

# English Vowel Duration in Textsetting

by

Nicole Gilroy, B.A.

A thesis to be submitted to the Faculty of Graduate and Postdoctoral  
Affairs in partial fulfillment of the requirements for the degree of

Master of Arts

in

Applied Linguistics and Discourse Studies

Department of Linguistics and Language Studies

Carleton University

Ottawa, Ontario

© 2021

Nicole Gilroy

## Abstract

The rhythmic structure in music, referred to as meter, consists of alternating strong and weak beats and higher-level structure. Likewise, in language there is a similar structure: stress is usually an alternation of strong and weak beats, though less regular than in music. In texts that have both linguistic and musical structure, the two rhythms prefer to match but the alignment is not always perfect, though it is systematic: this alignment is regulated by the textsetting grammar. I have investigated one aspect of a textsetting grammar, namely how vowel length differences interact with this system. More specifically, the effects of inherent durations were explored using minimal pairs differing in vowel tenseness, and minimal pairs with word-final /t/ or /d/ were used to probe the effects of voicing-induced allophonic length. This study explores the matching preferences between these linguistic tokens and slots in music which are metrically strong (and also long in duration), and slots which are metrically weak (and also short in duration). These experiments test both metrical structure and duration in music simultaneously and do not distinguish between them. The results show that English speakers prefer to match shorter vowels with short musical notes and longer vowels with long musical notes. The tenseness and allophonic lengthening conditions are both significant, and the participants were more strongly guided by the tenseness of the vowel than the voicing-induced allophonic length.

*Keywords:* Meter, Textsetting, Tense/lax Vowel Distinction, Allophonic Lengthening

## **Acknowledgements**

I would like to acknowledge and extend my gratitude to everyone who has been involved in this thesis. I am incredibly grateful for your contributions, involvement and support. I would like to specifically acknowledge and thank Ida Toivonen for her ongoing assistance throughout my independent research. Without her guidance I would not be in the academic position I am today, and I am forever grateful for her encouragement and support. I would also like to acknowledge and thank Lev Blumenfeld for his guidance throughout my graduate studies. I am grateful for his expertise which has been integral to my learning.

I would like to thank my committee members for taking the time to read and review my work. I am grateful to have people in my community that are willing to share their knowledge and experience with me.

Chase, thank you for being my support system and helping me through this entire process. Thank you for not giving up on me when I gave up on SPSS. I am sure it was not easy at times, and I appreciate all that you have done to help me get to where I am today.

## Table of Contents

<b>ABSTRACT</b> .....	<b>II</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>III</b>
<b>LIST OF TABLES</b> .....	<b>VI</b>
<b>LIST OF FIGURES</b> .....	<b>VII</b>
<b>LIST OF APPENDICES</b> .....	<b>VIII</b>
<b>CHAPTER 1: INTRODUCTION</b> .....	<b>9</b>
1.1 INTRODUCTION.....	9
<b>CHAPTER 2: LITERATURE REVIEW</b> .....	<b>12</b>
2.1 ENGLISH VOWELS .....	12
2.1.1 <i>The tense/lax distinction</i> .....	13
2.2 VOICED AND VOICELESS CONSONANTS .....	16
2.3 ALLOPHONY.....	17
2.4 MINIMAL PAIRS.....	18
2.5 METER .....	19
2.5.1 <i>Iambic Pentameter</i> .....	22
2.5.2 <i>Trochaic Tetrameter</i> .....	23
2.6 TEXTSETTING.....	24
2.6.1 <i>Lengthening</i> .....	26
2.7 LANGUAGES WITH CONTRASTIVE VOWEL LENGTH .....	28
2.7.1 <i>Languages with contrastive vowel lengths and their relationship with textsetting</i> .....	29

<b>CHAPTER 3: EXPERIMENT 1.....</b>	<b>31</b>
3.1 METHODOLOGY .....	31
3.1.1 <i>Participants</i> .....	31
3.1.2 <i>Experimental Design</i> .....	31
3.1.3 <i>Procedure</i> .....	35
3.2 RESULTS .....	35
3.3 DISCUSSION .....	40
<b>CHAPTER 4: EXPERIMENT 2.....</b>	<b>43</b>
4.1 METHODOLOGY .....	43
4.1.1 <i>Participants</i> .....	43
4.1.2 <i>Experimental Design</i> .....	43
4.1.3 <i>Procedure</i> .....	44
4.2 RESULTS .....	45
4.3 DISCUSSION .....	50
<b>CHAPTER 5: GENERAL DISCUSSION .....</b>	<b>52</b>
<b>CHAPTER 6: CONCLUSION.....</b>	<b>58</b>
<b>REFERENCES.....</b>	<b>60</b>
<b>APPENDICES.....</b>	<b>65</b>
APPENDIX A: FULL LIST OF MINIMAL PAIRS USED.....	65
APPENDIX B: EXPERIMENT 1 FULL LIST OF VOWEL DURATIONS .....	67
APPENDIX C: QUALITY ASSURANCE .....	71

## List of Tables

TABLE 1. ENGLISH VOWEL CHART .....	12
TABLE 2. ENGLISH TENSE AND LAX VOWELS .....	14
TABLE 3. MINIMAL PAIRS USING TENSE AND LAX VOWELS.....	18
TABLE 4. MINIMAL PAIRS USING WORD FINAL ALVEOLAR STOPS .....	19
TABLE 5. MUSIC NOTATION.....	32
TABLE 6. MINIMAL PAIR TYPES.....	33

## List of Figures

FIGURE 1. METRICAL HIERARCHY .....	20
FIGURE 2. IAMBIC PENTAMETER .....	23
FIGURE 3. TROCHAIC TETRAMETER .....	24
FIGURE 4. HIERARCHICAL CONSTITUENT STRUCTURE .....	25
FIGURE 5. LENGTHENING IN TEXTSETTING .....	27
FIGURE 6. TWINKLE TWINKLE LITTLE STAR .....	31
FIGURE 7. EXAMPLE OF MATCH STIMULUS .....	34
FIGURE 8. EXAMPLE OF MISMATCH STIMULUS .....	34
FIGURE 9. COUNT DATA OF VARIABLE COMPARISON .....	36
FIGURE 10. DURATION RATIO .....	38
FIGURE 11. QUARTER NOTE .....	39
FIGURE 12. HALF NOTE .....	40
FIGURE 13. STIMULUS FOR EXPERIMENT 2: BLANKS IN THE MUSICAL PHRASE .....	44
FIGURE 14. MATCH MISMATCH DATA .....	45
FIGURE 15. MATCH/MISMATCH RATES FOR GROUP 1 (BIT/BID).....	46
FIGURE 16. MATCH/MISMATCH RATES FOR GROUP 2 (BEAT/BEAD) .....	47
FIGURE 17. MATCH/MISMATCH RATES FOR GROUP 3 (BIT/BEAT).....	48
FIGURE 18. MATCH/MISMATCH RATES FOR GROUP 4 (BID/BEAD).....	49
FIGURE 19. VARIABLE COMPARISON .....	50

## **List of Appendices**

APPENDIX A: FULL LIST OF MINIMAL PAIRS USED .....	65
APPENDIX B: EXPERIMENT 1 FULL LIST OF VOWEL DURATIONS .....	67
APPENDIX C: QUALITY ASSURANCE .....	71



## Chapter 1: Introduction

### 1.1 Introduction

Music and language are rule-governed systems that appear in every human society (Patel, 2008). Music produced synchronously with lyrics offers an opportunity to investigate the interaction of music and language. Philosophers like Plato, Diderot and Rousseau discussed the power of music and how musical composition resembles speech sounds. In addition, these philosophers introduced the first notion that music played an integral role in the emergence of human language (Thomas, 1995; McPherson, 2019), and thus questions of language-music interactions are fundamental to understanding both.

Parallels between music and language include complex and meaningful sound sequences and so comparisons between the two domains have continued to spark the interests of linguists, psycholinguists, cognitive scientists and neuroscientists over many decades (Dell & Halle, 2005; Halle & Keyser, 1966; Halle & Lerdahl, 1993; McPherson, 2019; Rodríguez-Vázquez, 2010). In particular, these topics have been studied in the generative tradition, an approach I follow here (Blumenfeld, 2016; McPherson, 2019). There are two structures of similar nature within music and language. There is a rhythmic structure in music, referred to as meter, which consists of alternating strong and weak beats and higher-level structure. Likewise, in language there is a similar structure: stress is usually conceived as an alternation of strong and weak beats, though it is less regular than in music (Hayes, 1989). When speakers use their intuition to arrange syllables of novel lines in music into a pattern, they are engaged in the act of textsetting (Dell, 1989; Dell & Elmedlaoui, 2008; Dell & Halle, 2005; Rodríguez-Vázquez, 2010). In texts that have both linguistic and

musical structure, the two rhythms prefer to match but the alignment is not always perfect; this alignment is regulated by the textsetting grammar (Halle & Lerdahl, 1993).

In this thesis I investigate the specific, narrow issue of placement of monosyllabic words on strong or weak beats. In the experiment, these musical strong and weak beats contain long and short notes, respectively, thus on the musical side, duration and strength are conflated and not distinguished. The goal is to determine whether English speakers prefer to match shorter vowels with short/weak notes and longer vowels with long/strong notes in music, and secondly if length differences that are inherent properties of the vowels behave differently from length differences that are a result of a phonological process.

Chapter 2 provides a review of the literature on English vowels, English consonants, allophony, minimal pairs, meter, textsetting, and quantity languages. As a result, this chapter is divided into seven sections. The first section discusses English vowels and their articulatory properties. This section introduces inherent vowel duration properties and expands on the distinctions made between the lengths of English vowels, which is a primary focus for this thesis. The second section discusses English consonants and illustrates how voicing characteristics of consonants can affect properties of the surrounding English vowels. The third section discusses allophony, which builds on the preceding two sections to arrive at the phonological process used in the two experiments. The fourth section discusses minimal pairs which are integral to the understanding of the experiments presented in this thesis. Minimal pairs are used within the two experiments to create environments where allophonic lengthening occurs. The fifth section discusses meter and how linguistic properties interact with meter. In the sixth section I discuss textsetting and illustrate examples of textsetting from previous literature. In the seventh section I expand on textsetting and

how it is realized in languages that have duration as a phonetic property. The two experiments conducted use textsetting as a way to observe the interaction between music and language. These seven sections together provide a brief overview of the previous literature within this domain.

Chapter 3 describes the first experiment. The participants were asked to listen to pairs of short song lines and judge which line they preferred. Short vowel words were sung on quarter notes and long vowel words were sung on half notes in the matched case, and in the mismatched case short vowel words were sung on half notes and long vowel words were sung on quarter notes. Thus, each pair of stimuli included one song line with two sung words containing vowels that “matched” notes with respect to length, and the other song line included two words that did not match notes in length. Within this chapter there is a detailed description of the participants recruited, the experimental design, the procedure, how data analysis was conducted, the results, and a brief discussion.

Chapter 4 describes the second experiment, where participants were asked drag and drop one member of a minimal pair into a musical line. The musical line contained one blank on a quarter note and one blank on a half note. Within this chapter there is a detailed description of the participants recruited, the experimental design, the procedure, how data analysis was conducted, the results, and a brief discussion.

Chapter 5 is a general discussion of the results of the experiments in regard to previous literature. This chapter reviews the key findings and discusses the implications and limitations of the experiments.

Chapter 6 is the conclusion of the thesis. In this chapter I discuss the main ideas presented and elaborate on directions for future research.

## Chapter 2: Literature Review

### 2.1 English Vowels

The English phonemic inventory is comprised of vowels and consonants each with specific characteristics. A phoneme is the smallest unit that can distinguish meaning. For example, the words *coat*, and *goat* differ in only their initial sound /k/ vs. /g/, and because these two words differ in meaning, /k/ and /g/ are different phonemes (Plag et al., 2015). In this section I will introduce the phonetic characteristics of vowel phonemes.

The first distinction that can be made when discussing vowel placement in the oral cavity is a high/low dimension, corresponding to the height of the body of the tongue during the articulation of a vowel (Sundberg, 1969). Using this height distinction vowels can be plotted within the vowel space as high, mid or low, as seen in Table 1 below. The second distinction is a front/back dimension which refers to the extent to which the body of the tongue lies towards the front, middle or back of the vowel space (Sundberg, 1969). Using this backness distinction vowels can be plotted within the vowels space as front, mid or back, as seen in Table 1 below.

Table 1. English Vowel Chart

	Front	Central	Back
High	/i/, /ɪ/		/u/, /ʊ/
Mid	/e/, /ɛ/	/ə/, /ʌ/	/o/
Low	/æ/		/ɑ/

English Vowels have a rounding characteristic based on whether the lips are rounded during the production of the vowel. In Canadian English, there are only four vowels that have lip rounding, and they are all made with the tongue at the back of the mouth: /u/ as in *boot*, /ʊ/ as in *book*, /o/ as in *boat*, and /ɔ/ as in *bore* (Sundberg, 1969).

Lastly, English vowels can be distinguished by their tenseness. The tenseness of a vowel is integral to the studies conducted in this thesis and is discussed in the following section.

### ***2.1.1 The tense/lax distinction***

In addition to height, backness, and rounding, English vowels can be classified as tense or lax, in which tense vowels are produced with more extreme articulatory movements than lax vowels. Tenseness is discussed in relation to the tension of the tongue during the production of the vowel. When articulating a lax vowel, the muscles of the vocal apparatus are relatively loose. In contrast, when articulating tense vowels, the tongue and oral cavity are relatively tense (Leung et al., 2016; Plag et al., 2015; Sundberg, 1969). In addition, the vowel space covered by tense vowels is typically larger than that covered by lax vowels (Sundberg & Skoog, 1999). Table 2 provides a list of the English tense and lax vowels with examples.

Table 2. English Tense and Lax Vowels

Tense Vowels	Example	English Gloss	Lax Vowels	Example	English Gloss
i	/lid/	lead	ɪ	/lɪd/	lid
eɪ	/het/	hate	ɛ	/bɛt/	bet
u	/pul/	pool	æ	/bæt/	bat
ow	/bot/	boat	ʊ	/pʊt/	put
ɑ	/brat/	brought	ʌ	/bʌt/	but
			ə	/əbʊt/	about

Unlike height, backness, and rounding, tenseness does not have such obvious phonetic correlates. It does, however, have a clear role in English phonology, in that the tense/lax distinction matters for stress assignment, as well as for phonotactics (Peterson & Lehiste, 1960). Lax vowels act as phonologically short and tense vowels as phonologically long. For example, lax vowels, unlike tense vowels, do not attract stress, and lax vowels are reduced to [ə] or [ɪ] in open unstressed syllables. In addition, lax unreduced vowels cannot occur word-finally, cf. *be* [bi] but [\*bɪ]. This means that the tense and lax vowel categories are realized as separate categories in English grammar (Leung et al., 2016; Peterson & Lehiste, 1960; Plag et al., 2015; Sundberg, 1969).

The vowels /ɪ/ - /i/ and /u/ - /ʊ/ were selected for the experiments conducted in this thesis to contrast vowel differences in English; this includes backness, rounding and tenseness (Peterson & Lehiste, 1960). Only high vowels were selected as there are no mid or low vowels in English with a tense/lax distinction; these vowels only occur as part of the diphthongs. Leung et al. (2016) discuss that the minimal pairs /ɪ/ - /i/ and /u/ - /ʊ/ are the

easiest minimal pairs to observe. High, front vowels /i/ and /ɪ/ only contrast is their tenseness. These vowels are very common in English and made it easy to come up with minimal pairs for the conducted experiments. The high, back, rounded vowels /u/ and /ʊ/ were the second pair of vowels selected. These vowels have the same characteristics as one another aside from one being tense and the other being lax (Klatt, 1973; Peterson & Lehiste, 1960), which makes them perfect candidates for analysis.

In this thesis tense and lax vowels are used for their accompanying length properties. In English, vowels do not have phonemic duration properties that can distinguish them from each other based only on this feature (Klatt, 1973). However, tense vowels are typically longer than their lax vowel counterparts. Leung et al. (2016) discuss how this effect is the result of longer excursions for articulators to reach the more extreme tense vowel target positions.

Previous research conducted by House (1961) has shown that the amount of durational change for tense and lax vowels differs under the influence of contextual factors such as consonantal context and speaking rate. He ran an experiment looking at average durations of 12 English vowels measured in bisyllabic nonsense utterances. These environments consisted of the voiced and voiceless versions of three stops, one affricate, and three fricative consonant articulations. The data shows that the average duration of a vowel increases with vocalic tenseness and with vocalic openness. Openness corresponds to a vowel sound in which the tongue is positioned as far as possible from the roof of the mouth; open vowels are also referred to as low vowels. Closeness corresponds to a vowel sound in which the tongue is positioned close to the roof of the mouth; closed vowels are also referred to as high vowels (Plag et al., 2015).

English lax vowels, also described as *short* vowels, are in fact characterized as having the shortest average durations of all the English vowels (Klatt, 1973; Leung et al., 2016). The average difference between the tense category and the lax category is 100ms when produced in isolation (House, 1961), and the general duration range for lax vowels is less than that of tense vowels. House (1961) suggests that as the rate of utterance increases, the change in the duration of an unstressed syllabic nucleus is greater than the change in the duration of the syllabic nucleus under stress. Furthermore, the average duration of a stressed vowel is approximately 50% longer than the average unstressed vowel.

The stimuli used for the experiments reported in this thesis consisted exclusively of words with a single, closed syllable (i.e., monosyllabic words ending with a consonant). This eliminated the confounds of word-internal stress and openness of syllables.

## **2.2 Voiced and Voiceless Consonants**

Voicing is a characteristic of consonants that corresponds to the presence of vocal fold vibration during articulation. In English, some consonants, called obstruents, display a voiced/voiceless contrast, while other consonants, called sonorants, and vowels do not (Plag et al., 2015).

For the two experiments reported in this thesis, voiced and voiceless alveolar plosives were used in the majority of minimal pairs, in word final position, to create an environment where vowel lengthening occurs (House & Fairbanks 1953; House, 1961; Klatt, 1973). The process of lengthening is discussed in the following section on allophony.



### 2.3 Allophony

In phonology, an allophone is one of a set of possible spoken sounds, or *phones*, which can realize a single phoneme. For example, in English, there is an allophonic variation of /t/ (as in *stop* /stop/) and the aspirated form [t<sup>h</sup>] (as in *top* [t<sup>h</sup>ɒp]). Both /t/ and aspirated [t<sup>h</sup>] are allophones because they are a realization of the same phoneme (House & Fairbanks, 1953; Klatt, 1973; Plag et al., 2015). Replacing one phoneme with an allophone of the same phoneme does not change the meaning of a word, but the result may sound non-native or unintelligible. In languages such as Thai and Hindi, the two phones [t] and [t<sup>h</sup>] are different phonemes (House & Fairbanks, 1953; Klatt, 1973; Plag et al., 2015). Likewise, in Spanish, [d] and [ð] are allophones of the phoneme /d/, while these two are different phonemes in English. The specific allophone selected in a given situation is often predictable from its phonetic context (Plag et al., 2015; Rodríguez-Vázquez, 2010).

In English there is a phonological process of allophonic variation, or more narrowly, allophonic lengthening (House & Fairbanks, 1953; Klatt, 1973; Lunden, 2017). This occurs when there is a voiced consonant in coda position immediately after the vowel. As a result, vowels that come before voiced consonants are longer than vowels that come before voiceless consonants (House & Fairbanks, 1953; Klatt, 1973). House (1961) reports that the average duration of lax vowels before a voiceless consonant (140 milliseconds) is approximately 40% of that of tense vowels before a voiced consonant when used in concise and slow speech (340 milliseconds). Vowels are at their maximum length when they are in open syllables or closed by a voiced consonant, but they are considerably shorter when they occur in a syllable closed by a voiceless consonant (Lunden, 2017).

This thesis investigates how the allophonic variation of vowels, specifically allophonic lengthening, interacts with textsetting. More specifically, the experiments analyze whether a vowel that does not undergo lengthening is preferred on a shorter note in music, and if a vowel that does undergo lengthening is preferred on a longer note. To explore these possible realizations minimal pairs of monosyllabic words have been selected and are discussed in the following section.

## 2.4 Minimal Pairs

When one phoneme is substituted for another, we get a different word. For example, if we substitute the phoneme /d/ in *mid* (/mid/) for /t/, we get the word *mitt* (/mit/). Word pairs like *mid* (/mid/) and *mitt* (/mit/) are examples of minimal pairs (Klatt, 1973; Ryan, 2014). Minimal pairs differ by one sound and also differ in meaning. Table 3 below provides examples of minimal pairs due to vowel tenseness. The only difference between the example words in each row is the tenseness of the vowel. In the example pair for /i/ and /ɪ/ we see the same /b/ phoneme in onset position and the same /d/ phoneme in coda position. The second example is a word pair containing the vowels /u/ and /ʊ/. Both of these examples have the phoneme /f/ in onset position and /l/ in coda position, resulting in identical the environments for both vowels.

Table 3. Minimal Pairs Using Tense and Lax Vowels.

Tense vowel	Example	English gloss	Lax vowel	Example	English Gloss
/i/	/bid/	bead	/ɪ/	/bɪd/	bid
/u/	/ful/	fool	/ʊ/	/fʊl/	full

Table 4 below provides examples of minimal pairs containing /d/ or /t/ in word final position.

Table 4. Minimal Pairs Using Word Final Alveolar Stops

Vowel	Word final /d/	English gloss	Word final /t/	English gloss
/i/	/mid/	mead	/mit/	meet
/ɪ/	/mɪd/	mid	/mɪt/	mitt
/u/	/bud/	booed	/but/	boot

In Table 4, vowels in words with word-final /t/ do not undergo lengthening as they have a voiceless phoneme in coda position. However, the vowels the column for word final /d/ does undergo allophonic lengthening due to the voiced consonant in coda position.

## 2.5 Meter

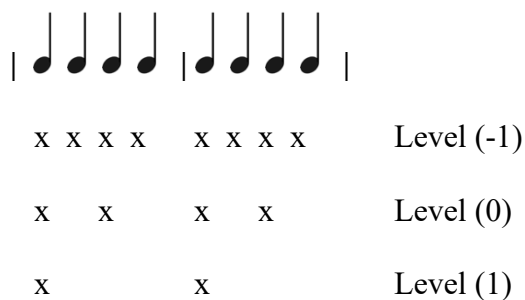
Meter, as defined by Hayes (1989), is the study of how any sort of rhythmic patterns are realized and represented in linguistics. The study of meter, in the generative tradition, began with foundational papers by Halle & Keyser (1966), and the basic idea has not changed since then: meter is a rule-governed system consisting of a template of strong and weak beats, and a system of rules/constraints that matches that template with a linguistic representation (Dell, 2009; Dell & Elmedlaoui, 2008; Dell & Halle 2005; Halle & Idsardi, 1995; Halle & Keyser, 1966; Hayes, 1989; Rodríguez-Vázquez, 2010). The rhythmic structure in music, referred to as metrical structure, also consists of a template of alternating strong and weak beats. Likewise, in language there is a similar structure: stress is usually conceived as an alternation of strong and weak beats, though less regular than in music

(Hayes & Kaun, 1996). In the ideal case, the metrical templates match the stress pattern; however, actual lines may show misalignments.

Dell & Halle (2005) discuss that in languages with stress, such as English, there is a preference called stress to beat matching, which associates stressed syllables with strong metrical positions. Stress to beat matching is strict in English, where certain stress/beat mismatches are prohibited.

Textsetting works analogously: the metrical template is supplied by the music and the text must also be aligned with it. Halle & Lerdahl (1993) analyze a sung-verse textsetting system using an algorithm that matches the two structures, using a grid representation. In the grid, units in rows depict a series of isochronous beats on a hierarchy of levels, and the columns indicate the strength of individual beats. This is used to determine which metrical positions best correspond to each syllable. Strong beats are those falling on level 0 or higher. Additionally, there is a beat addition rule that states that a level (-1) beat may be inserted between two level 0 syllables, in order to achieve a well-formed grid with a strong/weak alternation.

Figure 1. Metrical Hierarchy

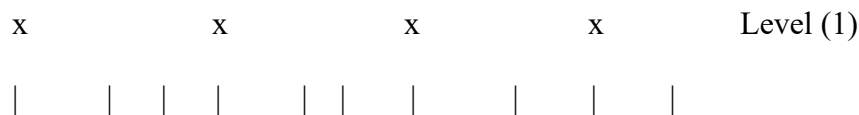


*Note:* Events that take place on multiple levels are assigned to the most prominent level of which they are situated.

The algorithm by Halle & Lerdahl (1993) provides the background on how the metrical template is realized in context. This algorithm refers to categorical properties, such as stress/no stress; however, the conducted experiments for this thesis use fine-grained properties of duration. This algorithm best describes a particular instance of an English textsetting grammar that Halle & Lerdahl investigated, shown below, and cannot be extrapolated universally.

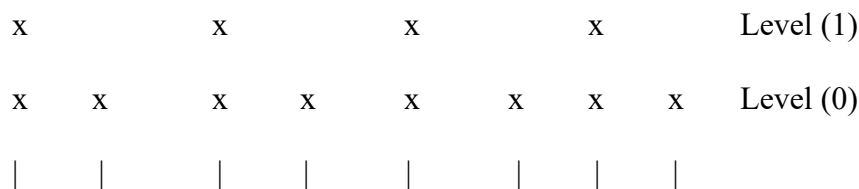
The algorithm is broken up into three (3) steps:

(1) You must first find the stressed syllables and assign them from left to right in unoccupied strong positions in the grid, to level 1.



What shall we do with a drunk- en sail- or

(2) The remaining syllables are placed onto the highest available grid level from left-to-right, on level 0.



What shall we do with a drunk- en sail- or



Hayes (1989) provides an example of this strong/weak alternating rhythm seen in Figure 2, using a line from *Ozymandias* by Percy Bysshe Shelly. Here we see that *line a* is an actual realization of iambic pentameter because it follows the rules/constraints of English iambic pentameter. However, *line b*, though it has the same number of syllables, does not follow iambic pentameter, though it still has strong/weak sequencing.

Figure 2. Iambic Pentameter

a. w s w s w s w s w s

My name is Ozymandias, king of kings

b. w s w s w s w s w s

Ozymandias by Percy Bysshe Shelly

(Hayes, 1989)

### 2.5.2 *Trochaic Tetrameter*

More specifically for this thesis, trochaic tetrameter exemplifies (mis)matching between two structures. Trochaic tetrameter uses the same metrical theory of iambic pentameter except the foot in the metrical template is left-headed (Blumenfeld, 2016; Halle & Keyser, 1971). In the conducted experiments for this thesis, a musical phrase from the children's song *Twinkle Twinkle Little Star* was used. *Twinkle Twinkle Little Star* uses trochaic tetrameter as it contains a line of four trochaic feet corresponding to the four strong beats in music. This is represented in Figure 3 below.

Figure 3. Trochaic Tetrameter

s w s w s w s  
twinkle twinkle little stár

(Halle and Keyser, 1971).

The grammar that I am investigating in the following experiments follows a trochaic pattern consisting of strong and weak beats, with the added component of music. As a result of the musical accompaniment, the experiments are asking participants to engage in the act of textsetting.

## 2.6 Textsetting

When speakers engage in the act of textsetting when they use their intuition to arrange syllables of novel lines, in music, into a pattern (Dell, 1989; Dell & Elmedlaoui, 2008; Dell & Halle, 2005; Rodríguez-Vázquez, 2010). Strictly speaking, it is textsetting as a whole is that is metrical or unmetrical, and that includes not just the words of the line, but also how those words fill the metrical positions of the template. Incidentally, some sequences of words may be such that there is no possible way to fit them into the template, for example, if they are too long or too short (Hayes & Kaun, 1996).

Textsetting alignments rely on two similarities between language and music. The first similarity is the hierarchical constituent structure: syllables and words in language, like the notes in a song, form groups and these groups form larger groups, resulting in a hierarchy, illustrated in Figure 4 (Dell, 1989; Dell & Elmedlaoui, 2008; Dell & Halle,





strong beats” (p. 2). However, the confounds of grouping are not relevant to the conducted experiments of this thesis.

These grouping practices are used, however, in an experiment by Hayes and Kaun (1996) investigating sung and chanted verse. Their research is based on a corpus of 670 English folksong lines, as well as chanted renditions of the corpus by ten native speaker consultants. They found that in sung and chanted verse there is a connection between the metrical pattern of the linguistic material that is either grouped at the phrasal level or at the right edge of a high-level constituency. An integral finding from this paper is that sung and chanted verse tends to match the number of beats allotted to a syllable, which fits with that syllable's natural linguistic duration, suggesting that syllables are phonetically lengthened at the end of a stanza (Hayes & Kaun, 1996). This is described in more detail below.

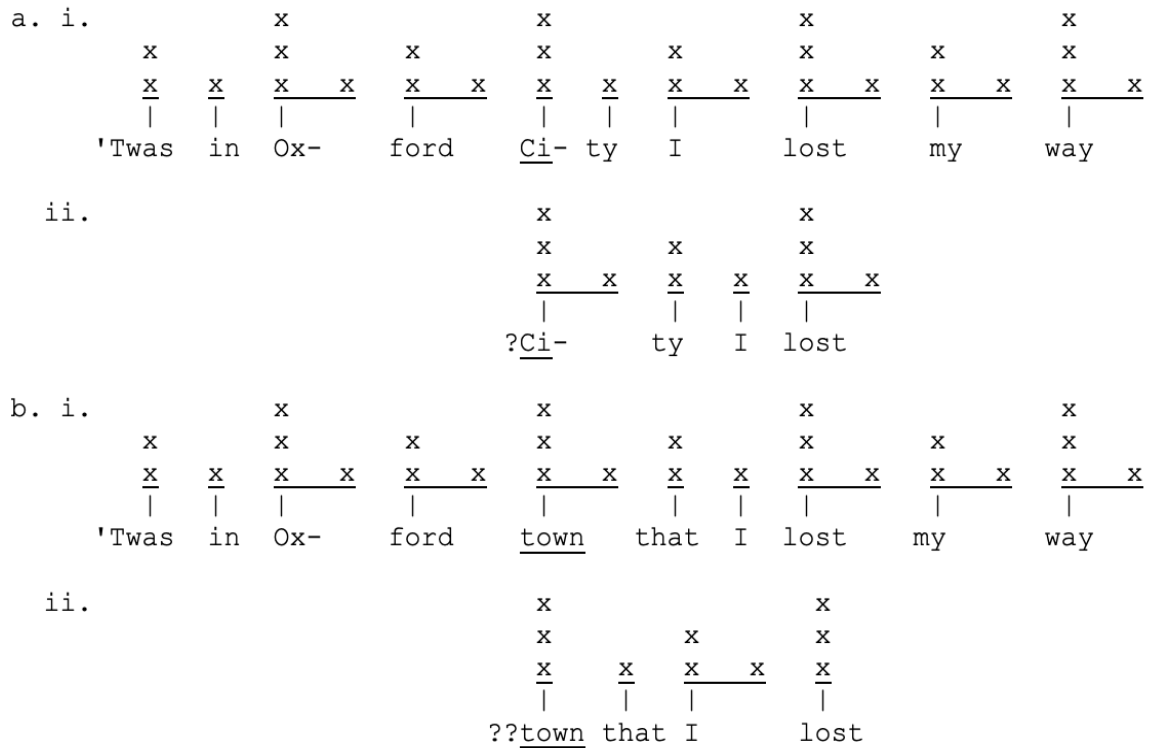
In this thesis I am looking at a specific, narrow, issue of placement of monosyllabic words on strong or weak beats that are determined by their phonetic duration.

### **2.6.1 *Lengthening***

Hayes and Kaun show that phonetic length plays a determining role in textsetting. English is described as a *stress-timed* language, and its syllables may vary dramatically in their duration. Hayes and Kaun (1996) discuss that it is possible to divide English syllables into the traditional categories of heavy vs. light, but this distinction is only one of the factors that determine the phonetic duration of English syllables. Additional factors are stress level (referring to the stress distinctions that determine vowel reduction), what the segment is, and most importantly the location of the syllable within prosodic structure. Further, the authors discuss how they expect that preferences would arise in how the natural duration

of a syllable is reflected in the textsetting. Figure 5 is an example taken from Hayes and Kaun (1996) where they contrast two syllables, one phonetically short and the other long, in the same metrical position:

Figure 5. Lengthening in Textsetting



(Hayes & Kaun, 1996)

The first syllable of city is phonetically short (lax vowel and open syllable) and seems to be set most naturally assigned to a single metrical position (3a.i). The syllable *town* is long (closed syllable) and is set most naturally to two positions (3b.i). The distinction becomes clearer when saying (3a.i) to the rhythm of (3b.i) and vice versa, as shown in Hayes & Kaun's (1996) examples labeled (ii).

The effect of phonetic duration is also demonstrated for syllable weight (Hayes & Kaun, 1996; Ryan, 2014). Syllables found at the right edges of the higher-level constituents are more likely to be textset across multiple beats. Moreover, heavy syllables like *town* are more likely to be textset on longer notes than light syllables. What makes this textsetting finding particularly interesting is that it shows that speakers are at least subconsciously aware of gradient durational differences, which can influence how language aligns in music.

As a result of these findings, we know that textsetting is not only sensitive to the phonological configurations of a word but also to phonetic details of speech in those configurations. This thesis delves into a specific aspect of this fine phonetic detail, specifically how vowel length differences, due to tenseness and lengthening, interact with this system.

## **2.7 Languages with contrastive vowel length**

Unlike English, in some languages, such as Arabic, Thai, Finnish, Lithuanian and Fijian, vowel length is contrastive and an important phonemic factor: in these languages, vowel length can change the meaning of the word (Tajima et al., 2006). Most languages that do distinguish vowel length phonemically usually only distinguish between short vowels and long vowels. Very few languages distinguish three phonemic vowel lengths, such as Estonian (Plag et al., 2015; Peterson & Lehiste, 1960).

Languages with contrastive vowel lengths are known for their use of different vowel durations which result in phonological distinctions. For example, Thai has nine short vowels with nine long counterparts. Long vowels in Thai have the same acoustic properties to their short counterparts with the only difference being that they are longer in duration. This acoustic difference is a primary perceptual cue for native listeners of Thai (Cooper et

al., 2017). Natural speech in Thai contains a high degree of variability with the rate of speech, which can have a significant effect on the duration. Cooper et al. (2017) states, “[...] an increase in speaking rate reduced the duration of Thai long vowels more than short vowels, whereas a slower speaking rate lengthened short vowels more than long vowels” (p. 66). The duration of a quickly produced long vowel may be shorter than a slow rate short vowel in plain speech. This suggests that speakers could be normalizing for duration as a function of speech rate. This is relevant to this thesis as it illustrates that speakers of these languages possess internal phonemic length categories. However, when languages with a phonemic length contrast engage in textsetting, there are specific ways in which they can distinguish vowels.

### ***2.7.1 Languages with contrastive vowel lengths and their relationship with textsetting***

The participants recruited for the conducted experiments in this thesis were not required to be native speakers of English. As such, these participants could be native speakers of languages with length contrasts. When language accompanies music in the form of lyrics, the duration of the vowels is determined not only by the spoken language, but by the length of the musical note (Dell, 1989; Dell & Elmedlaoui, 2008; Dell & Halle, 2005; Rodríguez-Vázquez, 2010). Speakers of languages with phonemic vowel length contrasts use various strategies to negotiate the tension between lexical duration and the demands of textsetting. De’Ath & Neubert (2019) discuss that listeners tend to perceive non-native length distinctions more continuously, rather than categorically like native listeners. German singers are not always able to fulfill the demands of German standard pronunciation (Hochdeutsch) in terms of vowel length contrasts. This contrast depends on distinguishing

the use of high and low vowels, however, it would be difficult for the listener to distinguish between high /i/ and low /e/, and narrowing the open vowels to create a distinction in most cases is not appropriate to standard pronunciation (De'Ath & Neubert, 2019). As a result, singers of languages with long and short vowel contrasts use word stress to indicate a long vowel even if the actual duration of the vowel is dictated by the musical note it is accompanying. There is a ratio between the length of a consonant or consonant cluster and the (stressed) vowel it precedes or follows, as set by the composer, that cannot be violated. The longer the vowel, the longer the consonant or consonant cluster must be (De'Ath & Neubert, 2019).

## Chapter 3: Experiment 1

### 3.1 Methodology

#### 3.1.1 Participants

For the purpose of this study, 170 participants were recruited. In this first experiment participants were asked to listen to audio files. Listening to the audio files would take 7 minutes, therefore a minimum time cut off was set to 15 minutes for the overall duration of the experiment. As a result, the data of 83 participants was immediately excluded from the results as their responses did not meet the minimal possible time to complete the survey successfully. The 87 remaining participant responses were selected to be included in the analysis. Participants were recruited using Carleton University's Cognitive Science recruitment system (SONA). As compensation for their participation, participants received 0.75 % of a credit towards a Cognitive Science course. The eligibility requirements stated that any student with an active SONA account could participate. Participants were not required to have any musical knowledge or training. Participants were not required to be native speakers of English; this allowed for a larger participant pool.

#### 3.1.2 Experimental Design



Participants were asked to listen to pairs of pre-recorded musical phrases recorded specifically for this experiment, set to the tune of *Twinkle Twinkle Little Star*, as shown in Figure 6 below.

Figure 6. Twinkle Twinkle Little Star



The selected musical phrase for the experiment contains notes of two types shown in Table 5.

Table 5. Music Notation

Musical note name	Notation length	Note
Quarter note	Half the duration of a half note (1 beat)	
Half note	The duration of 2 quarter notes (2 beats)	

The phrases were recorded as short songs sung with simple accompanying words designed for the experiment. The songs contained “la la” and two English words, one word inserted on the fourth note (a quarter note) and the other inserted on the fourteenth note (a half note), as seen in (1) and (2) below. On the screen, participants saw two play buttons, that allowed them to play the two songs as seen in (1) and (2). Which stanza corresponded to which play button was randomized for each participant. The participants could listen to the files as many times as they felt necessary, however, once the participant went to the next question they could not go back to re-listen to previous files. To indicate their response, participants selected the box beside the corresponding play button.

(1) la la la bid la la la la la la la la la la bead

(2) la la la bead la la la la la la la la la la bid

The English words inserted in the songs were monosyllabic minimal pairs that differed either in whether the vowel was tense or lax or whether the final consonant was voiced or voiceless. The members of the minimal pairs can be described as “long” and



“short” words, where the duration difference is due to one of two reasons: having a tense vs. lax vowel or having a voiced vs. voiceless consonant in the coda that results in allophonic lengthening of the vowel. Tense vowels are longer than lax vowels in English, and vowels are also longer before voiced than voiceless consonants. For the purpose of this experiment monosyllabic words with word-final /d/ or /t/ have been selected alongside the tense vowels /i/ and /u/ and the lax vowels /ɪ/ and /ʊ/. Ideally, there would have been only words with word-final /d/ and /t/ however, due to the lack of possible English words with /ʊ/, word-final /l/ and /k/ were also used<sup>1</sup>. See *Appendix A* for the full list of minimal pairs used.

As a result, the minimal pairs are of four types. Table 6 below provides a list of the possible word combinations participants were presented with along with an example pair.

Table 6. Minimal Pair Types

Type	Group name	Example
1	lax vowel with voiced/voiceless contrast	bit, bid
2	tense vowel with voiced/voiceless contrast	beat, bead
3	tense/lax contrast with voiceless consonant	beat, bit
4	tense/lax contrast with voiced consonant	bead, bid

For each of the 29 minimal pairs, two musical phrases were recorded, one where the duration of the words matched the duration of the note (long on long, short on short), and one where they mismatched (long on short, short on long). Examples of a match

<sup>1</sup> Creating pairs such as look /lʊk/, Luke /luk/ and pull /pʊl/, pool /pul/

stimulus and a mismatch stimulus are shown in Figures 7 and 8, respectively. Additionally, the long note happens to occupy a strong metrical position (two gridmarks), while short notes occupy a weak metrical position (one gridmark). Figure 7 is an example of match stimulus, where the lax vowel is on the quarter note and the tense vowel is on the half note, whereas Figure 8 is an example of mismatch stimulus, where the tense vowel is on the quarter note and the lax vowel is on the half note.

Figure 7. Example of match stimulus

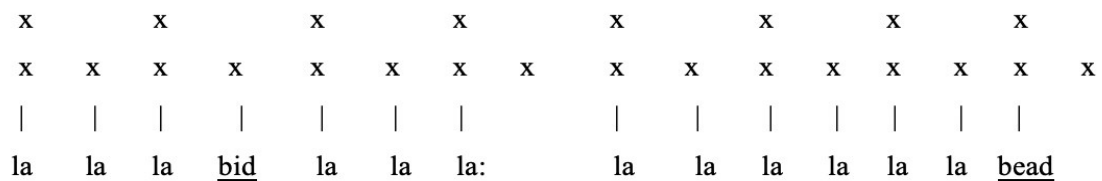
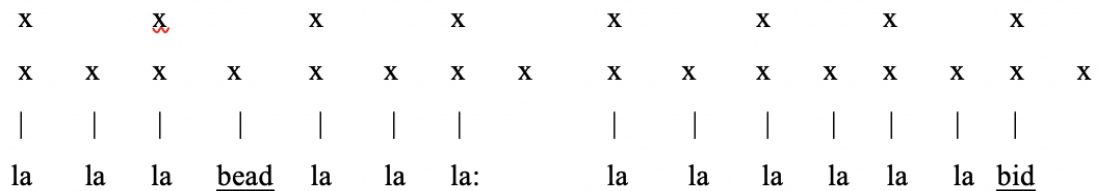


Figure 8. Example of mismatch stimulus



The task for experiment 1 was repetitive, and I needed to ensure that the participants were responding appropriately to their preferences and not clicking through the experiment aimlessly. Given that the average time to complete the study was low for the overall amount of work, I decided that using only the first 14 answers from each participant would provide data that more accurately represent the participants' judgements. Participants were asked to select their preferred stimulus for each match/mismatch pair. The order in which the

pairs were presented to the participants was randomized and only the first 14 answers from each participant were used for analysis. The order in which the two options were presented to the participants was randomized and it was therefore equally likely for a match or mismatch stimulus to be presented to the participant first.

### **3.1.3 Procedure**

Participants registered with a Carleton University SONA account were provided with a link to the external experiment on Qualtrics. First, they read the informed consent form, after which they proceeded to the instructions for the experiment. Participants were required to indicate that their audio was turned on so they would be able to hear the audio files. They were then instructed to complete experiment 1 which was followed by a 2-minute break.

## **3.2 Results**

All data was analyzed using IBM SPSS Statistics Data Editor. All the data was aggregated in excel into a summary format for data analysis. The results in this section are proceeding from more coarse-grained to more splintered. First, I looked at match/mismatch, then at tenseness and lengthening, then at the 4 groups presented in Table 6.

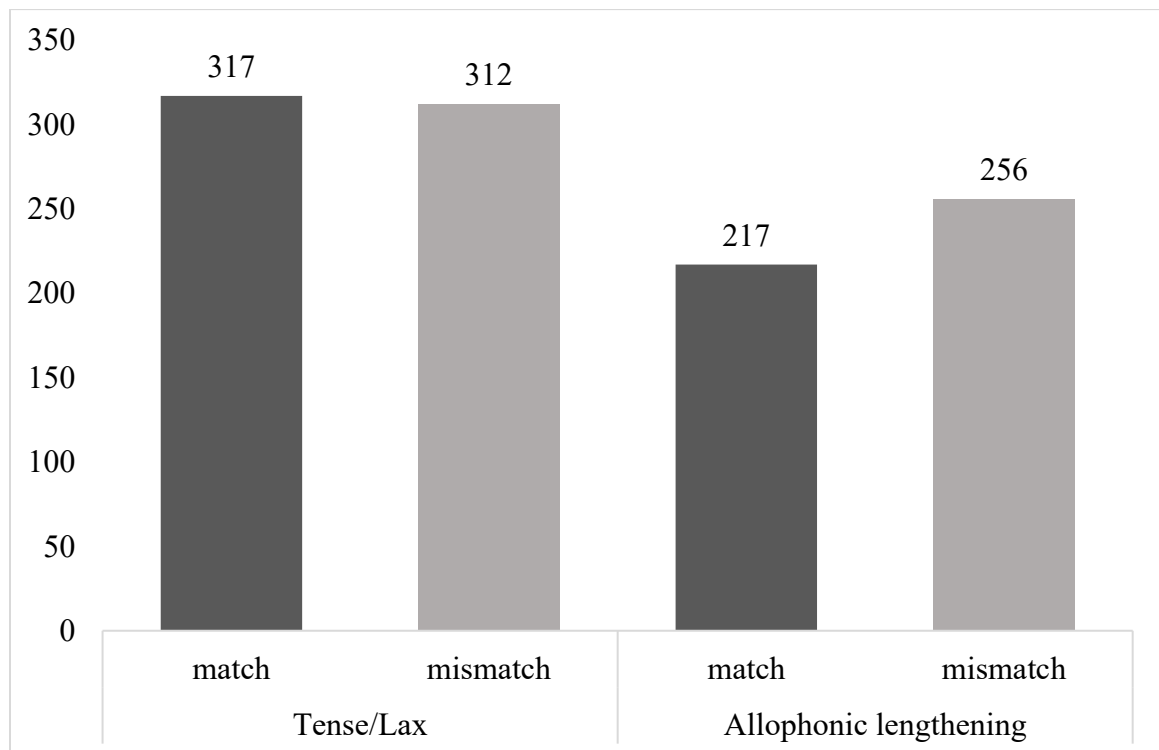
Participants selected the matched audio 534 times and the mismatched audio 568 times. In order to examine if the participants had a preference between the match stimuli and the mismatch stimuli a Chi Squared Goodness of Fit test was performed. This result was not significant,  $\chi^2(1) = 1, p = .31$ , illustrating that the participants did not have a preference for either the matched or mismatched audio files.

In the tense/lax condition participants matched a total of 317 times and mismatched a total of 312 times. In order to examine if the participants had a preference to match in the tense/lax condition a Chi Squared Goodness of Fit test was performed. This result was not significant,  $\chi^2(1) = 0.04, p = .84$ .

In the allophonic lengthening condition, the participants matched 217 times and mismatched 256 times. In order to examine if the participants had a preference to match in the allophonic lengthening condition a Chi Squared Goodness of Fit test was performed. This result was not significant,  $\chi^2(1) = 3.22, p = .07$ .

This illustrates that the participants did not have a match preference within the tense/lax category or the allophonically lengthened category. These results are illustrated below in Figure 9.

Figure 9. Count Data of Variable Comparison



I then conducted a Chi-Squared Goodness of Fit test for equal groups for each of the 4 groups illustrated in Table 6. This was done to see if there was an effect for match or mismatch within each of the 4 groups. When participants were presented with group 1, a lax vowel with a consonant voicing contrast (*bid/bit*), they matched 120 times and mismatched 134 times. There was no significant effect,  $N= 254$ ,  $\chi^2(1) = 0.8$ ,  $p = .38$ . When participants were presented with group 2, a tense vowel with a consonant voicing contrast (*beat/bead*), they matched 166 times and mismatched 141 times. There was no significant effect,  $N= 307$ ,  $\chi^2(1) = 2.04$ ,  $p = .15$ . When participants were presented with group 3, a tense/lax contrast paired with a voiceless consonant (*bit/beat*), they matched 151 times and mismatched 171 times. There was no significant effect,  $N= 322$ ,  $\chi^2(1) = 1.24$ ,  $p = .26$ . When participants were presented with group 4, tense/lax contrast paired with a voiced consonant (*bid/bead*), they matched 151 times and mismatched 122 times. There was no significant effect,  $N= 273$ ,  $\chi^2(1) = 3.08$ ,  $p = .08$ . This section illustrates that there was no significant effect found thus far. These results (or lack thereof) follow the same trend that was seen before I made the decision to include just the first 14 answers from each participant.<sup>2</sup>

After finding that there was no effect within each of the four word type categories, I analyzed the actual duration of the target vowels in the stimuli. To analyze the vowel duration I used a voice analysis software, *Praat* (Boersma et al., 2021). The duration of every vowel from each minimal pair was exported. This was done in order to examine

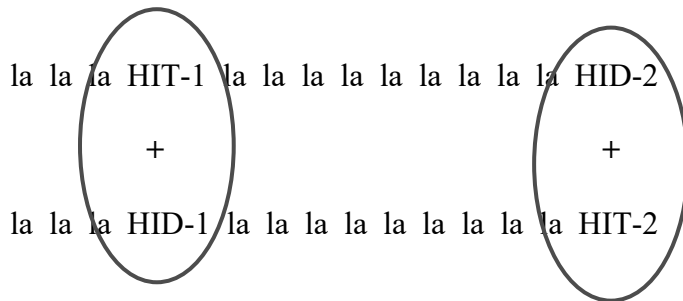
---

<sup>2</sup> It turns out that including each participants first 14 responses doesn't change the results, however the results from before the transformation are not included in this thesis.

whether the duration properties of the vowels, as they were recorded, had an effect on the participants' choices.

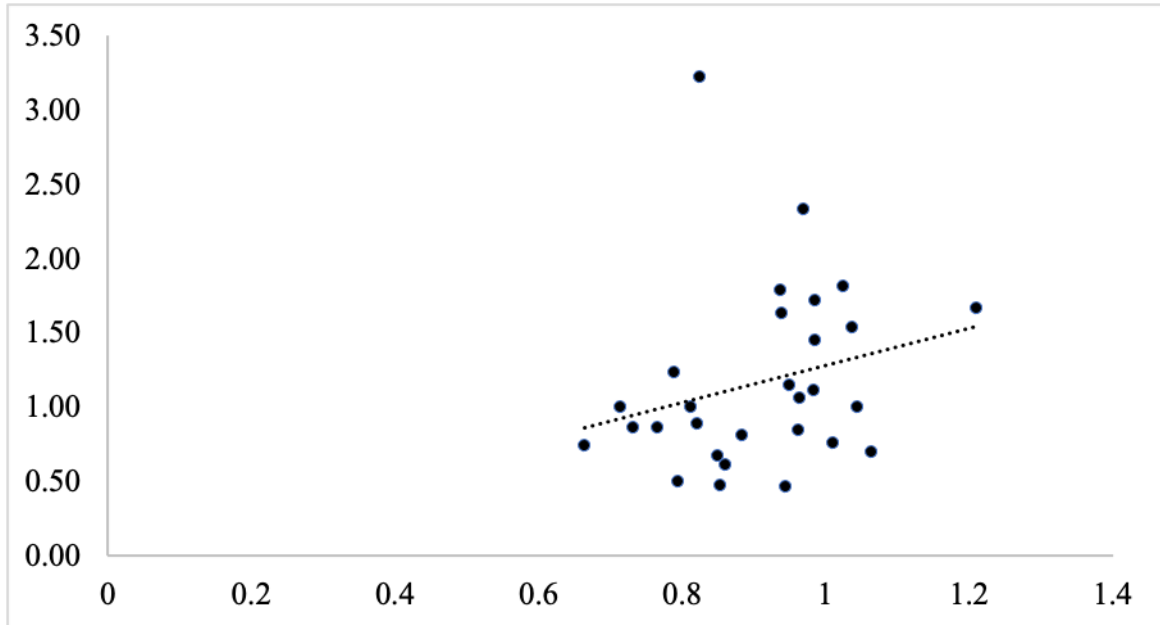
The vowels on the quarter note and the half note were analyzed separately. First, I calculated the ratio of the durations of the two vowels placed on the quarter note: one short and the other long. This provided one duration ratio per pair. I coded these pairs using a numerical system. The number that is paired with the word indicates which position the word was in: 1 being the quarter note and 2 being the half note. For example, HIT-1 represents *hit* on the quarter note and HIT-2 represents *hit* on the half note. HID-1 represents *hid* on the quarter note and HID-2 represents *hid* on the half note.

Figure 10. Duration Ratio



Given this format, the duration ratio of HIT-1 and HID-1 was calculated. Using these duration ratios, I ran a logistic regression and used the ratio as a continuous predictor variable of the participants' choice to match/mismatch. For the short condition we found that there was a significant effect,  $\chi^2(1) = 5.45, p < .005$ . The Nagelkerke  $R^2$  indicated approximately 0.6% of the variance in the short condition was accounted for by the predictor overall. For every 1 unit increase in the duration ratio, participants were 3.22 times more likely to get a match outcome than a mismatch outcome. This is illustrated in Figure 11 below.

Figure 11. Quarter Note



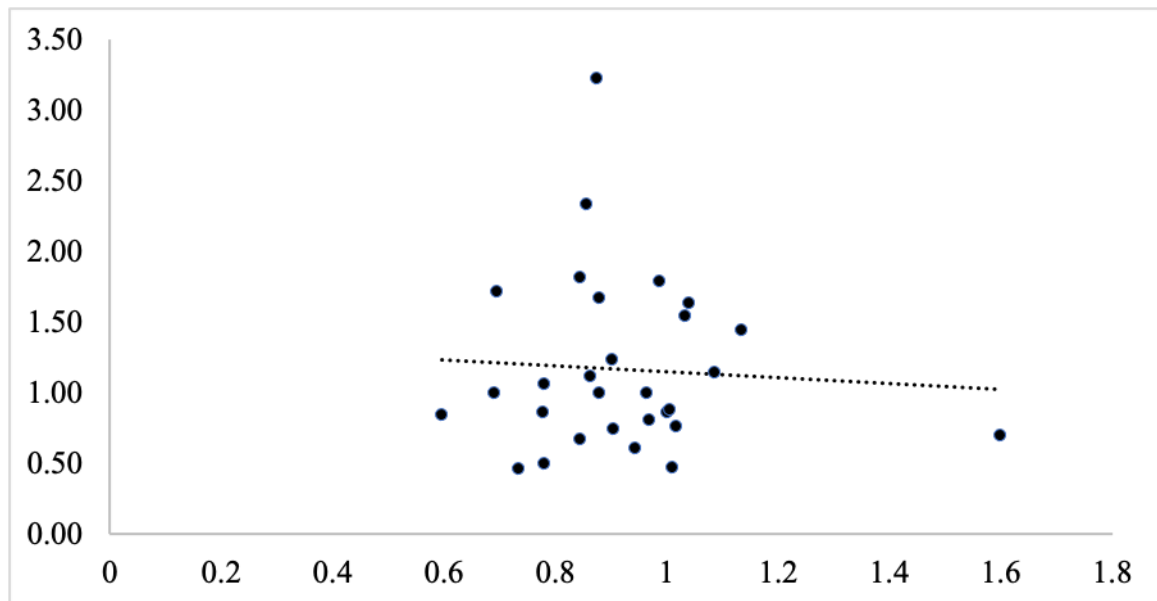
*Note:* The y-axis is the match/mismatch ratio, and the x-axis is the short/long duration ratio.

In this Figure each minimal pair is represented with a point in the graph. Along the x-axis is the duration ratio that was calculated for each pair, by dividing the short duration by the long duration. Along the y-axis is the match/mismatch ratio, calculated by dividing the number of matches by the number of mismatches. This result tells us that as the duration ratios get larger on the quarter note, participants become more likely to select a match outcome.

In order to analyze if the actual phonetic output and pronunciation of the vowels on the half note had an isolated effect on the outcome match, I took the ratio of the durations of the two vowels on the half note. The duration ratios were calculated in the same way as on the quarter note, however, now instead of HIT-1 and HID-1 I used the duration ratios of HIT-2 and HID-2. A logistic regression was then conducted on the duration ratios of the vowel pairs in the long condition. The long condition was not significant.  $\chi^2(1) = 0.158$ ,

$p = .691$ . The Nagelkerke  $R^2$  indicated 0% of the variance in the long condition was accounted for by the predictor overall.

Figure 12. Half Note



*Note:* The y-axis is the match/mismatch ratio, and the x-axis is the short/long duration ratio.

These results show that the actual duration of the vowels on the half note had no effect on the participants' choice to match or to mismatch.

The results from this experiment suggest that participants had no preference towards matching.

### 3.3 Discussion

This experiment examined whether participants preferred audio files that matched vowel length to the length of a musical note. Specifically, I was looking to see if participants would prefer shorter vowels on short notes and longer vowels on long notes. In English there are not phonemically long or short vowels; however, tense vowels in English



have a longer duration than lax vowels and vowels before voiced consonants are longer than vowels before voiceless consonants. This experiment yielded no significant results and did not confirm my prediction that participants would have a preference for matched over mismatched stimuli. Meaning, participants did not have a preference as to what word was placed on the quarter note and what word was placed on the half note. Likewise, participants did not have a preference towards matching in either the tense/lax condition or the allophonic lengthening condition.

Once finding that participants did not have a preference for which word went onto the long or short note, I analyzed the stimulus duration. This was done in order to see if the duration of the vowels, as they were recorded, had an effect on the participants' choices. I found that as the duration ratios on the quarter note get closer to 1, the participants were more likely to select the match. This means that there was a preference for longer vowels on the quarter note. However, for the half note there was no significant correlation between the vowel duration ratios and the matches/mismatches. The fact that the results indicate a preference for relatively longer vowels on the quarter note is interesting because the study was conducted with the idea that a low ratio would be expected to correlate with matching when considering the short note, while a high ratio would be expected to correlate with choice matching when considering the long note. In other words, we expected that participants would prefer a vowel with a relatively shorter duration on the quarter note and a vowel with relatively longer duration on the long note. Instead, there was a preference for longer duration on the quarter note, and there was no preference at all for the long note.

The findings in experiment 1 suggest the possibility that participants had no preference as to what was on the longer or the shorter notes. This does not support the original

hypothesis when creating the experiments, however, this result could be genuine, in that there really is no matching preference, or else due to external factors. Such factors could be participants using poor-quality headphones causing them to have difficulty hearing the audio files, the audio files may have lagged or skipped due to software error or the participants may not have actually listened to the sound files. Given the online format of this past year I was unable to perform the experiment in person with an ideal set up. These results could also be due to the actual task the participants were asked to perform. The task was repetitive: the participants had 29-word pairs and two songs per pair. The repetitiveness made the experiment extremely boring. This could be why, even with all the efforts made, the analysis yielded a null result.

## Chapter 4: Experiment 2

### 4.1 Methodology

#### 4.1.1 *Participants*

The participants in experiment 2 were the same as in experiment 1.

#### 4.1.2 *Experimental Design*

The musical phrase from *Twinkle Twinkle Little Star* that was used in experiment 1 was also used in experiment 2. However, for this experiment the minimal pairs were not already placed into the phrase. There was a play button beside the top line that the participants could press to listen to the song with the two target notes left silent. Each stimulus consisted of a minimal pair that differed in vowel length because of vowel tenseness or allophonic lengthening. Participants were asked to drag one member of the minimal pair into one of two boxes, which correspond to the missing notes labeled *blank 1* for the short note and *blank 2* for the long note, as seen in Figure 13 below.

Figure 13. Stimulus for Experiment 2: Blanks in the Musical Phrase

la la la \_\_\_\_\_ la la la la la la la \_\_\_\_\_  
(blank 1) (blank 2)

Please place the words **hit** and **hid** onto the blanks where you believe they sound the best.

Items	Blank 1	Blank 2
hit		
hid		

After one selection was made, the second member of the minimal pair was automatically placed into the remaining slot. The order in which the minimal pairs were presented to the participants was randomized. In addition, the way that the words in each pair were presented to the participant was randomized. Once the participant had completed the task and both terms were assigned to a note, participants would then click on the arrow bringing them to their next question. Once the participants arrived at the next question they could not go back and change their answers. There was no time limit. The participants repeated this task 29 times with 29 different minimal pairs. A full list of all the minimal pairs used are in *Appendix A*. The minimal pairs included in experiment 2 are the same pairs used in experiment 1.

#### 4.1.3 Procedure

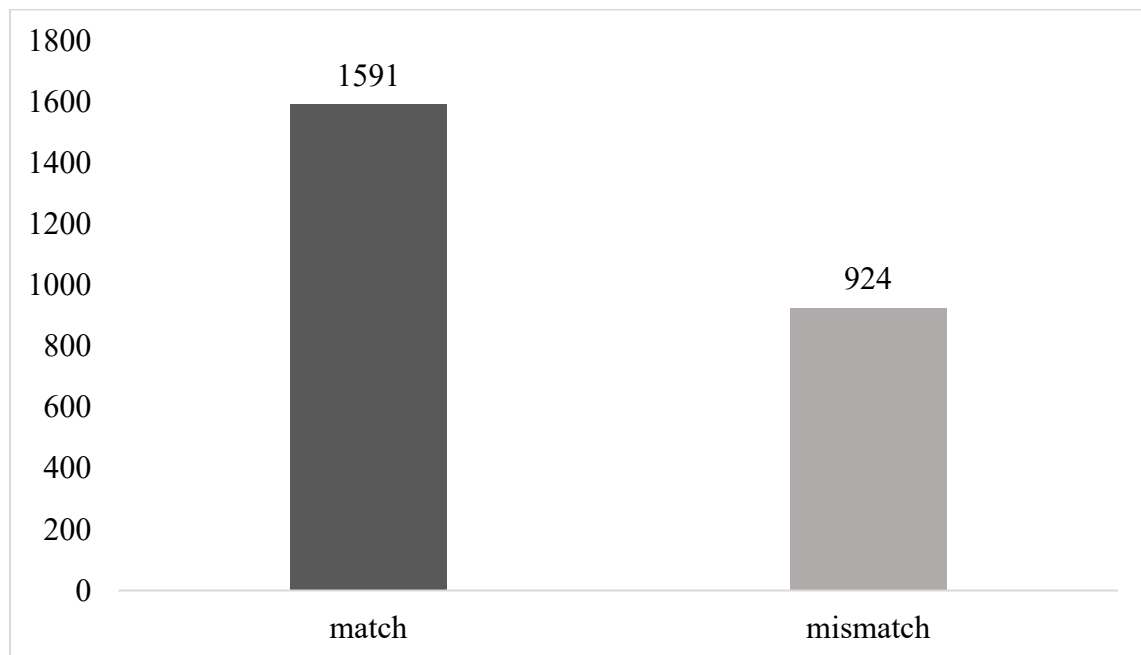
Following the 2-minute break after the first experiment, participants were asked to

begin the second experiment. Upon completion of the experiment, participants were presented with a debrief form and notified that they could no longer revoke their answers.

## 4.2 Results

All data was analyzed using IBM SPSS Statistics Data Editor. All the data was aggregated in excel into a summary format for data analysis. In order to examine if the participants had a preference between matching or mismatching, a Chi-squared Goodness of Fit test was performed. Participant matched vowel length to the length of the note 1591 times and mismatched 924 times. This result is significant,  $N= 2515$ ,  $\chi^2(1) = 176.9$ ,  $p < .01$ , illustrating that the participants matched vowel length to the length of the metrical note in a way that is more consistent than chance.

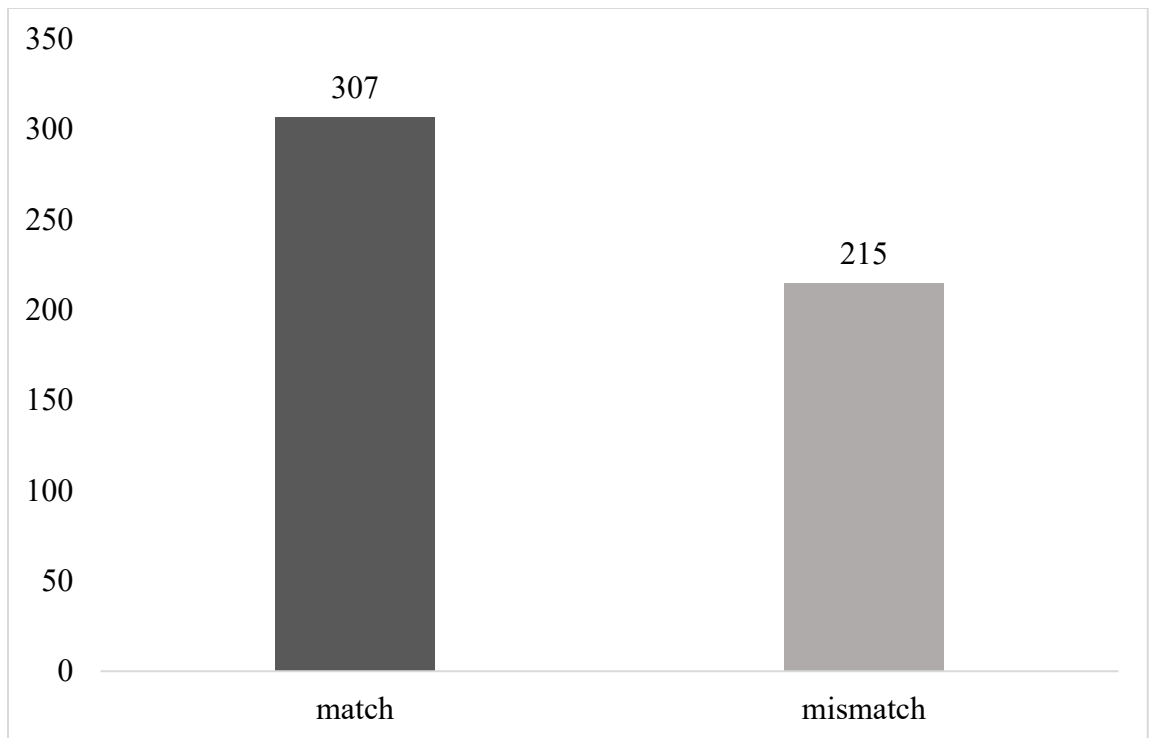
Figure 14. Match Mismatch Data



A Chi-Squared Goodness of Fit test for equal groups was then conducted for each of the four groups presented in Table 6, to test for significance within each group. When

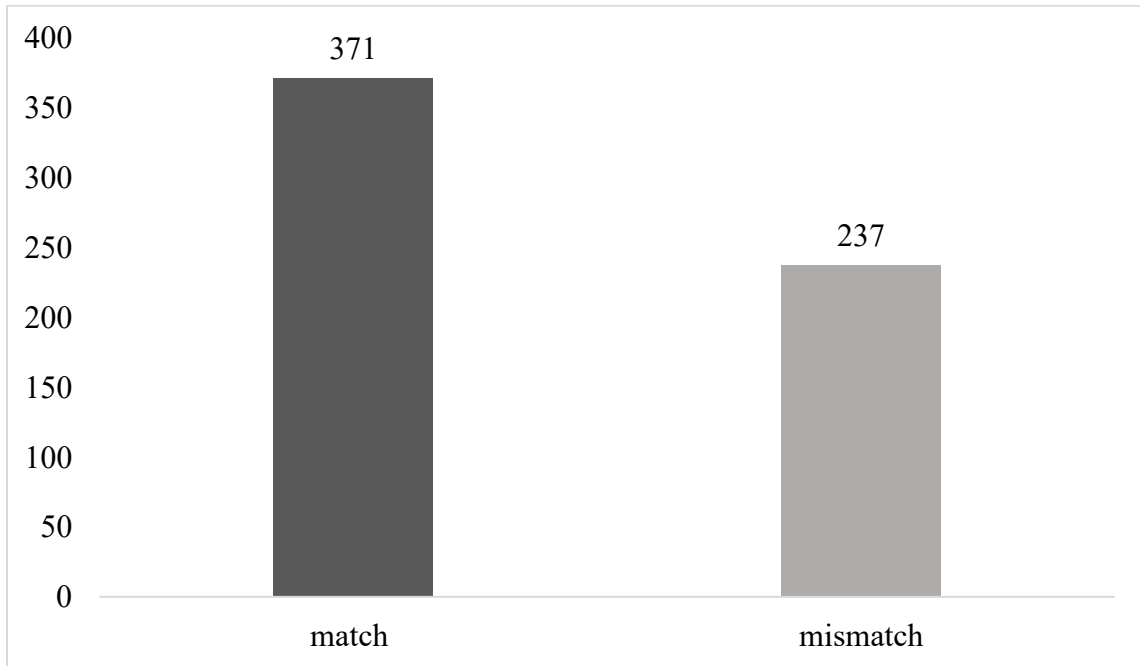
participants were presented with group 1, a lax vowel with a consonant voicing contrast, they matched 307 times and mismatched 215 times. There was a significant effect,  $N= 522$ ,  $\chi^2(1) = 16.2$ ,  $p < 0.01$ . Meaning, participants were more likely to match than mismatch when presented with a lax vowel with a consonant voicing contrast.

Figure 15. Match/Mismatch Rates for Group 1 (bit/bid)



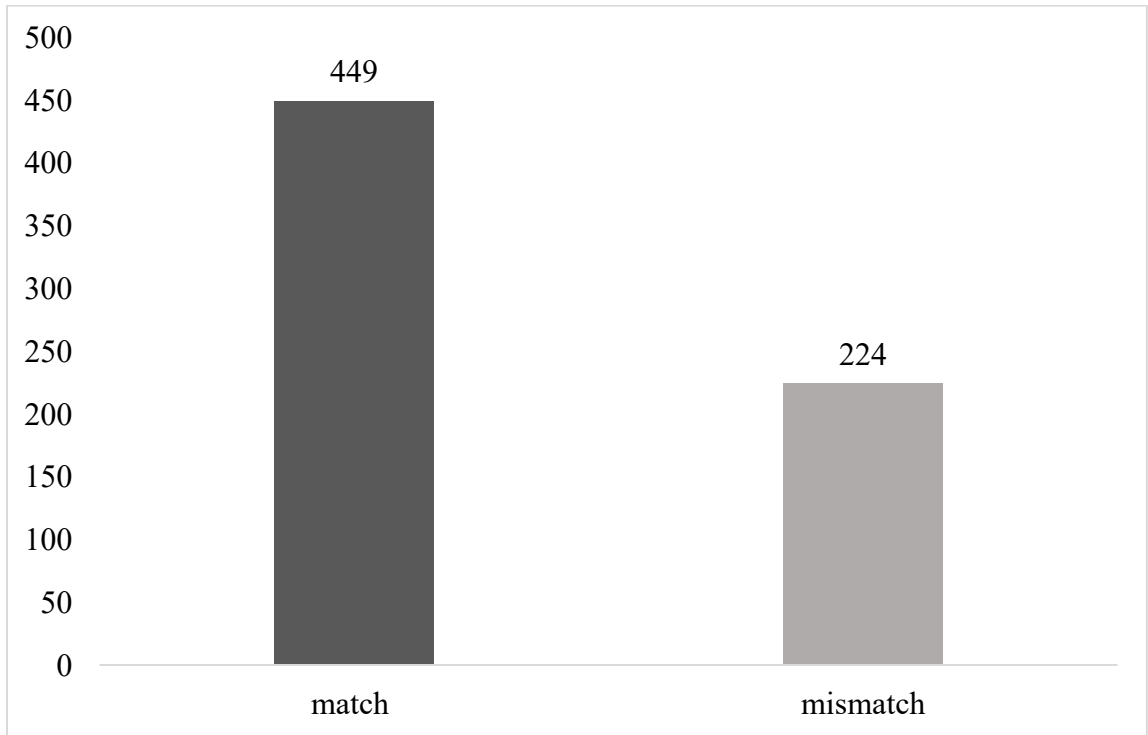
When participants were presented with group 2, a tense vowel with a consonant voicing contrast, they matched 371 times and mismatched 237 times. There was a significant effect,  $N= 608$ ,  $\chi^2(1) = 29.533$ ,  $p < 0.01$ . Meaning, participants were more likely to match when presented with a tense vowel with a consonant voicing contrast.

Figure 16. Match/Mismatch Rates for Group 2 (beat/bead)



When participants were presented with group 3, a tense/lax contrast paired with a voiceless consonant, they matched 449 times and mismatched 244 times. There was a significant effect,  $N= 693$ ,  $\chi^2(1) = 60.642$ ,  $p < 0.01$ . Meaning, participants were more likely to match when presented with a tense/lax contrast paired with a voiceless consonant.

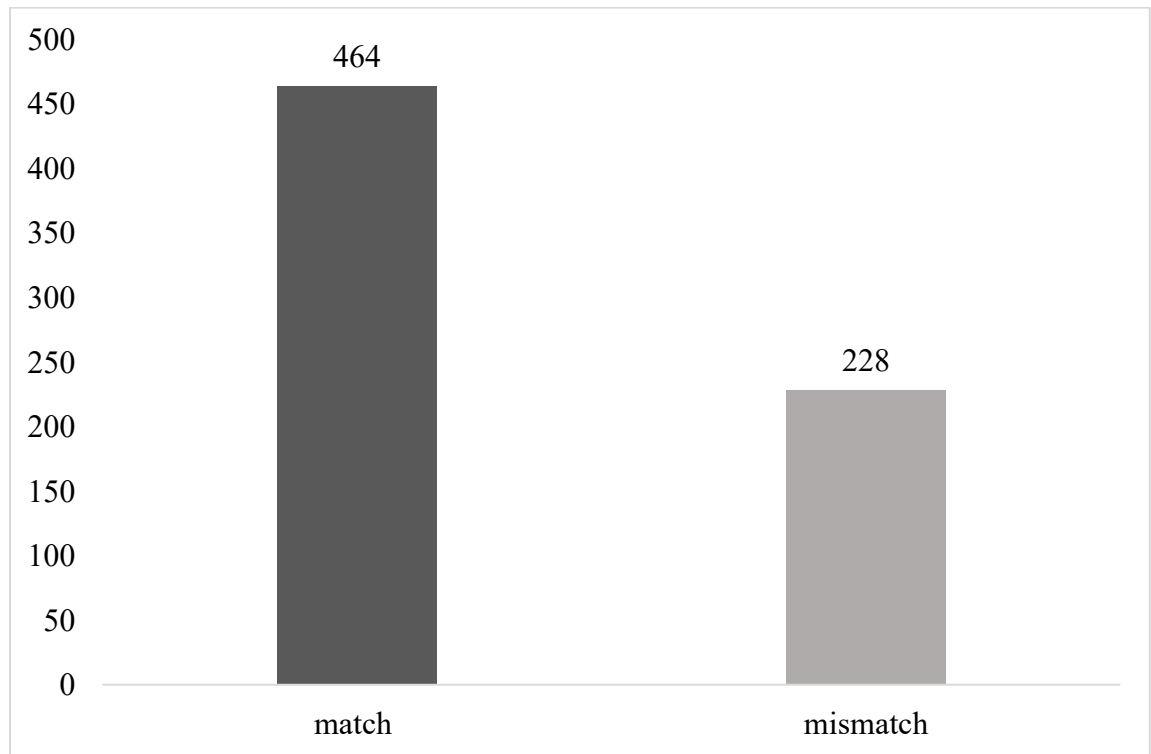
Figure 17. Match/Mismatch Rates for Group 3 (bit/beat)



When participants were presented with group 4, a tense/lax contrast paired with a voiced consonant, they matched 464 times and mismatched 228 times. There was a significant effect,  $N= 692$ ,  $\chi^2(1) = 80.486$ ,  $p < .01$ . Meaning, participants were more likely to match when presented with a tense/lax contrast paired with a voiced consonant.

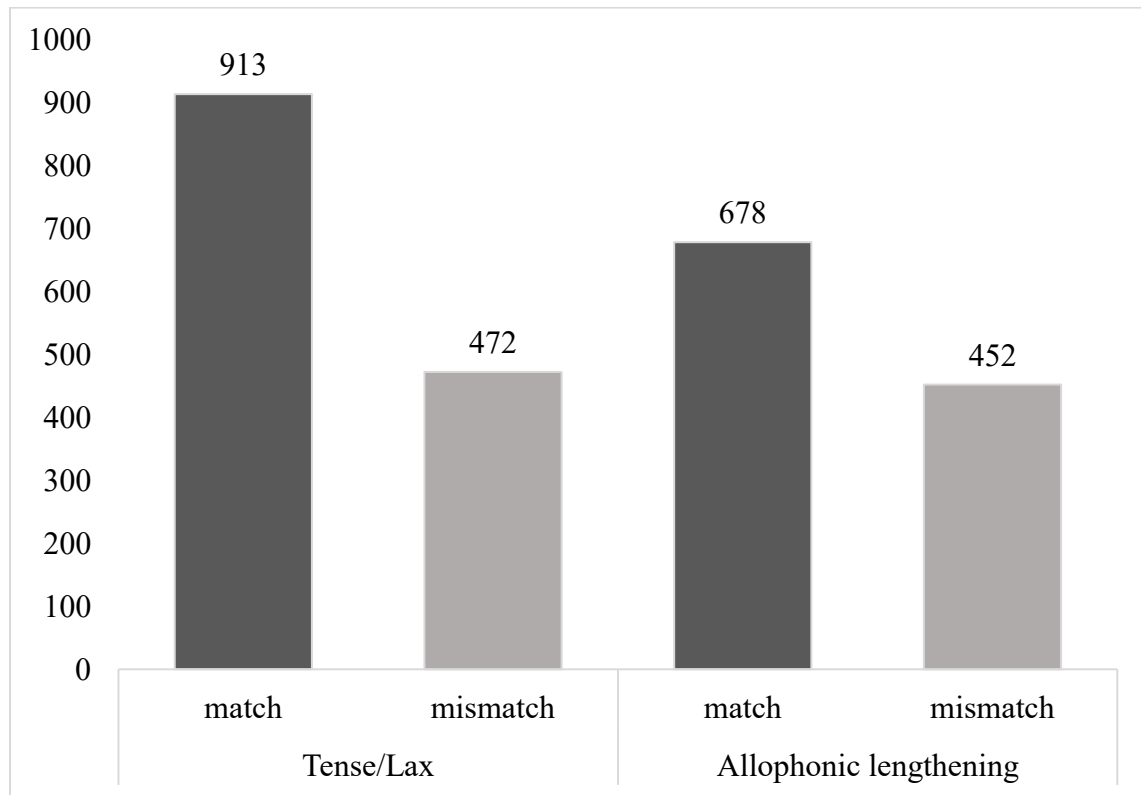


Figure 18. Match/Mismatch Rates for Group 4 (bid/bead)



In order to examine if either vowel tenseness or allophonic lengthening had a stronger effect, I conducted a logistic regression. The independent variable was the word pairs, coded for voicing or tenseness, and the dependent variable was the participants' choice (match or mismatch). I found a significant effect,  $\chi^2(1) = 9.372$ ,  $p < .01$ . The Nagelkerke pseudo/ $R^2$  indicated approximately 0.5% of the variance in the match condition was accounted for by the predictors overall. The effect of vowel tenseness was 1.29 times stronger than the effect of allophonic lengthening, though both groups were a significant predictor of the match/mismatch outcome. When participants were presented with the tense-lax vowel distinction they were 29% more likely to match than when they were presented with the allophonic lengthening condition.

Figure 19. Variable Comparison



While vowel duration and allophonic lengthening are both significant, when choosing which word to put where, the participants were more strongly guided by the tenseness of the vowel rather than the allophonic lengthening.

### 4.3 Discussion

This experiment examined whether allophonic lengthening and/or vowel tenseness had an effect on a participant's choice to match vowel length to the length of a metrical note. Specifically, I was investigating whether participants would prefer to match shorter English vowels to shorter notes and longer English vowels to longer notes in a musical phrase. This experiment illustrated that both the tense/lax and the allophonic lengthening

conditions are significant predictors of how the participants will respond. Participants preferred to match shorter vowels with the quarter note and the longer vowels to the longer note in a way more consistent than chance.

When participants were choosing which word to put where they were more strongly guided by the tenseness of the vowel than the allophonic lengthening.

These results are consistent with the existing literature on lengthening in textsetting. The previous research on textsetting shows that textsetting is sensitive not only to phonological configurations of stress but also to phonetic details in those configurations (Hayes & Kaun, 1996). As stated previously, this means that speakers are subconsciously aware of gradient durational differences, which can influence how language aligns with musical form.

## Chapter 5: General Discussion

In this section I discuss the results of the conducted experiments and whether they support the main hypotheses of this thesis. I explain what the results mean for theory and application and make links back to the previous literature, as well as discuss the experienced and potential limitations of this research.

This thesis was conducted to investigate the effects of vowel duration in a textsetting grammar. More specifically, this study explores the effects of inherent vowel durations using minimal pairs differing in vowel tenseness, and word-final /t/ vs /d/ were used to probe the effects of voicing-induced allophonic length. This was done in hopes to find matching preferences between linguistic tokens and slots in music. These experiments tested both meter and duration in music simultaneously and do not distinguish between them.

The results from experiment 1 do not support the main hypothesis of this thesis. It failed to show that tenseness or allophonic lengthening affect the outcome match or mismatch. This finding suggests the possibility that participants had no preference as to what was on the longer or the shorter notes. When creating this experiment, I was expecting to find that participants would like to match vowel length to the length of the musical note, however, that was not seen. The lack of results in experiment 1 could have been due to CVC weight asymmetry.

Lunden (2013) shows that in many languages CVC syllables behave as light word-finally but heavy otherwise. She suggests this CVC weight asymmetry is motivated by facts of phonetic length and human perception. Lunden (2013) conducted a production experiment in Norwegian which shows a proportional increase for the rimes of heavy

syllables across positions, however Norwegian does have a short/long vowel contrast so these findings cannot be easily extrapolated to English. Lunden's reasoning is as follows: weight theory illustrates that a syllable, in a given position, is only heavy if it is longer than the rhyme of a CV syllable in the same position. While a CVC syllable reaches this significant threshold non-finally, it does not do so word-finally. As a result, word-final CVC is lengthened due to final lengthening, and the long/short distinction is not as noticeable, resulting in no length contrast. This occurs in experiment 1, especially to the line-final (half-note) words: they're extra-long due to both the length of the note and the phrase-final lengthening, and so the long/short distinction may be harder to notice. This would cause it to be difficult to judge the well-formedness of the pairs of lines.

These findings illustrate a potential reason as to why there were no matching preferences in experiment 1. The stimuli for this experiment consisted of all monosyllabic words with some in phrase final position. From this we can take away that perhaps there was not a strong enough effect of duration to influence the choice; for example, even though *bit* is worse than *beat* in that strong position, it is still good enough to be accepted.

Since the tense/lax and allophonic lengthening categories did not turn out significant, I checked whether the actual durations were predictive of the participants' choice. The results show that there was a preference for longer vowels on the quarter note and that, for the half note, there was no significant correlation between the vowel duration ratios and the matches/mismatches. This is an interesting result because the study was conducted with the idea that a low ratio would be expected to correlate with matching when considering the short note, while a high ratio would be expected to correlate when analyzing the long note. In other words, we expected that participants would prefer a vowel with a relatively

shorter duration on the quarter note and a vowel with a relatively longer duration on the long note. Instead, there was a preference for longer duration on the quarter note, and there was no preference at all for the half note. This could be a result of the length variability of English vowel.

In English, tense vowels can be much longer than lax vowels, but under certain conditions, such as textsetting, they can become shortened to roughly the same length as a lax vowel (Nwe et al., 2010). A sung vowel is lengthened or shortened according to the duration of the corresponding musical note; however this does not change the tenseness or the laxness of the vowel with respect to vowel formants (Nwe et al., 2010). This makes English vowels easy to distinguish between, regardless of their length. In experiment 1 we see that the tense vowels have a more variable length than the lax vowels. The durations of the tense vowels range from 190ms to 310ms on the quarter note and 450ms to 860ms on the half note, while the durations of the lax vowels range from 170ms to 280ms on the quarter note and 500ms to 710ms on the half note. These vowel durations illustrate a large range on the half note and little variability on the quarter note. A full list of all vowel durations from Experiment 1 can be found in Appendix B.

The lack of findings in experiment 1 could have been the result of the linguistic factors discussed or there could be a more superficial reason for these results. Participants may have been using poor quality headphones causing them to have difficulty hearing the audio files, the audio files may have lagged or skipped due to software error or the participants may not have actually listened to the sound files. Epidemiological restrictions prevented the experiment from being conducted in person. Furthermore, these results could be a due to the actual task the participants were asked to perform. The task was incredibly

repetitive; the participants had 29-word pairs and two songs per pair, the repetitiveness made the experiment extremely boring. Even with all efforts made this could be why there was no match/mismatch effect or result.

Experiment 2 used the same phrase from *Twinkle Twinkle Little Star*; however, participants were not required to listen to sound files. This experiment examined whether participants would prefer to match shorter English vowels to shorter notes and longer English vowels to longer notes in a musical phrase. The participants were instructed to drag one member of the minimal pair, into the musical phrase, onto one of two blanks corresponding to the short and the long note. Since this was a binary choice the other member of the pair was automatically assigned to the remaining blank. The results show that participants do prefer to match shorter vowels with the quarter note and the longer vowels to the half note with better-than-chance consistency and illustrated that both the tense/lax and the allophonic lengthening conditions are significant predictors of how participants will respond. These results support the hypothesis of the experiment and the existing literature on lengthening in textsetting: longer syllables occupy a longer metrical span (Hayes & Kaun, 1996). If we only take into account experiment 2, the results suggest that the hypothesis about vowel length matching is true, which suggests that in English the duration of a vowel does have an effect of placement preferences within music.

However, the most interesting finding is that the length differences that are phonemic behave differently from length differences that are allophonic. Both the tenseness and allophonic lengthening conditions were found to be statistically significant, however, the results show that tenseness of a vowel has more of an effect than allophonic lengthening. This may be due to the fact that tense and lax vowels have a perceptual length difference,

in principle, that can be perceived auditorily by native speakers, while it is harder to detect length differences as a result of allophonic lengthening. For example, a native speaker of English would be more likely to detect a length difference between the words *bid* and *bead* but perhaps not, *bit* and *bid*.

These results illustrate that textsetting interacts with language not only on the surface and underlying forms, but that the underlying form matters more. The tense/lax distinction in English functions differently than vowel quantity, though it acts similarly by creating light and heavy syllables. However, quantity induced by voicing does not function in the same way. This is interesting because it delves into the interface between phonetics and phonology and enforces the idea that experimental phonetic research has informed phonology.

Kawahara (2011) discusses that phonological patterns can be explained in terms of articulatory and perceptual factors, and therefore purely phonological explanations without considering phonetic substances can be misinformed. Likewise, perceptual factors have been found to affect phonological patterns. Therefore, phonetic and psycholinguistic reasons are behind many, if not all, phonological patterns. According to Ohala (1990), excluding phonetic and psycholinguistic factors in phonological theory would be misguided. In experiment 2 we see that phonetics offers insights into phonological patterns. Furthermore, in experiment 2 we see that the phonetic factors have more of an effect than the phonological patterns.

When creating these experiments, I did not state that participants need to be native speakers of English, which could have contributed to the results obtained and be a limitation in the overall design of the experiments. Kluender et al. (1988) discuss how languages



with vowel quantity often do not experience voicing induced lengthening. English has a large vowel-length difference before voiced and voiceless consonants, but other languages tend to show a much smaller difference, or no difference, that is undetectable by listeners. For example, this lengthening effect does not occur in some dialects of Spanish, suggesting that durational variations are, in part, learned by English speakers and not by Spanish some speakers (Kluender et al., 1988). In future studies, asking the participants about their language background could help identify any potential issues presented in the data. Not having any language screening questions could have led to variability in the responses based on what the speakers L1 is.

In the existing literature that explores textsetting, the main focus is on syllables and how syllables stress assignment aligns with meter. The experiments presented in this thesis are particularly interesting because we can now look at the phonetic components that make up these syllables to decipher where and how they align within a musical phrase.

## Chapter 6: Conclusion

In music, the alternation of strong and weak beats at various hierarchical levels is known as metrical structure. Likewise, in language stress is a similar structure as it is usually conceived of as a strong and weak alternation, though less regular than in music. In texts that have both linguistic and musical structure, the two rhythms prefer to match but the alignment is not always perfect, though it is systematic. In this thesis, I have investigated one aspect of a textsetting grammar, namely how vowel length differences, due to tenseness and allophonic lengthening, interact with this system. More specifically, I have investigated the narrow issue of placement of monosyllabic words on strong or weak beats and their impact their duration. The results from Experiment 2 show that English speakers prefer to match longer vowels with long metrical notes and shorter vowels with short metrical notes. These results also show that the tenseness and allophonic lengthening conditions are both significant, however, when participants are choosing which word to put where in a musical phrase, they were more strongly guided by the tenseness of the vowel rather than the allophonic lengthening.

Future directions for this research could be to run the experiments again but with the half note not in phrase final position. As discussed, vowels lengthen in word-final CVC as well as phrase finally. Avoiding this position may help identify why there were no intuitive results found in experiment 1.

The experiments in this thesis are the first to bring together English vowel length, allophonic lengthening and meter into one experiment using textsetting. The existing literature that explores textsetting focuses on syllables and explores how syllables align with meter, however, this study provides evidence that phonetic properties correlate with

perceptual length differences and result in fine grained preferences within a textsetting grammar.

## References

- Blumenfeld, L. (2016). Generative metrics: An overview. *Language and Linguistics Compass*, 10(9), 413–430. <https://doi.org/10.1111/lnc3.12204>
- Boersma, P., & Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.42, retrieved 15 April 2021 from <http://www.praat.org/>
- Cooper, A., Wang, Y., & Ashley, R. (2017). Thai rate-varied vowel length perception and the impact of musical experience. *Language and Speech*, 60(1), 65–84. <https://doi.org/10.1177/0023830916642489>
- De'Ath, L., & Neubert, N. (2019). Reading and listening between the lines: Ideas on singing the short and open vowels [i], [o], and [Y], and the long and closed vowels [e:], [o:], and [ø:] in German. *Journal of Singing*, 76(5), 567-570.
- Dell, F. (1989). Concordances rythmiques entre la musique et les paroles dans le chant. l'accent et l'e muet dans la chanson francaise. In Dominicy, M. (Ed.), *Le souci des apparences*. Brussels: Editions de l'Université de Bruxelles, pp. 121–136.
- Dell, F., & Halle, J. (2005, April 8). Comparing musical textsetting in French and in English songs [Paper presentation]. *Typology of Poetic Forms*. <https://johnhalle.com/musical.writing.technical/fdell-jhalle-comparing-settings-4>
- Dell, F. (2009). Review article: meter in poetry. *Canadian Journal of Linguistics*, 54(2). 401–15. 2015.
- Dell, F., & Elmedlaoui, M. (2008). Poetic meter and musical form in Tashlhiyt Berber songs. Cologne: Rüdiger Köppe.

- Halle, J., & Lerdahl, F. (1993). A generative textsetting model. *Current Musicology*, 55, 3-23.
- Halle, M., & Idsardi, W. (1995). General properties of stress and metrical structure. *The handbook of phonological theory*, ed. by J. Goldsmith, 403–44. Oxford: Blackwell.
- Halle, M., & Keyser, S. J. (1966). Chaucer and the study of prosody. *College English*, 28(3), 187–219. <https://doi.org/10.2307/374038>
- Halle, M., & Keyser, S. J. (1971). Illustration and defense of a theory of the iambic pentameter. *College English*, 33(2), 154–176. <https://doi.org/10.2307/374744>
- Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky, & G. Youmans (Eds.) *Phonetics and Phonology: Vol. 1. Rhythm and meter* (pp. 201-260). Academic Press.
- Hayes, B., & Kaun, A. (1996). The role of phonological phrasing in sung and chanted verse. *The Linguistic Review*, 13, 243–304.
- House, A. (1961). On vowel duration in English. *The Journal of the Acoustical Society of America*, 33(9), 1174–1178. <https://doi.org/10.1121/1.1908941>
- House, A., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25(1), 105–113. <https://doi.org/10.1121/1.1906982>
- Kawahara, S. (2013). Experimental approaches in theoretical phonology. In M. van Oostendorp, C.J. Ewen, E. Hume K. Rice (Eds.), *The Blackwell companion to phonology: Vol. 4. Phonological Interfaces* (pp. 2283–2303). Blackwell-Wiley.

- Kawahara, S., & Braver, A. (2013) The phonetics of multiple vowel lengthening in Japanese. *Open Journal of Modern Linguistics*, 3(2), 141-148.  
doi: 10.4236/ojml.2013.32019.
- Klatt, D. (1973). Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America*, 54(4), 1102–1104.  
<https://doi.org/10.1121/1.1914322>
- Kluender, K., Diehl, R., & Wright, B. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, 16(2), 153–169. [https://doi.org/10.1016/S0095-4470\(19\)30480-2](https://doi.org/10.1016/S0095-4470(19)30480-2)
- Leung, K. K. W., Jongman, A., Wang, Y., & Sereno, J. A. (2016). Acoustic characteristics of clearly spoken English tense and lax vowels. *The Journal of the Acoustical Society of America*, 140(1), 45-58. doi:10.1121/1.4954737
- Lunden, A. (2013). Reanalyzing final consonant extrametricality: A proportional theory of weight. *The Journal of Comparative Germanic Linguistics*, 16(1), 1–31.  
<https://doi.org/10.1007/s10828-013-9053-3>
- Lunden, A. (2017). Duration, vowel quality, and the rhythmic pattern of English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1), 27.  
doi: <http://doi.org/10.5334/labphon.37>
- McPherson, L. (2019). Musical adaptation as phonological evidence: Case studies from textsetting, rhyme, and musical surrogates. *Language and Linguistics Compass*, 13(12). <https://doi.org/10.1111/lnc3.12359>

- Nwe, T. L., Dong, M., Chan, P., Wang, X., Ma, B., & Li, H. (2010). Voice conversion: From spoken vowels to singing vowels. *Journal of the Acoustical Society of America*, 1421-1426. doi:10.1109/ICME.2010.5582961
- Ohala, J. J. (1990). There is no interface between phonology and phonetics: A personal view. *Journal of Phonetics*, 18(2), 153–171.  
[https://doi.org/10.1016/S0095-4470\(19\)30399-7](https://doi.org/10.1016/S0095-4470(19)30399-7)
- Palmer, C., & Kelly, M. H. (1992). Linguistic prosody and musical metre in song. *Journal of memory and language*, 31(4), 525-542. [https://doi.org/10.1016/0749-596X\(92\)90027-U](https://doi.org/10.1016/0749-596X(92)90027-U)
- Patel, A. (2008). *Music, language, and the brain*. Oxford University Press.
- Peterson, G., & Lehiste, I. (1960). Duration of Syllable Nuclei in English. *The Journal of the Acoustical Society of America*, 32(6), 693–703.  
<https://doi.org/10.1121/1.1908183>
- Plag, I., Arndt-Lappe, S., Braun, M., & Schramm, M. (2015). *Introduction to English Linguistics* (3rd ed.). De Gruyter.
- Rodríguez-Vázquez, R. (2010). Text-setting constraints: A comparative perspective. *Australian Journal of Linguistics*, 30(1), 19–34.
- Ryan, K. (2014). Onsets contribute to syllable weight: Statistical evidence from stress and meter. *Language*, 90(2), 309–341.
- Smiljanic, R., & Bradlow, A. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics*, 36(1), 91-113.  
<https://doi.org/10.1016/j.wocn.2007.02.002>

- Sundberg, J. (1969). Articulatory differences between spoken and sung vowels in singers. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 10(1), 33-46.
- Sundberg, J., & Skoog, J. (1999). Formant frequencies in country singers' speech and singing. *Journal of Voice*, 13(2), 161-167.  
[https://doi.org/10.1016/S08921997\(99\)80020-4](https://doi.org/10.1016/S08921997(99)80020-4)
- Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., & Munhall, K. (2006). Effects of exposure and training on perception of Japanese length contrasts by English listeners. *The Journal of the Acoustical Society of America*, 120(5), 3209.  
<https://doi.org/10.1121/1.4788124>
- Thomas, D. A. (1995). *Music and the origins of language: Theories from the French enlightenment*. Cambridge: Cambridge University Press.



## Appendices

### Appendix A: Full List of Minimal Pairs Used

hit hid

hid heed

sit Sid

meet mead

bit bid

took toque

hit heat

sit seat

fit feet

greet greed

fit fid

look Luke

grit grid

mid mead

pull pool

seat seed

boot bood

grit greet

fid feed

mitt mid

heat heed

full fool

Sid seed

bit beat

bid bead

beat bead

mitt meet

grid greed

feet feed

## Appendix B: Experiment 1 Full List of Vowel Durations

Match			Mismatch		
Minimal Pair	Token	Duration/ms	Minimal Pair	Token	Duration/ms
hit hid	hit	174	hid hit	hid	258
	hid	623		hit	527
hid heed	hid	227	heed hid	heed	296
	heed	632		hid	597
sit Sid	sit	146	Sid sit	sid	225
	sid	687		sit	621
meet mead	meet	228	mead meet	mead	273
	mead	725		meet	639
bit bid	bit	233	bid bit	bid	227
	bid	549		bit	513
took toque	took	155	toque took	toque	239
	toque	676		took	553

hit heat	hit	185	heat hit	heat	220
	heat	613		hit	554
sit seat	sit	195	seat sit	seat	225
	seat	651		sit	565
fit feet	fit	207	feet fit	feet	207
	feet	728		fit	591
greet greed	greet	173	greed greet	greed	119
	greed	606		greet	439
fit fid	fit	204	fid fit	fid	211
	fid	599		fit	566
look Luke	look	171	Luke look	luke	244
	Luke	638		look	545
grit grid	grit	132	grid grit	grid	155
	grid	456		grit	442
mid mead	mid	273	mead mid	mead	294
	mead	694		mid	690

pull pool	pull	109	pool pull	pool	106
	pool	560		pull	543
seat seed	seat	223	seed seat	seed	271
	seed	667		seat	570
boot booted	boot	254	booted boot	booted	213
	booted	692		boot	591
grit greet	grit	173	greet grit	greet	153
	greet	476		grit	451
fid feed	fid	241	feed fid	feed	305
	feed	743		fid	680
mitt mid	mitt	220	mid mitt	mid	231
	mid	642		mitt	641
heat heed	heat	209	heed heat	heed	302
	heed	700		heat	640
full fool	full	188	fool full	fool	176
	fool	701		full	623

Sid seed	Sid	223	seed Sid	seed	270
	seed	590		sid	576
bit beat	bit	230	beat bit	beat	234
	beat	687		bit	663
bid bead	bid	264	bead bid	bead	235
	bead	868		bid	623
beat bead	beat	253	bead beat	bead	206
	bead	551		beat	530
mitt meet	mitt	237	meet mitt	meet	230
	meet	731		mitt	666
grid greed	grid	167	greed grid	greed	114
	greed	638		grid	453
feet feed	feet	236	feed feet	feed	282
	feed	667		feet	669

## **Appendix C: Quality Assurance**

This experiment required ethics approval through the Carleton University Ethics Board because the experiment required human participants, however, the anticipated risk associated with the experiment for participants was low. Participants were asked to listen to auditory stimuli then were asked numerous questions in regard to what they heard. These questions appeared on a computer or mobile screen in front of the participant. To obtain informed consent from the participants, I had them read the consent form at the beginning of the survey. This experiment did not target vulnerable populations and participants were able to withdraw their consent at any point before or during the experiment process.