituent parts (Tan et al, 2010). This modeled isochrony of tempo, however, does not give us the entire story.

Many musical and non-musical rhythms that we perceive as being isochronous are actually flexible and non-isochronous (Large & Jones, 1999). These temporal fluctuations are frequently used expressively in what are perceived as strictly metered pieces (Large et al., 1995). So if music is not isochronous, how can we, as listeners, perceive it as having a "strict tempo" and pick out so easily where the beats fall? The answer proposed by Large and Jones (1999) is the use of the dynamic attending theory of rhythm in which dynamic oscillators create the listeners' expectations of when the next beat or stress should fall. This dynamic model makes use of an entrainment model, which states that oscillators tend towards synchronization even when beginning out of synch (Large & Jones, 1999). These internal oscillators, or "attending rhythms," which create expectations about when future events will occur, are isochronous. Attending rhythms "attend" the external patterned stimuli to which they then entrain, allowing the listener to continue previously established rhythms, detect fluctuations within those rhythms, and adjust as rhythms continuously change. With a single oscillator, a listener may effectively

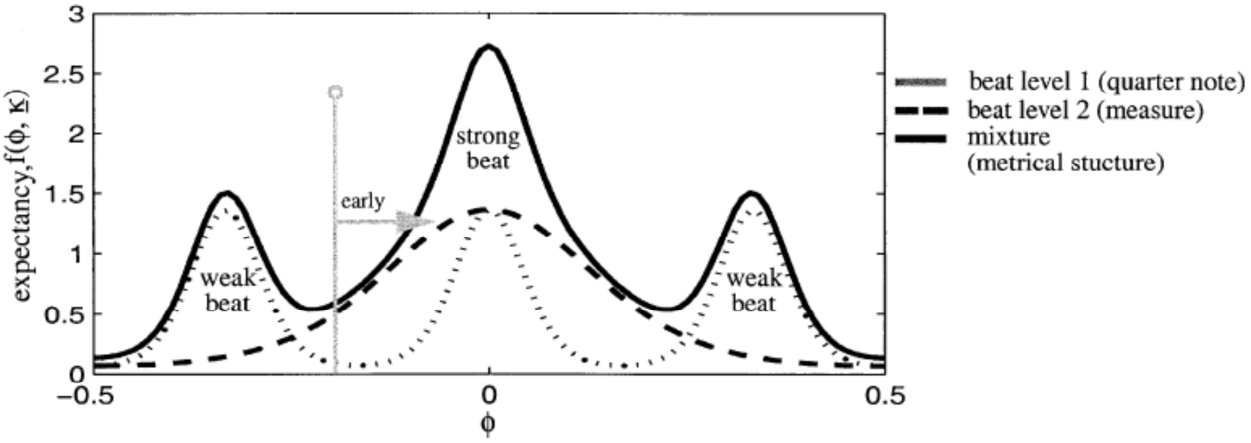


Figure 1: A diagram from Large & Palmer (2002) modeling a ternary meter (3 beats : 1 measure) with von Mises stresses. The dotted line represents the attending stress on each quarter note beat and the dashed line represents the attending stress on each dotted half note measure. The solid line shows the compound stresses of the two levels of the metrical hierarchy.

model and follow simple rhythms. Many musical meters, however, represent more complicated hierarchies of stress and timing. Dynamic theory therefore proposes that the listener has not one, but multiple oscillators that work in concert to represent these meters. If we begin with an oscillator at the measure with a set period and then add another oscillator at the beat level that oscillates at three times the frequency as the beat level such that their peaks coincide, they will create a compound oscillation representing a ternary metrical structure, seen in Figure 1.

The amplitude of the peaks of this compound wave is said to represent stress in musical meter, resulting in a two-tiered metrical hierarchy, which listeners may hold in their head and produce. Take then, for instance, the most common meter in all of Western tonal music, $\frac{4}{4}$ or common time (Tan et al., 2010) as seen in Figure 2. Common time in its simplest form represents a three-oscillator system. The oscillator with the lowest frequency peaks at the beginning of each measure, followed by an oscillator peaking on beats 1 and 3, then an oscillator peaking on every beat. This gives beat 1 highest stress prominence, followed by beat 3, then beats 2 and 4. In this way, the listener's mental model of a rhythm may be organized by periodic stresses of varying amplitude or "strength." This strength of metrical stresses is represented by increasing the volume and length of the beat but not, in most cases, the pitch (Drake & Palmer, 1990; Sloboda, 1983). These volume and pitch effects are disregarded here in any case as we examine the temporal lengthening effects. This lengthening of "strong beats" over "weak beats" naturally leads to metrical non-isochrony needed to represent the meter. However, since these oscillations are modeled isochronously at the tempo of the music, the global beat length average equals the isochronous rate of the tempo and the musical meter shows a **tendency towards isochrony**. This gives the meter the constant sense of forward motion. For a meter to be stable, it must articulate these hierarchical stresses while maintaining its alignment with the tempo.

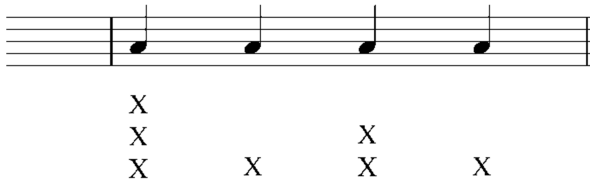


Figure 2: One measure of common time containing four beats with recurring stresses at the frequency of once every measure, once every half-measure, and once a beat.

With musical meter and stress defined in terms of temporal cues, the next question pertains to how this may be done for linguistic stress. Syllables may be prosodically stressed by

increasing the volume, pitch, or length of the syllable, depending upon the language in question (Lehiste, 1972). In 1977, Liberman & Prince argued that prosodic stress and boundaries should be modeled in branching tree structures similar but not identical to those used in the modeling of syntactic structure. Using this method, the assignment of prosodic stress, intonation, rhythm and prosodic boundaries could be modeled more accurately than by merely comparing them only to syntactic structure. In 1986, Beckman & Pierrehumbert proposed that below the level of a sentence, branching hierarchical trees contain intonational phrases, intermediate phrases, and prosodic words which possess prosodic stress assigned hierarchically on each constituent level. Kim & Cole (2005, 2006) found that while language does not demonstrate strict isochrony on the word or syllable level, the stress foot, a hierarchical group containing one prosodic stress, does show a tendency towards isochrony.

This similarity between musical tempos and prosodic feet particularly in poetry has also led to studies examining prosodic feet only in terms of musical terminology and notation (Lerdahl, 2001). With both of these tendencies towards isochrony interacting in poetic texts set to musical meter, one might expect to see significant regularity in the lengths of both segments and syllables when sung to a strict musical meter and tempo. To say this another way, the application of a strict musical meter over a text prevent fluctuations in segment and syllable lengths due to varying speech rates, causing singers to produce a given word at the same duration in multiple tests. This proposition shall be evaluated in Hypotheses 1 & 3 stated later.

Boundaries

Alongside the use of stress and isochrony in music to create musical meter, this paper examines the effects of prosodic boundaries, which are the cues defining categories and phrase

structure in prosody. The difference between these boundaries, as in the case of stress, appears most prominently in the temporal domain. Prosodic boundaries differ in manifestation and strength and appear perhaps most clearly through pause durations. These prosodic boundaries also demonstrate their relative strength by lengthening the ends of phrases preceding pauses in speech (Lehiste, 1977). The strength of these boundaries are determined by discourse structure (Swerts, 1997), syntactic structure (Selkirk, 1980; Ferreira, 1991), prosodic structure (Gee & Grosjean, 1983), phrase length (Krivokapić, 2007), and speech rate (Trouvain & Grice, 1999). Wightman et al. (1992) found that this lengthening in prosody is limited mostly to the rhyme of the preceding syllable and particularly to the preceding coda consonant. The marking of the strongest prosodic boundaries may be most easily seen in the placement of commas and periods in written language, particularly those representing clausal or sentential breaks. If you compare the sentences “Tess eats Anne’s beans.” and “No matter what Tess eats, # Anne’s beans all disappear.” the first sentence does not possess a strong syntactic or prosodic boundary after the word “eats,” whereas the second sentence has a clause boundary (the prosodic boundary is represented by a “#”). Through the manipulation of syntax, one may create a clause boundary between two words, which would then be realized as a strong boundary by the prosody. While this first sentence may possess a weaker type of prosodic boundary after the word “eats,” that word shall be defined here a “phrase-medial” for all but the strongest of boundaries.

The strength of prosodic boundaries can also be hierarchically modeled using a method proposed by Beckman & Pierrehumbert (1986). These prosodic boundaries represent the structural boundaries between the branches of the prosodic constituents starting at the level of the word and extending to the level of the level of traditional clausal and sentential boundaries, like those represented by commas and periods. To examine the strength of prosodic boundaries

between the different levels of prosodic hierarchy, Price et al. (1991) asked participants in an experiment to listen to and labeled "break indices" between orthographic words on a scale of 0-6, where 0 represented no break and 6 represented the largest breaks. More specifically within this system, the index 1 was assigned to prosodic word boundaries with no associated special prosodic effects. The index 2 represents a group of words inside of a larger unit, though its exact hierarchical structure remained elusive. Indices 3 and 4 have been assigned to intermediate and intonation phrases respectively in later analyses, such as Wightman et al. (1991). The index 5 was later described as representing the "superior major tone group" by Ladd & Campbell (1991). Index 6 was reserved for the boundaries at the end of sentences. Prosodic boundaries labeled as 4-6 are considered in this system as "strong." An example of such a labeling may be seen in Figure 3. This system was used to develop the 0-4 break index system used in the ToBI Annotation Conventions described in Beckman & Elam (1997).

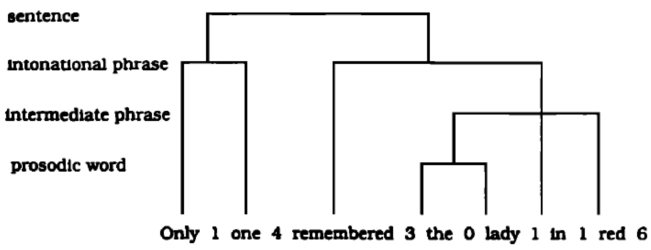


Figure 3: A sample labeling taken from the Price et al. (1991) experiment, found in Wightman et al. (1991).

Just as musical meter and linguistic prosody both possess hierarchical stress, so too do both possess boundary marking. The findings of pre-boundary lengthening in prosody parallel the findings in musical meter that events preceding phrasal boundaries lengthen both locally and globally (Palmer, 1989). These musical boundaries act as cues for the musical phrases described by Lerdahl & Jackendoff (1983).

Stress and boundaries in singing

In the use of temporal lengthening, both musical meter and linguistic prosody converge in the ways they mark stressed beats and syllables. In the case of singing, these two processes overlap. The question then arises how prosodically and musically stressed syllable interact within a singing context. To answer this problem, Palmer & Kelly (1992) conducted a study that examined the effects of matching and mismatching prosodic and musical metrical stress. They found that when prosodic and musical stress aligned, the resulting lengthening was greater than the lengthening of either one alone. In an examination of large corpora of English language songs, they also found that linguistically stressed syllables tended almost always to line up with strong beats. To these authors, this suggested that the stress systems of musical meter and linguistic prosody functioned separately and could then converge or diverge to make stress of greater or lesser strength. They also found, however, that in cases of mismatch between the linguistic and musical stress, the effects of the musical stress overrode the effects of the linguistic stress, implying that in the case of lengthening caused by stress in singing, the need to mark stress in the hierarchy of musical meter took precedence over the need to mark linguistic stress. Another way to understand the stronger lengthening of musically stressed syllables is to say that allowing the same amount of linguistic stress lengthening on a weak beat could destabilize the periodicity and tendency towards isochrony of the musical meter, causing metrical failure in extreme cases.

Linguistic stress, however, is not the only linguistic temporal effect that may interact with musical meter in singing as there is evidence that prosodic boundaries also have an impact on lengthening in music. Durational pauses at prosodic boundaries occur with great frequency when aligned with musical boundaries. In the research of Falk (2009), prosodic boundaries of varying

strengths tended to align with phrasal boundaries within the musical meter of children's songs—a similar finding to the alignment patterns observed for linguistic and musical stresses by Palmer & Kelly (1992). Yet the existence of these shared boundaries again poses the question of how mismatches of prosodic boundaries and musical boundaries play out in the context of musical meter. In particular, if a prosodic boundary appears at a point where no musical boundary occurs, what must be the lengthening effect of a mismatched prosodic boundary, given that the singer is attempting to adhere to the stress marking and tendency towards isochrony of the musical meter? Prosodic boundaries occurring phrase-medial, however, may be reasonably expected to show shorter durational pauses at the prosodic boundary because of tendency towards isochrony in musical meter, unless a piece is occurring at a very slow tempo. Prosodic boundaries, however, might be expected to appear through the lengthening of pre-boundary segments and syllables before stronger prosodic boundaries, since pre-boundary lengthening shares temporal characteristics with both linguistic and musical stress lengthening. This pre-boundary lengthening may also be expected to be less great than in normal speech due to its conflict with the musical meter. However, examining the presence of this pre-boundary lengthening in strictly tempo-ed music is the purpose of this paper.

Given the previously observed dominance of musical meter over prosody in the case of stress mismatches, it stands to reason that prosodic boundaries in the middle of musical phrases might not be expressible in the temporal domain. Experience, however, tells us that singers can express a great deal of syntactic and semantic subtlety through the lyrics of a song. Perhaps then the pre-boundary lengthening in prosodic boundaries could be a functioning disambiguation cue in musical meter. The hypothesis to be tested then is that, if a prosodic boundary occurs in the middle of a musical phrase, it will still have a lengthening effect on the segments and word

before it, and that thus musical meter does not entirely block the production of prosodic boundaries.

If it is the case that prosodic boundaries are expressible in the temporal domain while the structure of musical meter is maintained, then one would expect the following observations, designated as H1-H4, if the hypothesis above is correct:

H1 Segments at word boundaries should be of consistent length in the context of a strict musical meter.

H2 Segments immediately preceding prosodic boundaries will undergo lengthening greater than their normal temporal duration in a strict rhythmic context.

H3 Syllable lengths should be of consistent, though non-isochronous lengths in the context of a strict musical meter.

H4 Syllables immediately preceding prosodic boundaries will necessarily be longer than their normal temporal durations even in a strict rhythmic context because they contain lengthened final segments.

As mentioned before, it is musical meter's tendency towards isochrony that leads to the proposition that, over many different productions of the same phrase, a word or segment may be expected to have similar durations (this is stated in H1 and H3). If syllables and segments are reproduced consistently, then any lengthening cause by prosodic boundaries (H2 and H4) would then be contrary to the natural length of syllables prescribed by the musical meter's stress and tempo, representing a partial bending of the needs of the musical meter to allow for the temporal cues of linguistic prosody. H3 and H4 propose that, if H1 and H2 are true and pre-boundary segments do lengthen, then pre-boundary syllables will lengthen due to the presence of

lengthened pre-boundary segments and that these pre-boundary syllables will not be the domain for compensatory shortening to compensate musically for prosodic boundary lengthening.

METHODS

The Stimuli

This experiment exploits structural ambiguity in the four-syllable target phrase "Tess eats Anne's beans" to create specific prosodic boundaries after specific words. By creating a certain context, "Tess eats Anne's beans" may be manipulated such that the prosody of the contextualized phrase places a prosodic boundary after each constituent of the four-word phrase in turn. This one phrase is used four times so as to control for the lengthening effects of the four-beat structure of the musical meter and to control. From this point forward, each two-line phrase containing a target phrase shall be referred to as a "couplet" so as to disambiguate them from experimental "conditions," which refers to the actual specific placement of the prosodic boundaries within the target phrase. The four separate couplets appear with the target phrase underlined and the prosodic boundaries marked with "#."

Couplet 1: I'm on my way to the Farmer's Market to buy some beans for Tess and Anne.
I realize, if <u>Tess eats Anne's beans, #</u> it will destroy our baking plan.

Couplet 2: Whenever I buy my friends some beans, I always find I have a mess.
No matter what <u>Tess eats, # Anne's beans</u> all find their way into distress.

Couplet 3: I ultimately make this deal that they must keep about the beans:

If once again Tess eats Anne's, # beans are going to be swapped for greens!

Couplet 4: Yet all of this is just a trick, and our dog Max is duping us.

He sneaks behind Tess, # eats Anne's beans and thus creates the whole big fuss.

The words in the target phrase were chosen for their labelability according to guidelines laid out in Turk et al. (2006), specifically by maximizing the number of fricatives in codas before words beginning with vowels.

To test the separate Hypotheses 1-4, these couplets, though gathered in one experiment, are analyzed in two separate ways. In all the following analyses, the independent variable is the presence of a prosodic boundary following a syllable or segment and the dependent variable is the length of syllables and segments. Because H1 & H2 pertain to segmental lengthening before prosodic boundaries, they will be examined in an analysis referred to here as Segmental Length Analysis (SLA). This analysis pertains more directly to the effects of prosodic boundaries on words. H3 & H4, on the other hand, pertain to the length of words and beats before prosodic boundaries, and thus it shall be referred to here as the Beat Length Analysis. This second analysis deals more with the effects of prosodic boundaries on beat lengths (BLA). If pre-boundary beats and words show lengthening effects, then to keep the meter stable, compensatory shortening must happen later on in the phrase to account for the conflicting lengthening due to the prosodic boundary.

In the SLA, the fricatives at the end of “Tess,” “eats,” and “Anne’s” in Couplet 1 shall be used as the no-boundary control, and the fricatives before each boundary in Couplets 2-4 shall be used as the boundary cases for comparison. The length of the fricatives when in a pre-boundary position and when not in a pre-boundary position will be compared to one another to check for lengthening of segments before a boundary. Table 1 below represents the method by which the words will be analyzed. Only underlined fricatives will be examined, and each column contains two fricatives that will be compared, one in pre-boundary position and one not in pre-boundary position. For the evaluation of H1, the consistency of length for fricatives by word will be ascertained, and for the evaluation of H2, the effect of lengthening for fricatives before boundaries may be ascertained.

Table 1

Couplet 1 (Control)	*Tess*	*Eats*	*Anne’s*
Couplet 2	Tess	**eats #**	Anne’s
Couplet 3	Tess	Eats	**Anne’s #**
Couplet 4	**Tess #**	Eats	Anne’s

In the BLA, each couplet contains one pre-boundary word and three phrase-medial words. Each word in the target word has one pre-boundary couplet and three phrase-medial couplets. This creates a 4x4 Latin square such that the length of each pre-boundary word may be evaluated against the same word phrase-medially to check for lengthening of the syllable or beat. This is shown in Table 2. This method of analysis, namely comparison of syllable lengths across couplets, does represent a break from the normal analysis method for prosodic boundaries (comparison of syllable lengths of the words surrounding boundaries.) Because of the tendency towards isochrony provided by the musical meter, however, this comparison of inter-couplet word lengths seems appropriate to check for the word length consistency predicted in H3 and

syllable lengthening predicted in H4. Table 2 represents a vertical comparison of inter-couplet words as opposed to a horizontal comparison of boundary adjacent words.

When analyzing the length of syllables in a musical context, there is a question of whether the lengthening effects on a linguistic word and those on a musical beat differ in their pre-boundary lengthening effects. It also bears remembering that, because all the words in the test phrase are monosyllabic and assigned each to their own beat, the lengths of a "word," of a "syllable," and of a "beat" may be co-indexed with one another; though the lengths of a "word" and of a "beat" may differ as described below. Because of the finding that the onset of musical beats is most closely aligned with the onset of a syllable's vowel (Allen, 1972), all the words in all the couplets will be evaluated both according to their "word length" (the interval from the onset of each word's **initial segment** segments to the onset of the following word's **initial segment**) and to their "beat length" (the interval from the onset of each word's **vowel** to the onset of the following word's **vowel**). If findings concerning the lengthening of linguistic words and musical beats before prosodic boundaries differed significantly, then H3 and H4 must be evaluated separately for both measurements. H3 will be evaluated as to whether words and beats in set metrical positions are of consistent length between couplets, and H4 will be evaluated as to whether words and beats in set metrical positions will lengthen before prosodic boundaries.

Table 2

Couplet 1	Tess	Eats	Anne's	**beans #**
Couplet 2	Tess	**eats #**	Anne's	beans
Couplet 3	Tess	Eats	**Anne's #**	beans
Couplet 4	**Tess #**	Eats	Anne's	beans

These couplets, once made, form an eight line poem necessary to create the environments necessary for each test phrase and to give the experiment the hallmarks of a song. Each couplet

is organized with a poetically metered first line to give the semantic context needed for the changing syntax in the second line containing the target phrase. This first line also contains a rhyming word at the end which rhymes with a word at the end of the second line, thus giving each couplet an element of closure with each rhyme and mapping out a classic antecedent-consequent structure for the musical setting. This extra first line also helps to distract the participant from the continuous repetition of the target phrase in the second line by providing novel semantic material. Because of the amount of context needed for each condition, the couplets must presented in the same order to every subject in order to create a logical "plot," or else the contextualization of the conditions would not make sense. Each couplet was presented separately in order to prevent comparison by the participants between conditions to find similarities.

Each of these couplets was written to fit a simple harmonized melody written to be both memorable and easy to sing. The melody was composed in $\frac{4}{4}$ time because of the centrality of that time signature in Western music. Both the melodic contour and harmony follow common schemes in order to make them easy for the participant. The melody created an overall arc with its peak at the site of the target phrase before descending to its starting point. Its melodic range was a minor 7^{th}, having as its lowest note the fifth scale degree and as its highest, the fourth scale degree. This range of less than an octave makes it easily singable by most trained singers and lies well within the comfortable ranges of basses and altos when in D major and of tenors and sopranos when in G major. The two lines of each couplet line up with four bars each, thus creating an eight-bar phrase of an antecedent-consequent idea where the two musical phrases have similar melodic ideas to cohere them. A breath mark appears between the two musical phrases, and the participants were advised to shorten the beginning of the second line by an

eighth note to make room for a breath so that the singer could make it through each line without breathing during the target phrase.

With the couplet created for this melody, which remained constant throughout each couplet and condition, the target phrase was always positioned in the same metrical and melodic position within the song. As appears in Figure 4, the target phrase always began on the fourth beat of the fifth measure and proceeded through the third beat of the sixth measure of each couplet, thus ensuring that all four beat types were represented in the test phrase and the musical stresses lined with the natural linguistic stresses of "Tĕss éats Ănne's béans." The entire test phrase was sung on the same notated pitch, so that there would be as little effect from pitch as possible on the production of the test words. As a result, the independent variable in each condition would be the location on the prosodic boundaries since the musical pitch, tempo, and location of the words within the rhythmic hierarchy were standardized. This, in turn, means that the acoustic correlate of effective boundary production would be temporal changes in the test words. These inter-word time intervals acted as the dependent variables.

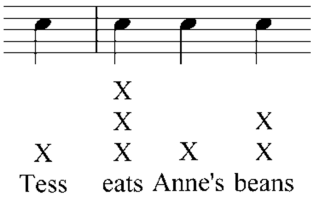


Figure 4

Test Procedure

These music and couplet groups were entered into the musical notation software Sibelius 6 to produce standard notational materials. Each couplet was placed on a page with a musical introduction to orient the listener to the key and to familiarize them with the tempo. The appropriate tempo for the piece was set at 160 beats per minute to align with the jaunty feel of the song. This tempo marking is perceived to be slightly faster on a scale of slow-to-fast tempos by musicians (Tan et al., 2010) but sits within the range of words per minute recommended for

the reading of audiobooks for ease and comprehensibility (Williams, 1998). These pages of music, each with the same melody and harmony but with a different couplet, were printed out in the keys of D major and G major, resulting in eight pages. A sample page is attached at the end of this paper. MIDI files of the harmonic accompaniment were then created in both of the two keys and saved on the test computer.

These participants were brought into a quiet room and asked to stand in front of a computer. Each participant was shown a MIDI file in G major for the sopranos and tenors and D major for the altos and sopranos and asked whether the range and tempo were acceptable. When showing the participant the melody, the experimenter asked the participant to observe all breath marks and note values literally. After familiarizing themselves with the melody and lyrics of the first couplet of the experiment, the participant was asked to sing the couplet four times in a row confidently and with the intent of narrating a story to someone who has never heard the story before. Each of these reiterations of a couplet shall be referred to as “repetitions” for further reference. If during the recording of the couplet, the participant made an error in singing or failed to produce a prosodic boundary at the correct time, he or she was asked to repeat the couplet another time to obtain four repetitions of each couplet. This process was then repeated for the other three couplets. The participants sang along with the MIDI recordings in the appropriate key through a set of Klipsch Reference One headphones in order to maintain a steady tempo and key. Palmer & Kelly (1992) found that singers singing with and without accompaniment produced similar prosodic and rhythmic effects in each case. Participants were recorded on the experimenter’s computer using Praat (Boersma & Weenink, 2013) and the computer’s internal microphone. The recordings were made at a 44,100 Hz sampling rate in stereo sound.

Participants

8 undergraduate participants, 7 male, 1 female, between the ages of 20 and 23 were solicited via email from the large population of highly skilled singers at Yale College to take part in the experiment for no monetary benefit. All of the participants came from audition-only choral singing groups which teach traditional choral technique and possessed sufficient musical skill to complete the task. These participants were naïve to the purpose of the experiment and were not shown the music beforehand.

Data Labeling

After these recordings were made, each recording was labeled in Praat. For the purpose of the SLA, the lengths of all segments adjacent to word boundaries between the target words in Couplet 1 were labeled. In Couplets 2-4, only the lengths of segments adjacent to prosodic boundaries were labeled. This gave pairs of pre-boundary and non-pre-boundary fricatives for the words “Tess,” “eats,” and “Anne’s” to be analyzed (refer to Table 1). For the BLA, the lengths of all words and all beats were labeled in all couplets (refer to Table 2). For the vowel-initial words, the glottal stops at their beginning were considered to be initial consonants, which the word length labelings (initial segment onset to initial segment onset) and the beat length labelings (vowel onset to vowel onset) took into account. This labeling is shown in Figure 5.

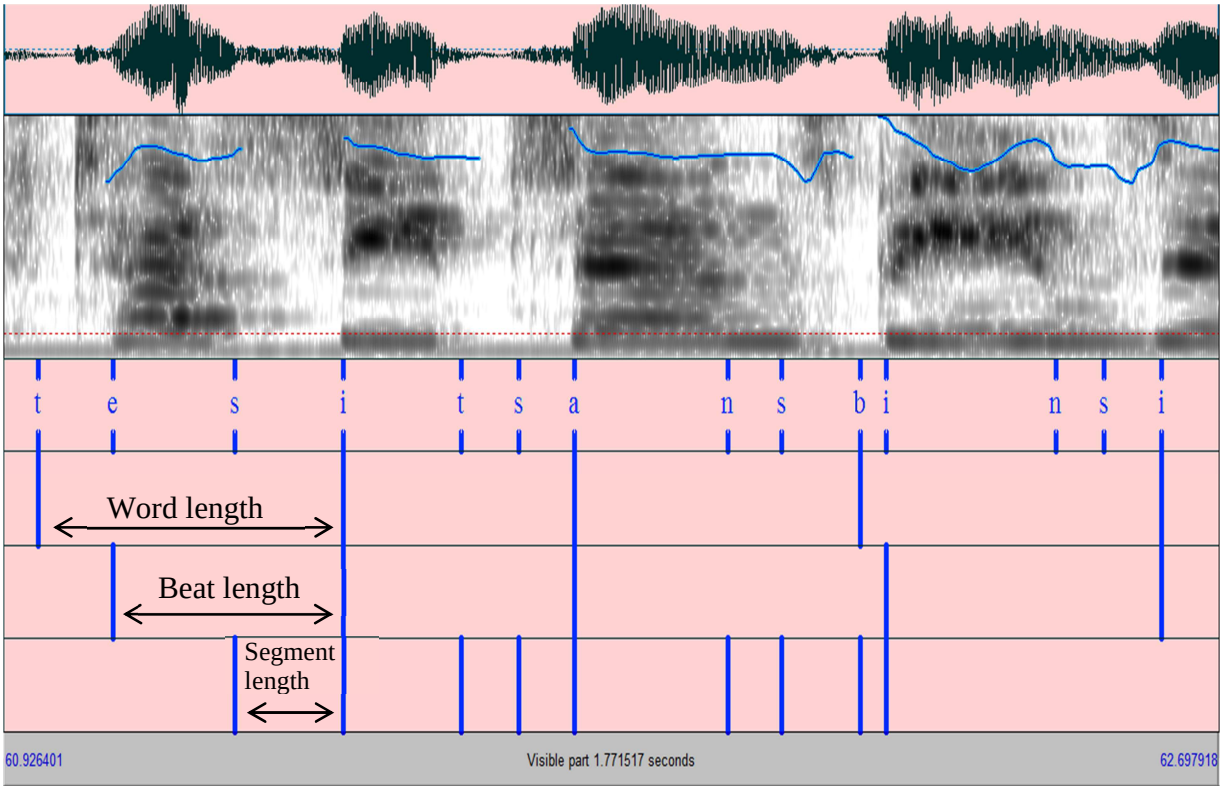


Figure 5: A sample labeling of Couplet 1 showing the waveform of the test phrase, the sptrogram, a tier for word length, a tier for beat length, and a tier for segmental length.

One of the subject's data was not labeled and discarded because in about half of his recordings, the nasalization of the preceding vowel was too strong to allow for reliable labeling. Thus only 7 subjects' data was labeled.

Statistical Analysis

The data were analyzed in R (R Development Core Team, 2013). The data was pooled between subjects because, while the subjects' data all tended in the same direction, their data individually did not show significant patterning. Dependent variable length values were converted to z-scores for each subject separately. Repeated-measures ANOVAs were performed on the pooled data. For the SLA, a one-factor ANOVA was performed on the segmental interval data to test for the effect of whether the segment preceded a prosodic boundary (prosodic

boundary). Then a two-factor ANOVA was performed on the segmental data to test for an interaction between the word in which the segment was the coda (test word) and whether the segment preceded a prosodic boundary. For the BLA, one-factor ANOVAs were performed separately on both the word interval data and the beat interval data to test for the effect of whether the interval preceded a prosodic boundary (prosodic boundary). Then two-factor ANOVAs were performed on both the word interval data and the beat interval data to test for an interaction between what syllable was being examined (test syllable) and whether the syllable preceded a prosodic boundary. Significance was set at $p<0.05$.

Results

Segment Length Analysis

The results of the pooled data analysis show effects in segmental data. In this data, 144 data points were analyzed. The one-factor ANOVA found an effect of prosodic boundary. This may be seen in Table 3 below. The direction of the prosodic boundary analysis showed larger z-scores of segment lengths before prosodic boundaries than phrase-medially. The two-factor ANOVA did not find an interaction between whether a prosodic boundary followed a coda and in what word the coda appeared. Data for the two-factor ANOVA all tended towards coda lengthening before prosodic boundaries by each word, if not significantly. Graphs of both of the ANOVAs may be seen in Figures 6 and 7 below.

Table 3: ANOVAs for the SLA data

Factor	**Segmental length data**
Prosodic Boundary (PB)?	F (1, 137)=9.209 **p=0.00288**

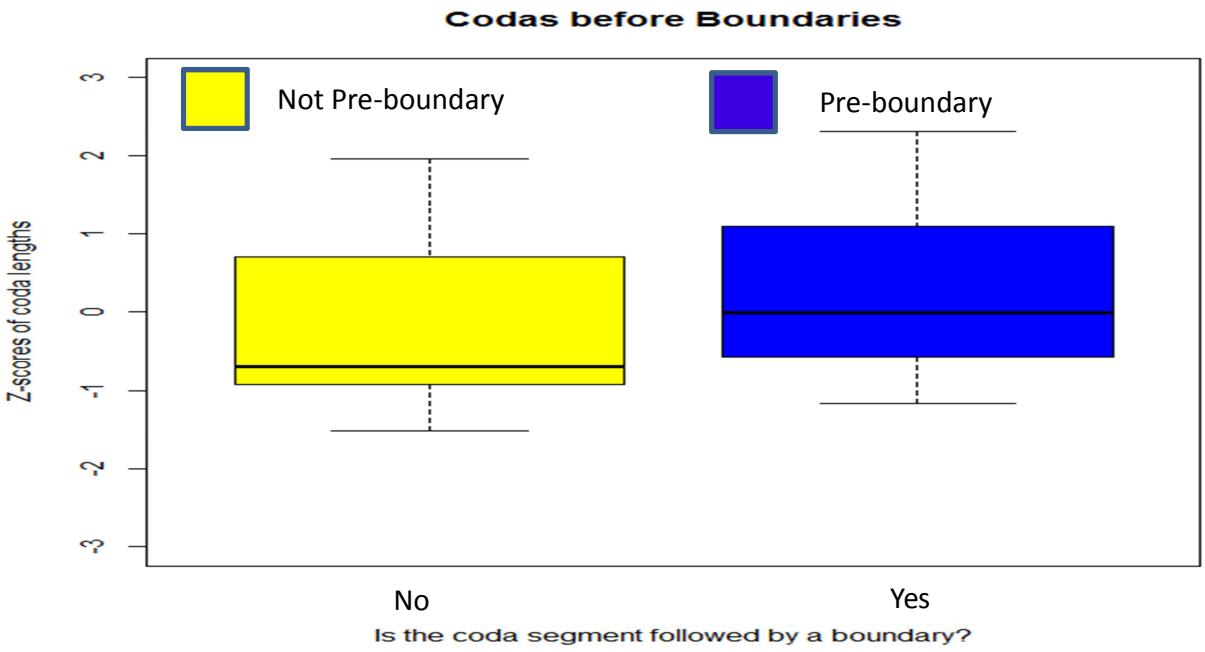


Figure 6: A graph of the z-scored lengths of segments preceding prosodic boundaries versus those phrase-medially.

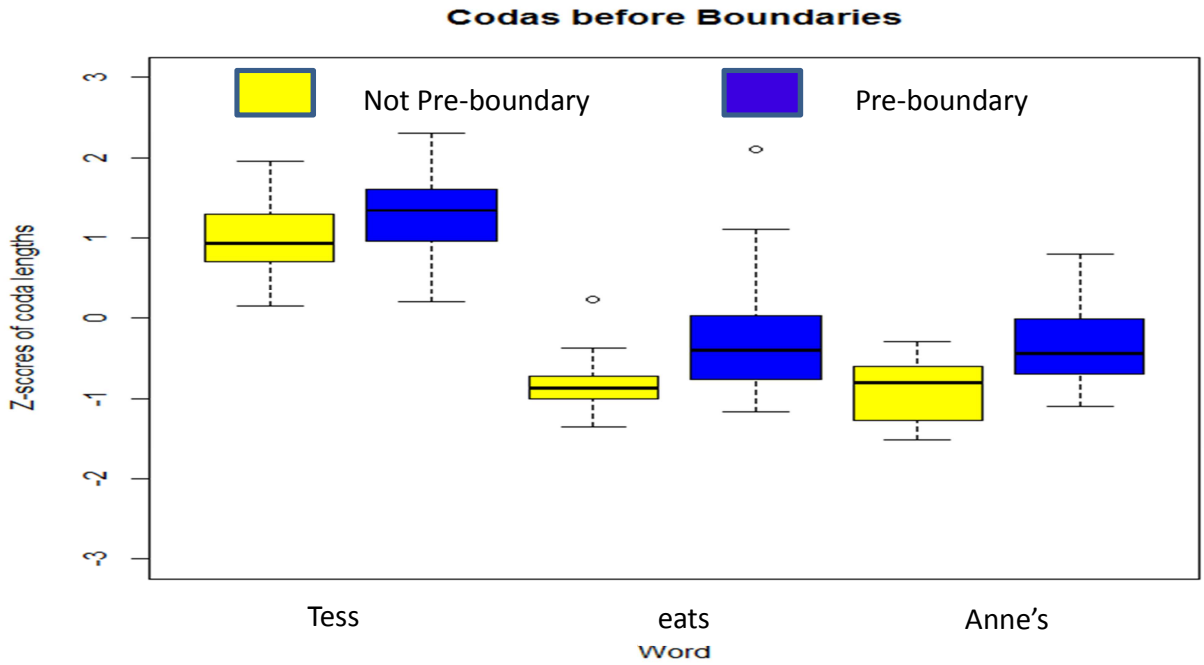


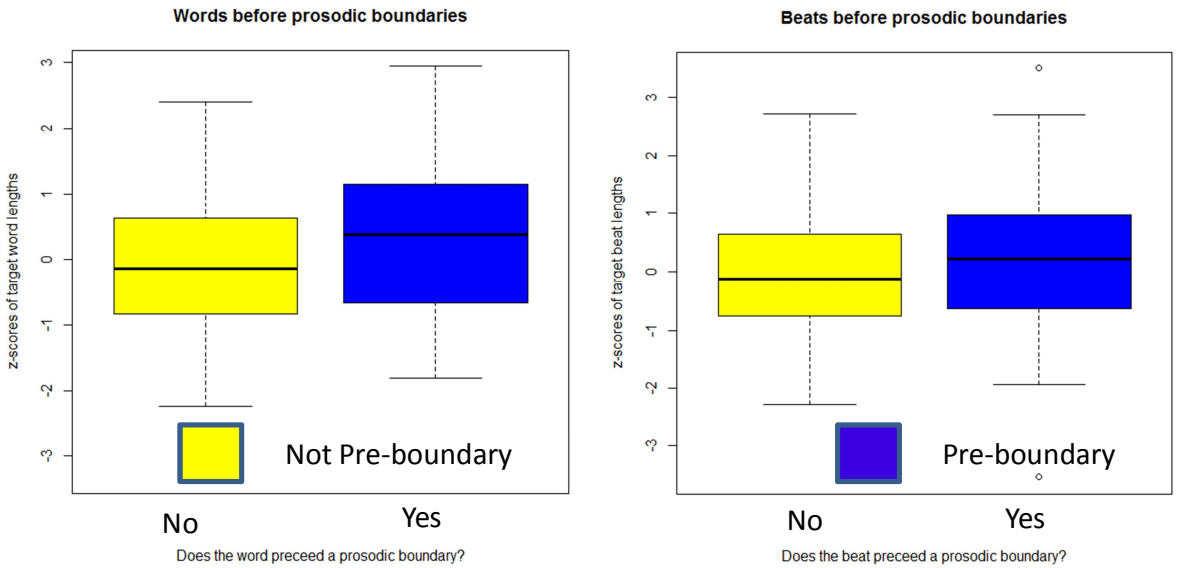
Figure 7: A graph of the z-scored lengths of segments preceding prosodic boundaries versus those not preceding prosodic boundaries broken down by word. These factors were not found to interact, but do show tendencies towards longer codas in the case of each word.

Beat Length Analysis

The results of the pooled data analysis show a number of different effects in both the word length and beat length data. In this data, 448 data points were analyzed. Globally, word length averaged 0.401 seconds per monosyllable in the target phrase and the beat length averaged 0.375 seconds per monosyllable in the target phrase. The one-factor ANOVAs found effects of target word and prosodic boundary for both the word length data and the beat length data. The two-factor ANOVA for word length found an interaction between whether a prosodic boundary followed the word and in what word was being examined. The two-factor ANOVA for beat length did not find an interaction between whether a prosodic boundary followed the beat and in what beat was being examined although the findings approached significance and the data tends in the right direction as seen in Figure 11. Table 4 and Figures 8-11 show these findings below. Lower z-scores represent shorter interval lengths and higher z-scores represent longer interval lengths. In almost every case, both words and beats are longer when preceding prosodic boundaries than when not. Intervals preceding prosodic boundaries are significantly longer than those that do not precede a prosodic boundary.

Table 4: ANOVAs for the BLA data

Factor	**Word length data**		**Beat length data**	
PB?	F (1, 440)=10.23	**p=0.00148**	F (1, 440)= 7.225	**p=0.00746**
PB?	F (1, 434)=42.08	**$p=2.39\times10^{-10}$**	F (1, 434)=19.127	**$p=1.53\times10^{-5}$**
Test Syl?	F (3, 434)=159.99	**$p<2\times10^{-16}$**	F (3, 434)=114.271	**$p<2\times10^{-16}$**
PB x Syl?	F (3, 434)=5.15	**p=0.00165**	F (3, 434)=2.553	p=0.055



Figures 8 & 8: Graphs of the z-scored lengths of words and beats preceding prosodic boundaries versus those not preceding prosodic boundaries.

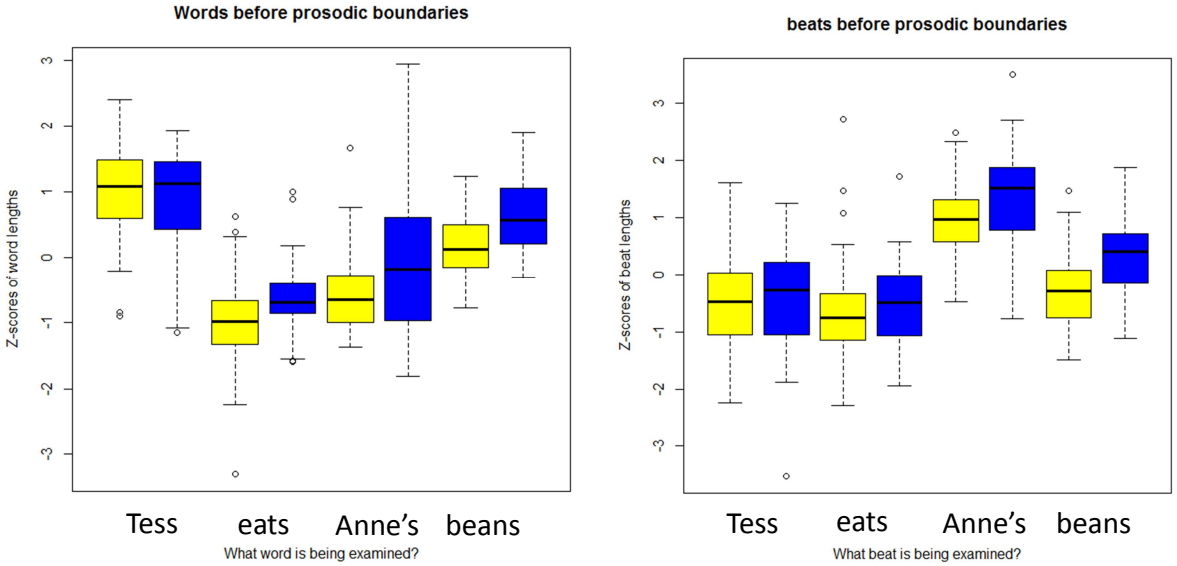


Figure 9 & 11: A graph of the z-scored lengths of differing words and beats.

Discussion

Segment Length Analysis Findings

The finding that a coda of a given word retains a consistent length when not interacting with a prosodic boundary (H1) is suggested by Figure 7, though not confirmed by a two-factor ANOVA. This suggests that the subjects across couplets and repetitions may produce consistent coda lengths for a given word and thus suggest the influence of the musical meter (i.e. the tendency towards isochrony of the tempo caused consistent patterning in the lengths of the fricative segments at the ends of the words "Tess," "eats," and "Anne's"). From this data, however, it is impossible to prove conclusively whether the codas are of consistent length by word, perhaps because of the small sample size.

As predicted in H2, however, the presence of prosodic boundaries did indeed lengthen the segments preceding the boundaries significantly, meaning that the need to keep the musical tempo "steady" and to mark musical stresses in the musical meter did not completely override the temporal requirements for marking the prosodic boundaries. The fact that these codas do lengthen before prosodic boundaries means that the tendency of musical meter towards isochrony was not strong enough to prevent the marking of these prosodic boundaries in the temporal domain. To examine the potentially destabilizing effects of the musical meter that these boundaries may have, we must first examine the effects of prosodic boundaries on the words and beats.

Beat Length Analysis Findings

The fact that the lengths of each word map across the study together significantly, as shown by the two-factor ANOVAs and Figures 10 and 11, demonstrates the effect of the musical

meter and supports H3. Where in normal speech the same word might be expected to vary in length when produced by different speakers and even by the same speaker, here the attending rhythms of the singer cause the participant to keep pace with the tempo of the computer recording and produce the words at a length calculated to make them arrive appropriately on the beat. While the particular necessities of articulation require certain words to be longer than other words, each word is being made relatively consistently. For example, in the case target words labeled with word intervals, the words "Tess" and "beans" will naturally be longer than "Anne's" and "eats" due to presence of onset consonants in the first pair and not the second. Similarly, we should expect that the target word "Anne's" will be the longest in beat interval data because that length also contains the [b] formerly possessed by "beans." In the beat intervals, it should also be noted that the [t] of "Tess" is not contained in the test word "Tess" or in any other test word interval. See the Table 5 below for more specific parsing of the word and beat intervals.

Table 5

Beat	3	4	1	2	3	4
Test Word		Tess	eats	Anne's	Beans	
Word Length		tɛs.	its.	ænz.	binz.	
Beat Length	**t**	ɛs.	its.	**ænz.b**	inz.	

This is not to say that the musical meter was not expected to have an effect of the production of the song; just that this level of consistency should be present before claims about changes in the length of words due to the presence of prosodic boundary can be made.

The finding by others (Allen, 1972) that the global beat established by the accompaniment is more closely tied to the inter-vocalic intervals seems to be consistent with the findings of this experiment. The tempo marking, to which the participants were entraining, was set to 160 beats per minute, which is equal to one beat per 0.375 seconds. If the beats in the musical meter tends towards the isochrony of the metronome marking, then the articulations most closely associated with the beat should be expected globally to average closely to the isochronous tempo marking. When all the inter-vocalic intervals are averaged, we find that 0.37518 seconds per syllable in the test phrase; whereas, the inter-onset intervals average to be 0.40126 seconds—longer than the intended beats per minute. If we assume that the claim that the beats are timed with the vowels of the target words, then this extra time in the inter-onset average almost certainly stems from the [t] of "Tess," which, according to the vowel-timed model, belongs to the previous beat.

Because of the finding that words and beats appear to be of consistent length phrase-medially in a strict tempo (H3) and the finding that words and beats lengthen before prosodic boundaries (H2), H4 seems to be confirmed. It does not appear that, with the lengthening of codas before prosodic boundaries, post-boundary words compensatorily shorten to preserve beat or word length normally prescribed by the meter. As a result of the lengthening prior to prosodic boundaries, there must be some amount of compensation to deal with the lengthening of the pre-boundary target words as the singer attempts to entrain back to the accompaniment. Unfortunately, this study was not designed to test for the domain of post-boundary compensatory shortening. However, this could be examined easily in a later study given this study's finding that prosodic boundaries may have an effect after any beat.

It is also interesting to note that the singing of Couplet 3 proved more difficult for subjects than the other conditions. This appears to be the case because in that couplet, the prosodic boundary follows the word "Anne's," which, if examined by beat length, is already the longest beat despite being on a weak musical beat (2^{nd}) due to the need to articulate the complex coda cluster of the word "Anne's" and the [b] of the following word which would be included in the inter-vocalic length of the 2^{nd} beat. Also, in Couplet 3, this word receives a linguistic focus stress, thus lengthening it more than in the other Couplets even before pre-boundary lengthen is applied. Thus when a prosodic boundary is placed after the word "Anne's," the 2^{nd} beat becomes even longer on an un-stressed musical beat, causing destabilization of the overall musical meter and causing 3 subjects to have to retry that Couplet to get it right.

Further Studies

As mentioned earlier, further studies will certainly need to be undertaken in order to ascertain the effects, if any, of compensatory shortening after prosodic boundaries to regain the musical meter. This would be particularly interesting to undertake to see the effects of compensatory shortening on musically stressed beats to determine whether post-boundary compensatory shortening may overcome the musical metric hierarchy where mismatches of prosodic stress failed before. This would be an interesting finding because it would imply that the need to entrain to an isochronous accompaniment speed outweighs the need to represent the stress of the attending rhythm oscillators themselves.

An articulatory magnetometer study examining the temporal properties of the boundaries in articulation would be very interesting to examine the domain of the pre-boundary lengthening

in singing and to see how the syllables may be aligned to the phase of the attending rhythm and the isochronous metronome of the accompaniment. In choral pedagogy, singers are taught to spend much more time on the vowels of the words being sung and less on the consonants. Thus, seeing whether this affects the domain of the pre-boundary lengthening or even whether there are short pauses at the boundary that are harder to hear in recordings would be very informative. However, even more interesting from the perspective of music cognition would be examining the relationship of the phase couplings of the syllables being sung and the intended tempo to see how they might be task dynamically linked.

Summary

This study examined the production and appearance of prosodic boundaries in song under a strict musical tempo. This study proposed the hypothesis that prosodic boundaries regardless of what beat they followed would overcome the musical meter's tendency to entrain to an isochronous rhythm when the prosodic boundaries were mismatched with musical boundaries. The study found significant evidence to support this claim in lengthening found on pre-boundary segments and syllables. Because of consistency also found in the length of phrase-medial syllables and their lengthening before boundaries, these pre-boundary syllables appear to counteract musical meter's stability and its tendency towards isochrony. In one case this even appears to have caused destabilization in several trials.

Acknowledgements

I would like to thank everyone who has helped me or guided me on this whirlwind of a project, in particular my insightful advisors Jelena Krivokapić & Ian Quinn. Without the Yale

Linguistics Department and the classes taught by Raffaella Zanuttini & Maria Piñango, this study would have never taken place, and the analysis would have been insurmountable without the suggestions and help of the other senior linguistics majors. Particularly on the statistical side, Zachery Maher and the Yale StatLab have proved invaluable. On the proofreading side, my parents and house mates Ayanna Woods, Kyle Picha, and Peter Thompson have been boons. The experiemtn also would not have be possible without the preponderance of excellent singers at Yale, particularly those in the Society of Orpheus & Bacchus. Without all these peoples help, this paper would not exist, and I am deeply indebted to them.

Sources

Allen, G. D. (1972). The location of rhythmic stress beats in English: An experimental study I. *Language and Speech*, *15*(1), 72-100.

Beckman, M. E., & Elam, G. A. (1997). Guidelines for ToBI labelling. *The OSU Research Foundation*, *3*.

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology yearbook*, *3*(1), 5-70.

Boersma, Paul & Weenink, David. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.1.25, retrieved January 28, 2010 from http://www.praat.org/

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*, 219-252.

Drake, C., Dowling, W. J., & Palmer, C. (1991). Accent structures in the reproduction of simple tunes by children and adult pianists. *Music Perception*, *8*, 315-334.

Falk, S. (2009). *Musik und Sprachprosodie: kindgerichtetes Singen im frühen Spracherwerb* (Vol. 9). De Gruyter.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*(2), 210-233.

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive psychology*, *15*(4), 411-458.

Grosjean, F. & M. Collins. (1979). Breathing, pausing and reading. *Phonetica*, *36*, 98-114.

Halle, J. (2003, March). Constituency matching in metrical texts. In *Submitted for publication in the proceedings of the conference Words and Music, University of Missouri-Columbia*.

Kim, H., & Cole, J. (2005). The stress foot as a unit of planned timing: Evidence from shortening in the prosodic phrase. In *Proceedings of Interspeech 2005*.

Kim, H., & Cole, J. (2006, January). Rhythmic Shortening in American English. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society* (Vol. 42, No. 1, pp. 133-147). Chicago Linguistic Society.

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of phonetics*, *35*(2), 162-179.

Ladd, D. R., & Campbell, N. (1991). Theories of prosodic structure: evidence from syllable duration. In *Proceedings of the 12th International Congress of Phonetic Sciences* (Vol. 2, pp. 290-293).

Ladd, D. R., & Cutler, A. (1983). Models and measurements in the study of prosody. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. l-10). Berlin: Springer-Verlag.

Ladefoged, P. (2001). Vowels and consonants. *Phonetica*, *58*, 211-212.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological review*, *106*(1).

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music.*Cognitive Science*, *26*(1), 1-37.

Large, E. W., Palmer, C., & Pollack, J. B. (1995). Reduced memory representations for music. *Cognitive Science*, *19*(1), 53-93.

Lehiste, I. (1972). Timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51.

Lehiste, I. (1973). Phonetic disambigation of syntactic ambiguity. *Glossa*, *7*, 107–122.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of phonetics*.

Lerdahl, F. (2001). The sounds of poetry viewed as music. *Annals of the New York Academy of Sciences*, *930*(1), 337-354.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*, 249-336.

Mithen, S. J. (2005). *The singing Neanderthals: The origins of music, language, mind, and body*. Harvard University Press.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of experimental psychology: human perception and performance*, *15*(2), 331.

Palmer, C., & Hutchins, S. (2006). What is musical prosody?. *Psychology of learning and motivation*, *46*, 245-278.

Palmer, C., Jungers, M. K., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. *Journal of Memory and Language*, *45*(4), 526-545.

Palmer, C., & Kelly, M. H. (1992). Linguistic prosody and musical meter in song. *Journal of Memory and Language*, *31*(4), 525-542.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), B35-B45

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, *90*, 2956.

R Development Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Selkirk, E. O. (1980). *On prosodic structure and its relation to syntactic structure*. Indiana University Linguistics Club.

Selkirk, E. O. (1984). *Phonology and syntax*. Cambridge, MA: MIT Press.

Sloboda, J. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology, 35*, 377-396.

Sundberg, J. (1974). Articulatory interpretation of the "singing formant". *The Journal of the Acoustical Society of America, 55*, 838.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America, 101,* 514.

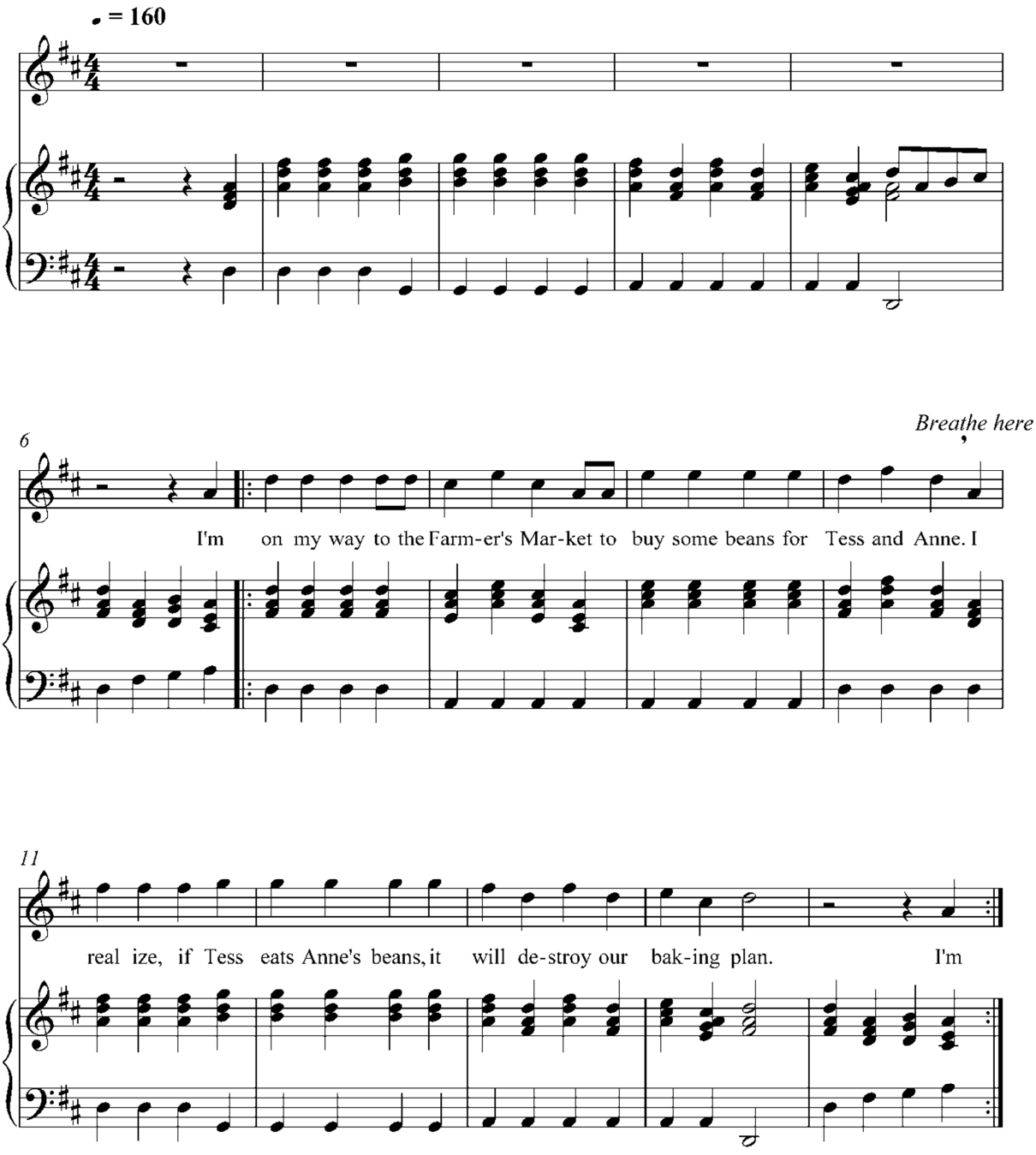
Tan, S. L., Pfordresher, P., & Harré, R. (2010). *Psychology of music: From sound to significance*. Psychology Press.

Trouvain, J., & Grice, M. (1999, August). The effect of tempo on prosodic structure. In *Proc. 14th Intern. Confer. Phonetic Sciences* (Vol. 1067, p. 1070).

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. *Methods in empirical prosody research, 3*, 1-28.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, *92*, 1707–1717.

Williams, J. R. (1998). Guidelines for the use of multimedia in instruction, Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, 1447–1451.


♩. = 160
6
Breathe here
I'm on my way to the Farm-er's Mar-ket to buy some beans for Tess and Anne. I
11
real ize, if Tess eats Anne's beans, it will de-stroy our bak-ing plan. I'm