

Tel Aviv University

Lester and Sally Entin Faculty of the Humanities

Department of Linguistics

**Rhythmic Similarities between Language and Music:
Jazz and Bluegrass Musicians as a Case Study**

MA thesis submitted by

Udi Wahrsager

Prepared under the guidance of

Dr. Evan Gary Cohen

July, 2021

Abstract

This work explores the interaction between speech and musical rhythm by analyzing comparable rhythmic patterns in utterances and musical phrases produced by musicians. Following Patel & Daniele (2003a), comparative studies of language and music have used the normalized Pairwise Variability Index (nPVI) for quantifying degrees of durational variability in spoken utterances and musical phrases. Speech nPVI measurements show that languages traditionally classified as stress-timed, such as English and Dutch, are characterized by greater variability (higher nPVI) of vowel durations across utterances, compared to languages classified as syllable-timed, such as French and Spanish (Grabe & Low, 2002; Ramus, 2002). Patel & Daniele used the nPVI to measure the variability of note durations in musical themes by British and French composers, and found a correlated pattern of higher nPVI values for English speaking composers in both music and speech. A similar pattern was found in studies of spontaneous speech and musical performance by musicians, who speak different dialects of English and specialize in different musical styles (McGowan & Levitt, 2011; Carpenter & Levitt, 2016).

In this paper, I follow these works by comparing musicians of two distinct styles of American music – jazz and bluegrass. Unlike previous studies, I calculate speech nPVI values based on syllable durations rather than vowel durations, under the view that the syllable is the basic rhythmic unit of speech and most comparable to musical tones (Patel, 2008). To minimize subjective judgement in syllabification, I rely on criteria used by automatic syllabification models (Bartlett et al., 2009). On the musical domain, I distinguish between rhythmic patterns derived from the underlying metrical structure of musical phrases, representable by music notation, and durational nuances in music performance. The current study focuses on rhythmic patterns of the second type by constraining the data to metrically uniform phrases composed of consecutive eighth note rhythms. I propose that a comparable method can be applied to

speech. The main results of my study are compatible with previous findings, showing a correlated pattern of higher speech and musical nPVI values for jazz musicians, compared to bluegrass musicians. This work elaborates the discussion on the possible connections between speech and musical rhythm. It illustrates how such connections can be formed on distinct levels of rhythmic knowledge, and proposes additional methods and measures for the study of these connections.

Acknowledgments

In this work, I was fortunate to combine my main fields of interest, which I pursued separately until recently – human language and music. After completing my Bachelor degree in the TAU linguistics department, I abandoned the study of linguistics for many years to pursue a career in music, which included a diploma in jazz performance in Berklee College of Music in Boston, and playing professionally as a flutist in Israel. Thoughts about the ways in which these two different, and yet similar domains interact, led me to the writing of this thesis.

I would like to express my sincere gratitude and appreciation to those who helped me along the long and winding road that led to the writing of this thesis: To Dr. Lin Chalozin-Dovrat for her insight in the early stages of the work, to Dr. Israela Becker for invaluable help with the statistic analysis of the data and her constant support, to my thesis supervisor Dr. Evan Gary Cohen for always keeping my feet on the ground and being the fastest e-mail responder in the academic profession, and to my private M.A advisor, my beloved Dr. Ayelet Even-Ezra.

Table of contents

| | |
|---|----|
| 1. Introduction..... | 1 |
| 2. Theories of speech rhythm..... | 6 |
| 2.1. Rhythmic typology and the stress/syllable-timing classification..... | 6 |
| 2.2. Phonological properties of rhythmic classes..... | 8 |
| 2.3. Acoustic correlates of speech rhythm..... | 10 |
| 2.3.1. Vocalic and consonantal intervals | 10 |
| 2.3.2. Durational variability and the nPVI measure | 12 |
| 2.4. The syllable as a rhythmic unit | 18 |
| 2.5. Chapter summary | 23 |
| 3. The comparative study of speech and musical rhythm | 24 |
| 3.1. The nPVI as a measure of metrical variability..... | 24 |
| 3.1.1. Patel & Daniele (2003a)..... | 24 |
| 3.1.2. Interfering factors | 27 |
| 3.1.3. Folk vs. classical music..... | 29 |
| 3.1.4. Metrical structure vs. rhythmic feel..... | 30 |
| 3.2. Durational variability in speech and musical performance | 33 |
| 3.2.1. McGowan & Levitt (2011)..... | 34 |
| 3.2.2. Carpenter & Levitt (2016)..... | 37 |
| 3.3. Durational variability vs. metrical uniformity | 39 |
| 3.4. Chapter summary | 42 |
| 4. Jazz and Bluegrass musicians as a case study | 44 |
| 4.1. Relevant background on jazz and bluegrass | 44 |

| | |
|--|----|
| 4.2. Material selection | 49 |
| 4.2.1. The musicians..... | 49 |
| 4.2.2. Recorded samples | 50 |
| 4.2.2.1. Diachronic considerations | 51 |
| 4.2.2.2. Metrical uniformity | 51 |
| 4.2.2.3. Tempo, fluency and sample size..... | 54 |
| 4.3. Segmentation..... | 55 |
| 4.3.1. Speech analysis..... | 55 |
| 4.3.1.1. Syllabification..... | 55 |
| 4.3.1.2. Acoustic analysis of speech samples | 59 |
| 4.3.2. Analysis of musical samples | 62 |
| 4.4. Results..... | 66 |
| 4.5. General discussion..... | 73 |
| 5. Conclusion | 78 |
| References | 79 |
| Appendix A – The musicians | 83 |
| Appendix B – Sources of recordings | 83 |
| I. Speech recordings | 83 |
| II. Music recordings | 84 |
| Appendix C – Data Analysis | 85 |
| I. Speech data..... | 85 |
| II. Musical data | 90 |

1. Introduction

Language and music share some properties that make them inherently similar. Both are universal and unique forms of human expression. We do not know of any human society that does not express itself by language and music, and we do not know of any other species that does so. Both language and music are mediated by sound. This has led some to the idea that the two may in fact be identical, namely that music is a form of language. This idea seems to be quite popular among musicians, as explicated by the world-renowned bass player and educator Victor Wooten:

“Music is a language. Both music and verbal languages serve the same purpose. They are both forms of expression. They can be used as a way to communicate with others. They can be read and written. They can make you laugh or cry, think or question and can speak to one or many. And both can definitely make you move”.¹

From a more scientific perspective, a similar idea has been introduced by Katz & Pesetsky (2011), in their provocative Identity Thesis for Language a’nd Music, stating that “all formal differences between language and music are a consequence of differences in their fundamental building blocks ... In all other respects, language and music are identical”. A more skeptic view on this matter is expressed by Jackendoff (2009). Jackendoff concludes a review of various parallels and non-parallels between language and music by stating that although they share “a considerable number of general characteristics”, these characteristics do not “indicate a particularly close relation that makes them distinct from other cognitive domains”. Jackendoff urges caution in “drawing strong connections between language and music, both in the contemporary human brain and in their evolutionary roots”. Yet even under this “sober” view, as Katz & Pesetsky define it, Jackendoff recognizes one “important formal parallel between the two domains, perhaps shared by only music and language”, which is “the extent to which

¹ "Music as a Language – Victor Wooten", <https://youtu.be/3yRMbH36HRE>.

phonology and music are structured by very similar metrical systems”. Similarly, Heffner & Slevc (2015) “review the evidence for a link between musical and prosodic structure and find it to be strong”. They note that a focus on the syntactic domain in the study of language and music parallels left “the prosodic patterns of loudness, pitch, and timing that make up the rhythms and melodies of speech” relatively understudied.

Empirical work on speech and musical rhythm aims to test the hypothesis that the music of a culture reflects the rhythmic patterns of its language (Patel & Daniele, 2003a). So far, this work focused on the rhythmic property of *durational variability* – the degree of durational contrast between adjacent units of sound. Different languages and different musical styles are characterized by different degrees of durational variability. The so-called “Morse-code” rhythm of languages classified as stress-timed (e.g., English, Dutch, Thai) emerges from durational contrast between full and reduced vowels and between complex and simple syllables (Dauer 1983, 1987). Lack of vowel reduction and simpler syllabic structure in languages classified as syllable-timed (e.g., French, Spanish) results in the more uniform “machine-gun” rhythm. Similarly, the melody of “Twinkle, Twinkle Little Star” would sound more “jazzy” when adjacent notes are played slightly unevenly, and more “classical” when notes are played very evenly.

In this paper, I propose that the durational variability of sound sequences, in both music and speech, represents the intersection of two rhythmic dimensions, which can be generalized under the terms *macro-* and *micro-rhythm* or *macro-* and *micro-timing*. Macro-rhythmic patterns can be thought of as durational tendencies of speech and musical underlying metrical structures. Metrically prominent rhythmic units in both domains, such as stressed syllables in language and downbeat notes in music, tend to be acoustically realized by longer durations. This is not always the case, though. Motoric, acoustic and stylistic factors may cause a stressed syllable to be pronounced shorter than an adjacent unstressed syllable, or a downbeat note to be played

shorter than an adjacent upbeat note. The precise durational ratios between adjacent sounds are determined also by micro-timing nuances of speech and musical performance, which cannot be captured by discrete metrical grid representations.

Several studies have used the normalized Pairwise Variability Index (nPVI) as a measure of durational variability across utterances (Low et al., 2000; Grabe & Low, 2002; Ramus, 2002). Acoustic measurements showed greater variability (higher nPVI values) of vowel durations in typical stress-timed languages such as English, compared to typical syllable-timed languages such as French (see section 2.3.2 for details). Patel & Daniele (2003a) found a similar pattern of variability in musical themes by national British and French composers (e.g. Edward Elgar and Claude Debussy). By applying the nPVI to music notation, they found greater variability in themes by English speaking composers than in themes by French speaking composers. Patel & Daniele see this as empirical evidence for a connection between speech and musical rhythm.

Because music notation represents proportional metrical relations, rather than absolute durations of sound, I suggest that Patel & Daniele's nPVI data reflect the degree of **metrical variability** in musical sequences, abstracted of their actual durations on surface. More recent studies directly compare **durational variability** in speech and music performance, by acoustic analysis of spoken utterances and musical phrases produced by individual musicians (McGowan & Levitt, 2011; Carpenter & Levitt, 2016). Similar to Patel & Daniele, these studies found correlated patterns of variability in music and speech. Musicians who speak relatively "more stressed-timed" dialects of English produced both utterances and musical phrases with higher nPVI values than musicians who speak "less stress-timed" dialects (e.g. American English vs. Jamaican English). If such findings indicate on a genuine connection between speech and musical rhythm, we may wonder how such a connection can be formed. Can we situate this connection within current linguistic and musical models? Based on the above distinction, I believe that rhythmic connections between the domains can be formed on both

the macro- and micro-timing levels. On the macro-timing level, language and music could share similar structural properties of metrical grid representations. On the micro-timing level, musicians may apply similar durational nuances when pronouncing utterances in their language and playing phrases on their instrument. This distinction has not been explored in previous research. In this paper, I explore the second possibility, and propose a framework for a direct comparison of rhythmic variability on the micro-timing level.

In previous studies, musical phrases incorporate some degree of metrical variability by being composed of mixed metrical values (quarter notes, eighth notes, triplets etc.). This inevitably affects the degree of durational variability on surface in these phrases. To neutralize this effect and the effects of other confounds, I collected a corpus of **metrically uniform** phrases, composed only of consecutive eighth note rhythms, from recordings by influential musicians of two distinct styles of American music – jazz and bluegrass. This type of data allows us to focus on micro-timing durational contrast between adjacent musical tones, as a possible connection with speech rhythm. In addition, I collected a set of recorded utterances produced by the same musicians.

Jazz and bluegrass are characterized by distinct types of “rhythmic feel”. The central rhythmic characteristic of jazz is “swing feel”, in which adjacent eighth notes are performed with noticeably uneven durations. Swing or unevenness of eighth notes is less characteristic of bluegrass rhythm. Jazz and bluegrass also evolved by distinct linguistic communities – jazz by African-American musicians, originally in New-Orleans and then around NYC, and bluegrass by musicians of Celtic descent in the Appalachian region. A sub-genre of jazz, known as “cool jazz”, is associated with a distinct group of European-American musicians centered around L.A. in the West Coast region. In a small-scale study presented here, I compare nPVI data obtained from acoustic analysis of speech and musical recordings by musicians of these three distinct styles. Its results are compatible with previous studies, and support the idea that

musicians may use similar rhythmic patterns to pronounce their language and to play their instrument.

The paper is structured as follows: Chapter 2 reviews relevant theoretical background in the study of speech rhythm – the rhythmic classification of languages and its phonological and acoustic properties. Chapter 3 reviews previous evidence on musical correlations to speech rhythm and discusses the distinction between metrical and durational variability as properties of different rhythmic domains. Chapter 4 presents a comparative study of spoken utterances and musical phrases by jazz and bluegrass musicians. The conclusion follows in chapter 5.

2. Theories of speech rhythm

The comparative study of rhythm in language and music derives its theoretical foundation from developments in the study of speech rhythm during the last decades. This chapter is an overview of previous studies in the field.

2.1. Rhythmic typology and the stress/syllable-timing classification

The systematic study of speech rhythm originated from intuitive observations by linguists about the rhythmic patterns of different languages. Lloyd James' (1940, cited by Abercrombie 1967) "machine-gun" vs. "Morse-code" metaphor expressed the intuition that a language such as Spanish has a more uniform and regular rhythm ("machine-gun") compared to a more abrupt rhythm in English ("Morse-code"). Pike (1945) proposed that this difference reflects a universal dichotomy of languages based on their timing mechanism. Under Pike's theory, the so-called Spanish machine-gun rhythm emerges from an *isochronous* or *periodic* timing of syllables, such that they are distributed more or less evenly across the utterance. The rhythm of English, on the other hand, is constrained by the isochronous timing of *interstress intervals*, i.e. the intervals between stressed vowels (roughly corresponding to metrical feet). To satisfy this constraint, syllables in long interstress intervals undergo phonological reduction and deletion, while in shorter intervals, syllables can be fully pronounced resulting in the so-called Morse-code pattern. For instance, Pike argued that the following sentences are pronounced with "more or less equal lapses of time between the stresses" despite a different number of syllables in each interstress interval:

(1) a. The **teacher** is **interested** in **buying** some **books**.

b. **Big** battles are **fought** **daily**.

For a more scientific terminology, Pike classified languages as having either a *syllable-timed* or a *stress-timed* rhythm. Abercrombie (1967) proposed that Pike's classification is based on universal physiological principles. Following R.H. Stetson's (1951) Motor Phonetics theory,

Abercrombie argued that syllables and stresses correspond, respectively, to *chest pulses* and *stress pulses* of the pulmonic air-stream mechanism, and are coordinated in such a manner that either the first type or the second type of pulses is isochronous, but not both. Abercrombie claimed that this classification applies to “every language in the world” and accounts for rhythmic similarities between seemingly unrelated languages, such as French, Telugu and Yoruba which he classified as syllable-timed versus English, Russian and Arabic which he classified as stress-timed. A third class of *mora-timed* languages was later proposed, following Ladefoged’s (1975) claim that in Japanese “each mora takes about the same length of time to say”.

Despite its popularity, empirical evidence failed to support Pike and Abercrombie’s theory as a whole. While the notion of rhythmic classes was supported by independent evidence from early language acquisition, speech isochrony in its strict phonetic sense, lacked empirical support. Abercrombie’s chest and stress pulses theory was very soon rejected (Ladefoged, 1967). Early instrumental measurements of interstress intervals already by Classe (1939) did not show evidence for “perfect isochrony”. Subsequent studies consistently failed to indicate on interstress isochrony in stress-timed languages or syllable isochrony in syllable-timed languages (see literary reviews in Lehiste 1977, Bertinetto 1989, Nespor 1990). To account for the intuitive impression of speech rhythm as isochronous, Classe (1939) proposed that isochrony operates as an underlying tendency of speech, which is often suppressed from surfacing but nonetheless has a perceptual effect on the listener. Perceptual experiments by Lehiste (1977) provided empirical support for this idea (see also Couper-Kuhlen, 1993 and Schreuder & Gilbers, 2004).

Perhaps the most compelling evidence for the notion of rhythmic classes and the stress-timing/syllable-timing classification was provided by Nazzi et al. (1998). Nazzi et al. found that newborns could discriminate languages of different rhythmic classes, but not languages of

the same rhythmic class. In one experiment they found that French newborns were able to discriminate English from Japanese sentences (stress-timed vs. mora-timed), but not English from Dutch (both stress-timed). In another experiment, French newborns could discriminate combinations of sentences in English, Spanish, Italian and Dutch only when these sentences were grouped by the same rhythmic class (e.g. English and Dutch vs. Spanish and Italian). The sentences were acoustically processed to reduce segmental information by a low-pass filter at a cut-off frequency of 400 Hz, suggesting that discrimination was made primarily by prosodic and rhythmic cues. Such findings led Ramus et al. (1999) to conclude that “the syllable-timing/stress-timing dichotomy may well be deeply anchored in the human perceptual system”. According to Beckman (1992), this typology seems to capture fundamental facts about the rhythmic patterns of languages, otherwise “it would have been relegated to the dustbin long ago, because taken literally as a statement about constant interval durations, as originally proposed, it was very soon proved false”.

2.2. Phonological properties of rhythmic classes

In Pike and Abercrombie’s theory, speech isochrony is a phonological primitive constraining phonological structure. Isochrony of interstress intervals forces more phonological reduction in stress-timed languages than in syllable-timed languages and results in greater complexity of syllable structure. With no clear evidence for speech isochrony, the opposite causality has been proposed, where speech rhythm emerges as the effect or by-product of a language’s phonological structure. For example, Nespor (1990) argues that the stress-timing vs. syllable-timing classification “is the result of a series of **non-rhythmic** phonological processes rather than the cause of these processes”. A similar idea is expressed also by Dasher & Bolinger (1982). Dauer (1983) believes that this classification has little to do with timing, but instead it concerns the status of stress in the grammatical system of languages. Dauer proposes a more gradient approach to rhythmic classification based on the notion of how *stress-based* a

language is. Typical stress-timed languages tend to be highly stress-based in the sense that stress plays a central role in their grammatical system. For instance, Dauer argues that most stress-timed languages have lexical or word level stress while other languages typically do not (e.g. only phrase level stress in French or a system of pitch accent in Japanese). In stress-timed (or stress-based) languages various phonological and other grammatical factors interact to support stress use. Among these factors, vowel reduction and syllable complexity contribute to greater contrast between stressed and unstressed syllables, compared to languages which have been classified as syllable-timed. In data collected by Dauer she found that Spanish and French have a low frequency of closed syllables (30% and 26%, respectively) while in English they are more frequent than open syllables (56%). Due to their high frequency in Spanish, CV syllables were both the most frequent stressed and unstressed syllables, as opposed to English in which CVC syllables were most frequently stressed and CV syllables most frequently unstressed. More complex syllables in English (CCVC, CVCC, CVCCC) were most of the times stressed, compared to a vast majority of unstressed onset-less syllables (V, VC). In general, stress-timed languages are known to assign stress by syllable weight (e.g., Arabic and Thai). Languages classified as stress-timed such as English, Swedish, Russian and conversational Thai contrast stressed and unstressed syllables by various processes of vowel centralization, modification and shortening in unstressed syllables, while the full forms are preserved in stressed syllables. Syllable-timed languages usually do not exhibit vowel reduction in unstressed syllables, with some exceptions such as Catalan and to some sort Brazilian Portuguese (Nespor 1990). In Spanish, although phonemic word-level stress exists, a weaker durational contrast was found between stressed and unstressed syllables than in English and German (see references in Dauer, 1983). For instance, medial open syllables in Spanish were found to be only 1.1 times longer than unstressed syllables, compared to 1.6 times longer in English and 1.5 longer in German. In general, stress-timed languages seem to be

characterized by a larger inventory of syllables, more vowel reduction and greater phonetic contrast between stressed and unstressed syllables. Dauer's proposal allows a more scalar typology of languages according to their overall stress-based character:

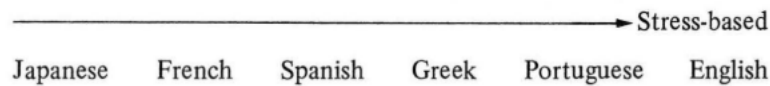


Figure 1: A comparison of languages according to their cumulative "stress-based" properties. Taken from Dauer (1983).

Unlike Pike and Abercrombie's dichotomous classification, Dauer's approach does not exclude "intermediate" or "mixed" languages, such as Polish with complex syllable structure and no vowel reduction, or Catalan with vowel reduction despite a relatively non-complex syllable structure, as was pointed out by Nespor (1990). In principle, we expect these languages to be located somewhere in the middle of Dauer's stress-based scale.

2.3. Acoustic correlates of speech rhythm

The ability of newborns to discriminate rhythmic classes in filtered speech suggests that acoustic cues for rhythmic classification are extractable from the acoustic signal. Because measurements of interstress intervals and syllable durations could not distinguish languages of different classes (Roach, 1982; Dauer, 1987), other acoustic measures have been proposed.

2.3.1. Vocalic and consonantal intervals

Ramus et al. (1999) assumed that the only perceptual distinction accessible to infants in early acquisition is between sequences of vowels and consonants. On this basis they measured durations of *vocalic intervals* (the duration between the onset and the offset of a vowel or a cluster of vowels) and *consonantal intervals* (the duration between the onset and offset of a consonant or a cluster of consonants) in eight different languages. Ramus et al.'s data were obtained from Nazzi et al.'s (1998) corpus plus additional recordings, and consisted of 5 utterances produced by 4 different speakers in each language, amounting to 20 utterances per language and (160 utterances total). The utterances were constructed as short news-like

sentences, roughly matched for the number of syllables and duration. Ramus et al. found two variables that seem to support the standard classification of languages as stress-, syllable- and mora-timed:

(i) proportion of vocalic intervals (%V) – the total duration of vocalic intervals in an utterance divided by the total duration of the full utterance.

(ii) variability of consonantal intervals (ΔC) – the standard deviation of the duration of consonantal intervals within the utterance.

These variables also seem to correlate with the underlying phonological properties of rhythmic classification suggested by Dauer. We expect vowel reduction and syllable complexity to be reflected in a lower %V and a higher ΔC in stress-timed languages compared to syllable-timed and mora-timed languages. That is, we expect stress-timed languages to have a smaller percentage of vowels and more variable durations of consonant clusters than in the other rhythmic classes. This seems to be supported by the chart in figure 2:

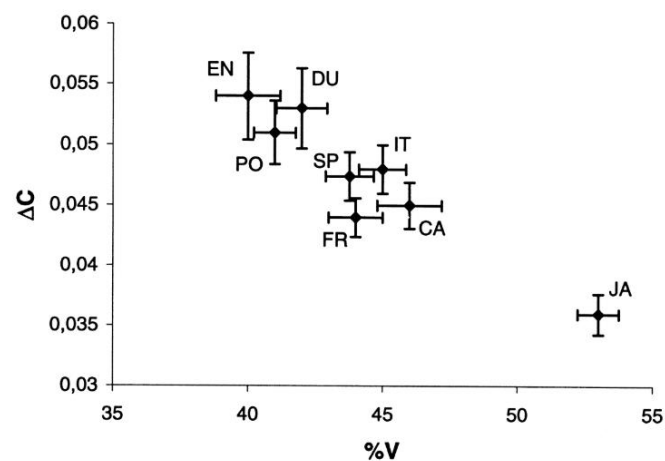


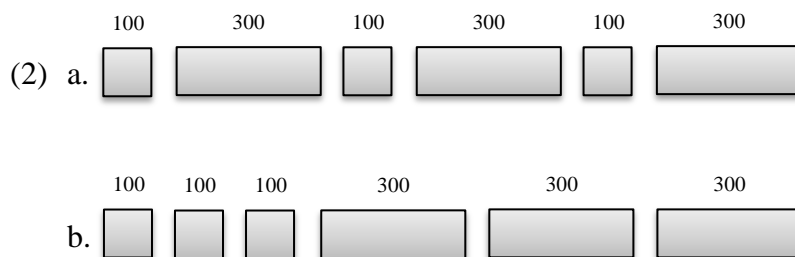
Figure 2: A comparison of %V and ΔC values in English, Dutch, Polish, Spanish, Italian, French, Catalan and Japanese. Taken from Ramus et al. (1999).

The distribution of languages by their %V and ΔC values corresponds more or less to their classification as stress-timed (English, Dutch), syllable-timed (Spanish, French, Italian) and mora-timed (Japanese). This chart does not indicate on a special status for the so-called mixed languages, Polish and Catalan. Polish seems to pattern here as stress-timed and Catalan as

syllable-timed. A distinct pattern for Polish did emerge by comparing ΔC with ΔV values (standard deviation of vocalic intervals) of the languages under study. This is further discussed in the following section.

2.3.2. Durational variability and the nPVI measure

Low et al. (2000) studied rhythmic differences between Singapore English (SE) and British English (BE). Low et al. argue that weaker phonetic contrast between full and reduced vowels contributes to the impression of SE as syllable-timed in comparison to the stress-timed BE. Reduced vowels in SE tend to be less shortened (and less centralized) than BE reduced vowels. As a result, Low et al. expected to find lower *vocalic durational variability*, i.e. less variation of vowel durations, in SE utterances than in BE utterances. Measurements of vocalic variability by standard deviation (ΔV) in Ramus et al. (1999) did not support a distinction between stress-timed and syllable-timed languages. Low et al. suggest that standard deviation could be inaccurate as a measure for durational variability. As they point out, the hypothetical sequences of vowel durations (2a) and (2b) have the same mean (200 ms) and the same standard deviation from the mean (100), yet their pattern of variation in is quite different:



Sequence (2a) resembles the “Morse-code” rhythm attributed to stress-timed languages. It alternates regularly from short to long vowels (100 ms to 300 ms). Sequence (2b) alternates only once from short to long. This could perhaps be the result of a sudden decrease in speech rate in the middle of the utterance. Apart from that, successive vowels in this utterance remain even. On the overall, (2b) resembles “machine-gun” or syllable-timed rhythm. What seems to

be at issue here is the specific linear order of each sequence and the variation between successive events. ΔV cannot capture this because standard deviation is indifferent to linear order. Instead, Low et al. use a Pairwise Variability Index (PVI) as a measure for durational variability. The PVI averages the absolute differences of successive pairs in a sequence of vowel durations. A higher PVI value indicates a higher degree of vocalic durational variability and vice versa. When applied to sequences (2a) and (2b), we see that the “stress-timed” sequence (2a) receives a considerably higher PVI value of 200, compared to a value of 40 for the “syllable-timed” (2b):

$$(3) \text{ a. } PVI(2a) = (|100-300|+|300-100|+|100-300|+|300-100|+|100-300|)/5 = 200$$

$$\text{b. } PVI(2b) = (|100-100|+|100-100|+|100-300|+|300-300|+|300-300|)/5 = 40$$

Example (2) also illustrates that changes in speech rate could obscure more underlying patterns of long-short alternation of successive speech sounds. While these patterns could be accessible to the listener by accommodating for speech rate changes, absolute durational values would fail to reflect them. For this purpose, a normalization component was added to the “raw” PVI (rPVI) calculation, by dividing the difference between each pair of durations by their average. The formula for normalized PVI (nPVI) calculation is given below (values are multiplied by 100 for readability):

(4) nPVI formula:

$$nPVI = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \frac{|d_k - d_{k+1}|}{\frac{d_k + d_{k+1}}{2}}$$

where d_k is the k^{th} interval in a sequence of m intervals

Low et al. recorded a corpus of 200 constructed utterances in BE and SE (100 each). These utterances were divided to two sets: (i) a reduced vowel set, in which some of the vowels were reducible vowels (could potentially undergo reduction), and (ii) a full vowel set, in which all

vowels were non-reducible. If BE and SE contrast as stress-timed and syllable-timed, we expect to find a significantly stronger effect for vowel reduction in BE compared to SE. This was supported by Low et al.’s findings:

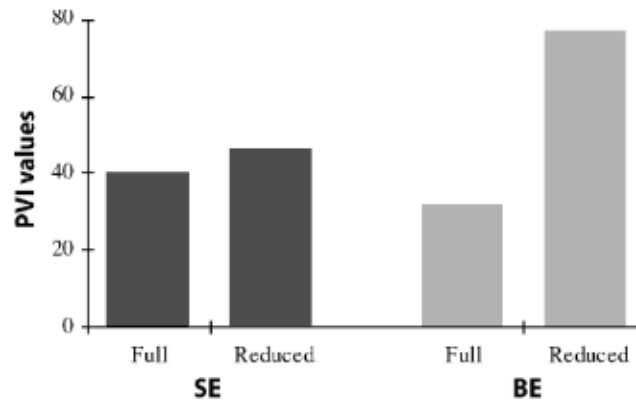


Figure 3: nPVI values of SE and BE utterances. “Full” stands for utterances containing only non-reducible vowels, and “Reduced” for utterances containing reducible and non-reducible vowels. Higher nPVI values indicate a higher degree of vocalic durational variability. Taken from Low et al. (2002).

Whereas BE exhibited a highly significant difference between the full and reduced vowel sets, no significant difference was found between the two sets in SE. In addition, the SE reduced vowel set had significantly lower durational variability (lower nPVI) than the BE reduced vowel set.

On a cross-linguistic level, Grabe & Low (2002) analyzed comparable utterances in 18 languages, produced by one speaker in each language. For each utterance they calculated its vocalic nPVI value and intervocalic, i.e. consonantal, rPVI value (unnormalized).² These data show that typical stress-timed languages are characterized by a high vocalic nPVI whereas typical syllable-/mora-timed languages are characterized by a low nPVI. According to Grabe & Low, no significant difference was found between these languages on the intervocalic level.

² See Grabe & Low (2002) section 2.4 for a discussion on the effects of normalization for vocalic and intervocalic intervals.

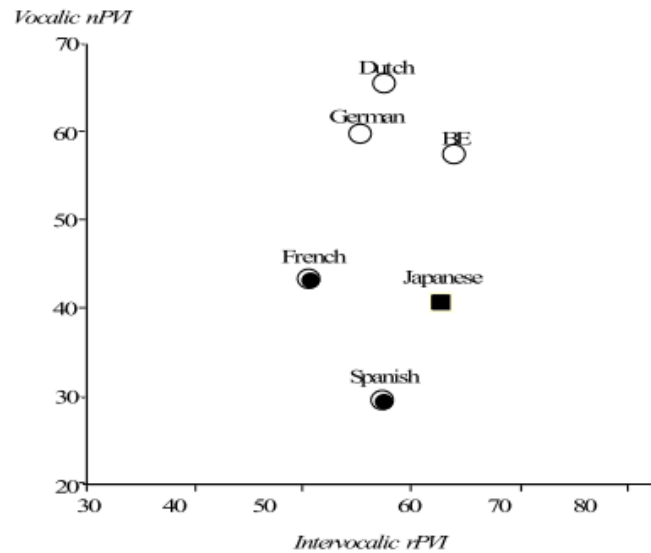


Figure 4: Vocalic nPVI values of typical stress-timed languages (Dutch, German, BE), syllabled-timed languages (French, Spanish) and mora-timed Japanese plotted against their intervocalic rPVI values. Taken from Grabe & Low (2002).

However, when less prototypical languages are taken into account, Grabe & Low argue that “a strict categorical distinction between stress-timing and syllable-timing cannot be defended”. Instead, their data point to a model in which languages can be “more or less stress-timed or syllable-timed” (as proposed by Dauer, 1983):

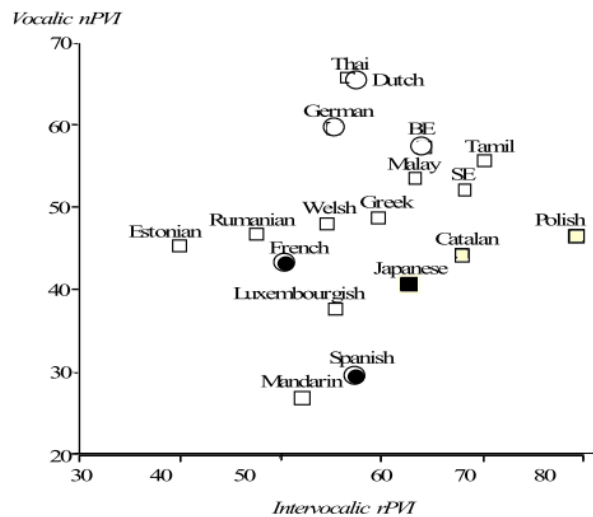


Figure 5: Vocalic nPVI and intervocalic rPVI measurements for 18 languages. Taken from Grabe & Low (2002).

Unclassified languages such as Welsh and Greek were found to considerably overlap “with the edges of the stress-timed and the syllable-timed group”. Japanese, according to their data, “is not in a rhythm class of its own”. These data also seem to support the Nespor’s (1990) notion of Polish as a “mixed” language. Complex syllable structure places Polish highest on the intervocalic axis, while lack of vowel reduction places it close to French on the vocalic axis. As for Catalan, while we would expect to find an opposite pattern, this does not seem to be the case. Despite having vowel reduction, Catalan patterns like French on the vocalic nPVI axis. Grabe & Low suggest that as in the case of SE, Catalan vowel reduction could be phonetically weaker than in stress-timed languages. Grabe & Low conclude that their data support a “weak categorical distinction” between languages which have been traditionally classified as stress-timed and syllable-timed, but that this distinction does not apply to all languages.

Finally, to compare the PVI measures with Ramus et al.’s (1999) intervallic measures, Grabe & Low computed %V and ΔC values over their corpus. This led to some mismatches such as Catalan and Japanese having lower %V (lower proportion of vowels) than German, and Polish having both a low %V and high ΔC (a typical stress-timed profile), contrary to Nespor’s prediction. In response, Ramus (2002) points to methodological differences between the two studies. Grabe & Low’s corpus was based on a single speaker per language. This required a normalization procedure of the PVI measure to minimize the effects of speech rate variability among individual speakers. Because standard deviation is sensitive to speech rate changes, ΔC measurements over Grabe & Low’s data gave strange results. Ramus et al.’s corpus, as opposed to that, was produced by four speakers per language and controlled for speech rate by matching the number of syllables and average duration of all utterances. Under these conditions there was less need for normalization, and standard deviation produced more consistent results. To illustrate this, Ramus compared measurements of vocalic and intervocalic standard deviation

(ΔV and ΔC) with vocalic and intervocalic PVI over Ramus et al.'s corpus. He found these results to be “strikingly similar”:

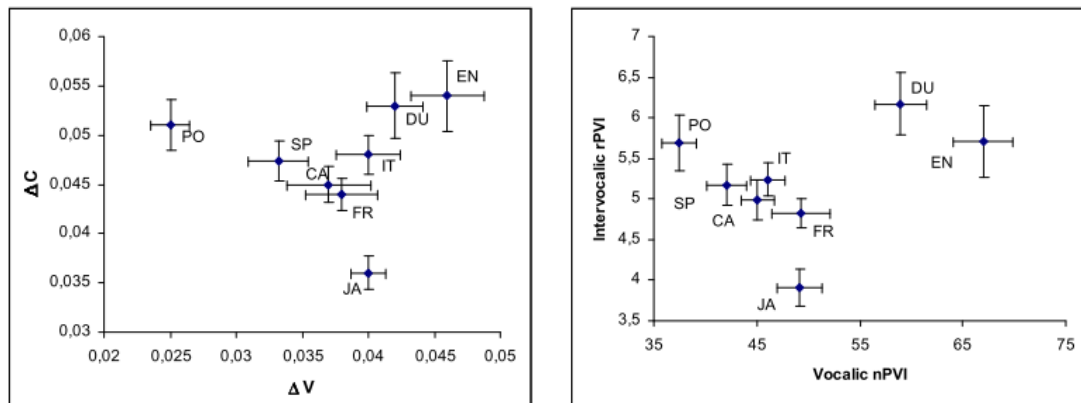


Figure 6: A comparison of vocalic and intervocalic variability in Ramus et al.'s corpus by standard deviation (ΔC and ΔV) and PVI measurements. Calculated by Ramus (2002).

In Ramus et al.'s corpus, the typical stress-timed languages (English, Dutch) and syllable-timed languages (Spanish, Italian, French) exhibit a similar pattern both by the standard deviation variables (ΔV and ΔC) and the PVI variables. Under both sets of variables Polish clearly patterns independently of other rhythmic classes, while Catalan seems to pattern in both cases as syllable-timed. Finally, both types of variables support here a distinct classification for Japanese. Ramus concludes that standard deviation measures may be effective only when “speech rate is strictly controlled”, and that controlling for speech rate is essential “either by constraining the corpus, or by using a normalization procedure”. This seems to make the nPVI a more flexible rhythmic measure, as it can be applicable also when speech rate cannot be controlled, for instance in spontaneous speech utterances. But as Ramus points out, standard deviation variables can also be normalized in a similar procedure to the PVI normalization. This has been subsequently suggested in the form of variation coefficient ΔV (varcoV) and ΔC (varcoC) variables (see Dellwo, 2006 and White & Mattys, 2007). There is, however, a more fundamental difference between the two measures. The intervallic measures used by Ramus et al. are purely phonetic. They are based on the distinction between vowels and consonants,

which is specific to speech. The nPVI on the other hand, is non-specific. Although it was originally developed as a measure of vowel durational variability, it can equally apply to any sequence of temporal events, including musical sounds. This led Patel & Daniele (2003a) and subsequent studies to use the nPVI as a comparable measure for speech and musical rhythm.

2.4. The syllable as a rhythmic unit

The previous sections discussed how the rhythmic patterns of speech correlate with phonological and acoustic parameters. These parameters, however, should not be confused with the rhythmic patterns themselves. Wagner & Dellwo (2004) showed that the %V/ Δ C measure can predict the rhythmic classification of languages as stress-timed and syllable-timed simply on the basis of phonetic transcription, rather than on the basis of durational measurements. Wagner & Dellwo transcribed short texts in English, German, Italian and French in broad phonetic transcription, and calculated %V/ Δ C based on their labeling as either “V” or “C”. Crucially, geminate consonants, diphthongs and tense vowels were considered single segments, and reduced vowels were not distinguished from full vowels. Figure 7 shows that a similar distribution of stress-timed vs. syllable-timed languages on the %V/ Δ C space was found based on this abstract, non-acoustic representation:

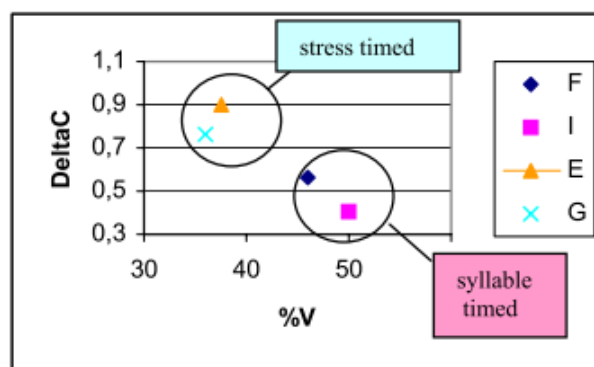


Figure 7: %V vs. Δ C values for stress-timed (English, German) and syllable-timed (French, Italian) languages based on phonetic transcription of texts in these languages. From Wagner & Dellwo (2004).

Wagner & Dellwo conclude that while the %V/ Δ C provides good indication for rhythmic classification, it is essentially “an alternative representation of syllable complexity and variety”

and “has nothing to say about rhythmic properties of the respective languages in the time domain”. A true rhythmic measure, they say, should capture the sequential nature of rhythmic patterns. PVI measures, in comparison, are sequential, but do they reflect the actual rhythmic sequencing of speech? Like the %V/ΔC, PVI measures have been applied to vocalic and intervocalic intervals separately. Accordingly, they correspond to different durational patterns. To illustrate this, consider the following example:

- (5) a. $\overset{150}{\text{'C}} \overset{50}{\text{V C C . C ə C}}$
 $\underline{10} \quad \underline{10} \quad \underline{10} \quad \underline{10} \quad \underline{10}$
- b. $\overset{130}{\text{'C}} \overset{100}{\text{V C . C}} \overset{100}{\text{V . CV}}$
 $\underline{10} \quad \underline{10} \quad \underline{10} \quad \underline{10}$

Sequence (5a) represents a stress-timed language, composed of a stressed CVCC syllable followed by an unstressed CVC syllable with a reduced vowel. In a stress-timed language we expect the stressed full vowel to be considerably longer than the unstressed reduced vowel. I arbitrarily assigned a duration of 150 ms to the full vowel and 50 ms to the reduced vowel. Sequence (5b) represents a syllable-timed language. As such, it does not contain reduced vowels and has smaller durational contrast between stressed and unstressed vowels (130 ms vs. 100 ms, arbitrarily). For simplicity, let us also assume that all consonants are equally 10 ms long. For both sequences, we can calculate PVI values on separate vocalic and consonantal tiers, as illustrated in (6):

- (6) a. $\underline{10} \quad \underline{30} \quad \underline{10}$ intervocalic rPVI = 20
 $\underline{150} \quad \underline{50}$ vocalic nPVI = 100
- b. $\underline{10} \quad \underline{20} \quad \underline{10}$ intervocalic rPVI = 10
 $\underline{130} \quad \underline{100} \quad \underline{100}$ vocalic nPVI = 13.04

For the stress-timed sequence (5a) we get a 20/100 intervocalic rRPVI/vocalic nPVI ratio, compared to 10/13.04 for the syllable-timed (5b). These ratios reflect differences in syllable

complexity and vowel durations, but do they also represent the rhythmic patterns of these sequences? To some extent, they do. The vocalic tiers in (6a,b) resemble the stereotypic Morse-code/machine-gun durational patterns. The problem is that the vocalic and intervocalic tiers in (6) do not combine to a single rhythmic pattern. If anything, they seem to conflict with one another. On the **vocalic tiers**, we get a long-short(-short) pattern, against a short-long-short pattern on the **intervocalic tier**. The underlying problem seems to be the complexity of speech rhythm, which cannot be reduced to the segmental level. Without some higher-order organizing principle, vowels and consonants cannot combine to a single rhythmic pattern. It seems that rhythm must involve some level of abstraction beyond the phonetic-acoustic surface.

Patel (2008) defines rhythm as the systematic patterning of sound in terms of timing, accent, and grouping. According to Jun (2014), “rhythm is the temporal organization of speech perceived by a regular occurrence of events, whether the event is auditory or visual and whether the acoustic medium is timing, fundamental frequency (F0), or amplitude”. Jun distinguishes two levels of rhythm. Micro-rhythm is typically formed by prosodic units below the word level, for our purposes – syllables and feet. Macro-rhythm is a tonal rhythm, perceived by changes of pitch contour beyond the word level:³

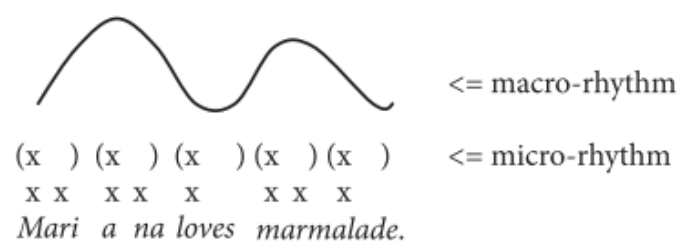


Figure 8: Macro- and micro-rhythm levels in speech utterances. From Jun (2014).

This makes the syllable the minimal rhythmic unit of speech. Jackendoff (2009) similarly notes that “in phonology, the minimal metrical unit is the syllable”.

³ In this paper I use the term *micro-* and *macro-rhythm* in a more general sense than Jun (2014). See section 3.1.4, footnote 6.

The Morse-code/machine-gun distinction originally concerned the durational patterns of syllables, not segmental intervals. Yet, Patel (2008) points out that “surprisingly, there has been little empirical work comparing sentence-level variability in syllable duration among different languages”. Recall that the vocalic/intervocalic dichotomy was introduced as a hypothesis on early language acquisition. In their study, Ramus et al. (1999) propose this distinction under the assumption that “the infant primarily perceives speech as a succession of vowels of variable durations and intensities, alternating with periods of unanalyzed noise (i.e. consonants)”. However, this distinction does not necessarily reflect the rhythmic perception of adult speakers, which may be guided by more abstract phonological principles in addition to purely phonetic cues. Based on previous evidence, Low et al. (2000) suspected that “the basis for attributing stress- or syllable- timing may involve vowels rather than syllables”. While they were able to show that **vocalic** PVI measurements distinguish SE from BE as relatively syllable-timed, they did not show whether this can be similarly supported by **syllabic** PVI measurements. The reason for this may be practical more than theoretical. As Low et al. point out, “syllable boundaries are notoriously difficult to determine in English, and one cannot necessarily assume that consistent syllable duration measurements can be taken”. Patel believes, though, that such difficulties “should not stop research into syllabic duration patterns, because these patterns are likely to be perceptually relevant”.

In (7), I calculated the **syllabic** nPVI values of sequences (5a,b), grouping vocalic and consonantal durations together, based on their underlying syllabic structure. Unlike the intervallic approach in (6), this method provides a single numeric value for each sequence, representing a single underlying rhythmic pattern:

$$\begin{array}{l}
 (7) \text{ a. } \quad \frac{180}{\text{' C V C C}} \quad \frac{70}{\text{C ə C}} \quad \text{syllabic nPVI} = 88 \\
 \\
 \text{b. } \quad \frac{150}{\text{' C V C}} \quad \frac{110}{\text{C V}} \quad \frac{110}{\text{C V}} \quad \text{syllabic nPVI} = 15.38
 \end{array}$$

Because syllabification is not acoustically grounded, criteria for marking syllable boundaries are less rigid than for segment boundaries. They vary under different phonological theories and are inevitably influenced by the researcher's subjective judgement. For instance, should intervocalic consonants be analyzed as the coda of the preceding vowel, onset of the following vowel or as ambisyllabic consonants (see Nesbitt, 2018)? This issue is irrelevant under a segmental analysis. On the other hand, the status of glides in pre-vocalic and post-vocalic positions (*queen* vs. *how*) is more problematic for a segmental analysis than for a syllabic analysis (see Ramus et al., 1999). Eventually, each method has its advantages and drawbacks. While practical considerations definitely play a role, I believe that the choice should ultimately be made on a theoretical basis. From a theoretical standpoint, I argued that syllable durations are more faithful to the underlying rhythmic patterns of speech than vocalic/intervocalic intervals. In addition, for the purposes of this study, syllables seem more comparable with other, non-linguistic rhythmic units, in this case musical tones. The vocalic/intervocalic distinction has been used quite effectively for the rhythmic classification of different languages. It is questionable, however, that it is a relevant distinction when comparing linguistic rhythm with other rhythmic domains such as musical rhythm. Comparative studies of speech and musical rhythm typically ignore intervocalic durations and focus on vowel durations as the correlates of musical tones. But this choice is again more practical than theoretical. Patel & Daniele (2003a) note that vocalic nPVI measurements have been used in comparison with durations of musical tones because "vowels form the core of syllables, which can in turn be compared to musical tones". This seems evident even by the simple fact that the notes of a melody are vocalized as syllables rather than vowels. That is, we tend to sing in *la-la-la* or *na-na-na*, not in *a-a-a*. While vowel durations provide an approximation of syllable durations, a comparison of musical tones to syllables seems more direct and thus preferable.

2.5. Chapter summary

This chapter discussed main concepts in the theory of speech rhythm, which include rhythmic classification and its phonological and acoustic correlates. I argued here that while the vocalic/intervocalic distinction proved useful for the study of these concepts, the syllable should nevertheless be considered as the basic rhythmic unit of speech, especially when compared to musical tones and other non-linguistic rhythmic units. For this reason, I chose to base my study, described in chapter 4, on a comparison of syllable and note durations. Before that, previous comparative work on speech and musical rhythm is discussed in chapter 3.

3. The comparative study of speech and musical rhythm

This chapter discusses two approaches within the comparative study of speech and musical rhythm. These approaches are based on two different uses of the nPVI as a comparable rhythmic measure between these domains. One approach, originally proposed by Patel & Daniele (2003a, henceforth P&D) measures nPVI values based on the orthographic representation of musical rhythm in music notation. The second approach, proposed by McGowan & Levitt (2011) and Carpenter & Levitt (2016), obtains nPVI measurements from acoustic durations of notes in music performance. I suggest that the first approach uses the nPVI as a measure of **metrical variability** in the underlying rhythmic structure of musical phrases. This is discussed and explained in section 3.1. The second approach uses the nPVI as a measure of acoustic **durational variability**, similar to how it has been used in speech analysis. This approach is reviewed in section 3.2. In section 3.3, I discuss how these two approaches relate to two different rhythmic dimensions: (i) underlying metrical structure, and (ii) timing nuances of rhythmic performance. The chapter ends with a short summary in section 3.4.

3.1. The nPVI as a measure of metrical variability

Following P&D, a series of studies applied the nPVI to the analysis of musical rhythm as it is represented in music notation. This form of analysis emphasizes structural and metrical aspects of musical rhythm in comparison to speech.

3.1.1. Patel & Daniele (2003a)

P&D's innovation was in providing an empirical framework for comparing speech and musical rhythm. This framework is based on the idea that the nPVI can be used as a comparable rhythmic measure for comparable linguistic and musical data. As their linguistic data, P&D used Ramus' (2002) vocalic nPVI measurements of English and French. P&D chose English and French as typical stress-timed and syllable-timed languages, with significantly different

degrees of durational variability. They argued that a similar pattern of variability in music by English and French speaking composers can empirically support the claim that “the prosody of a composer’s native language can influence the structure of his or her instrumental music”. P&D’s focus on instrumental music was guided by the intuition that instrumental music can reveal deeper connections to language than vocal music, in which they see an “obvious route” between the domains. As a comparable musical corpus, P&D collected excerpts of musical themes by British and French composers from Barlow & Morgenstern's (1983) “A Dictionary of Musical Themes”, a reference book listing important themes in the classical music literature. Figure 9 illustrates P&D’s method of nPVI calculation for musical themes:

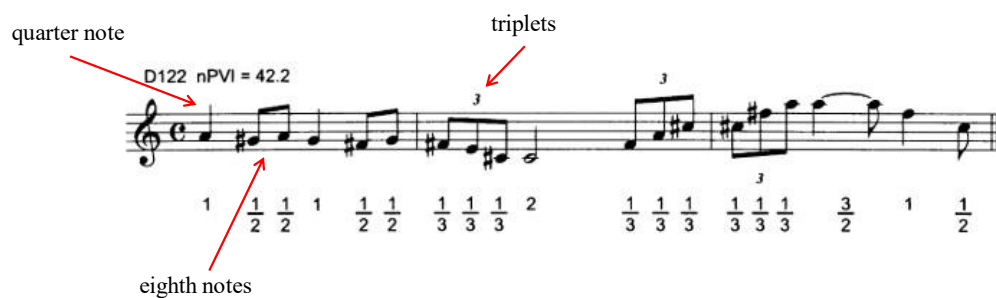


Figure 9: P&D’s nPVI calculation based on notated rhythm. Note values are calculated in proportion to the first note of the phrase, arbitrarily given a value of 1.

Whereas in speech nPVI values are obtained from acoustic measurements in production, P&D calculate musical nPVI based on the rhythmic values of notes in music notation. In music notation, rhythmic values do not represent absolute durations of sounds measurable in units of time. Instead, rhythmic values such as *quarter notes* and *eighth notes* represent proportional relations of beats on an abstract metrical grid. A more accurate term for durational values in P&D’s method would therefore be *metrical values*. Under P&D’s method, each note is assigned a numeric value proportional to the metrical value of the first note in the phrase. For example, in figure 9 the first note is a quarter note, arbitrarily assigned a value of 1. The following two notes in the phrase are eighth notes and therefore receive a value of 1/2. The triplets in the second bar, which are a ternary division of the quarter note, receive a value of 1/3, and so forth.

On average, P&D found a greater nPVI value of 46.9 for English musical themes, compared to 40.9 for French themes. When compared to Ramus' (2002) English and French speech nPVI data, a similar pattern emerges:

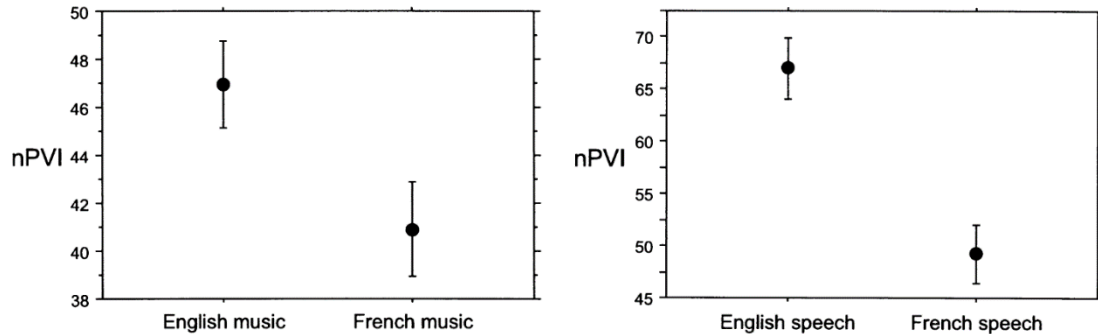


Figure 10: P&D's English and French musical nPVI values compared with Ramus' (2002) speech nPVI values for English and French.

Although the difference in musical nPVI values is notably smaller, P&D found this difference to be statistically significant. They consider this as “an empirical basis for the claim that spoken prosody leaves an imprint on the music of a culture”.

In a replication study, Huron & Ollen (2003) applied P&D's method to the entire Barlow & Morgenstern dictionary. Using music analysis software developed by Huron (1994), Huron & Ollen analyzed a set of 7748 instrumental musical themes by composers of various nationalities, including 737 themes by English composers, (average birth year = 1782, range = 1540-1913) and 1188 themes by French composers (average birth year = 1814, range = 1601-1912). Figure 11 shows Huron & Ollen's findings for composers of various nationalities, including English and French:

**Mean Agogic Contrast for Musical Themes According to
Composer Nationality**

| Nationality | Normalized Pairwise Variability Index (nPVI) | Standard Error | No. of Musical Themes |
|---------------------------|---|----------------|--------------------------|
| American | 46.3 | 1.12 | 415 |
| Austrian | 42.0 | 0.65 | 1194 |
| Czech | 47.1 | 1.56 | 232 |
| English | 45.6 | 0.90 | 737 |
| French | 43.7 | 0.73 | 1188 |
| German | 42.0 | 0.59 | 2006 |
| Hungarian | 45.4 | 1.62 | 244 |
| Italian | 42.7 | 1.06 | 529 |
| Polish | 45.1 | 1.74 | 254 |
| Russian | 39.8 | 0.82 | 736 |
| Scandinavian ^a | 45.9 | 1.79 | 141 |
| Spanish | 42.5 | 1.85 | 108 |

^aDanish, Swedish, Norwegian (Finnish excluded).

Figure 11: Average nPVI values for musical themes by composers of various nationalities. From Huron & Ollen (2003).

Like P&D, Huron & Ollen found a significantly higher nPVI value for English themes than for French themes, a finding they consider to be “consistent with the view that spoken prosody leaves an imprint on the music of a culture, at least in the case of English and French”. Although significant, the difference between English and French values found by Huron & Ollen is even smaller than in P&D's study (English nPVI = 45.6, French nPVI = 43.7).

3.1.2. Interfering factors

A comparison of notated classical music with contemporary speech performance involves a diachronic dimension. In the case of Huron & Ollen, musical data ranges over almost four centuries of music. Aware of the fact that “languages are known to change over historical time in terms of sound structure”, and “since measurements of speech prosody are based on contemporary speech”,⁴ P&D restricted their corpus to composers “from a relatively recent musical era”. Still, the average birth year of these composers is 1871, leaving about a century of **phonological change** between their linguistic and musical data. To avoid possible effects of phonological changes, contemporary speech can be compared synchronically with

⁴ Ramus' (2002) nPVI data are based on recordings from Nazzi et al's (1998) corpus.

contemporary musical data. This, in turn, could involve another interfering factor, which I call here **intercultural influences**. Contemporary musicians of the globalized era absorb various influences from different cultures, which may weaken the dominance of culture-specific characteristics on their music, including possible influences of their native language. In their study, P&D focus on composers who spanned the turn of the 20th century, during the era of *musical nationalism*, in which more distinct national styles evolved and “speech prosody has been thought to play a role”. Patel & Daniele (2003b) show that nPVI averages of German speaking composers (of German and Austrian nationalities) in Huron & Ollen’s corpus tend to increase over time, for example:

| Composer | Birth Year | nPVI Average |
|-----------------|-------------------|---------------------|
| J.S. Bach | 1685 | 20.7 |
| W.A. Mozart | 1756 | 31.9 |
| L.V. Beethoven | 1770 | 36 |
| R. Strauss | 1864 | 51.3 |

Figure 12: A diachronic nPVI comparison of four German speaking composers, representing different periods in classical music history - J.S. Bach (Baroque), W.A. Mozart (Classical), L.V. Beethoven (early Romantic), R. Strauss (late Romantic). Based on Patel & Daniele (2003b).

They speculate that lower values for pre-national composers (e.g. J.S. Bach, W.A. Mozart) could reflect strong Italian influences over German compositional style. Some composers who follow the rise of European nationalism (e.g., L.V. Beethoven, R. Strauss) have much higher nPVI averages, as if their style has become more “stress-timed”. While this is only a speculation, Patel & Daniele’s “Italian hypothesis” illustrates the possible effects of intercultural influences on musical style. By controlling for this factor, stronger correlations between language and music could possibly emerge.

3.1.3. Folk vs. classical music

At least intuitively, folk musical styles seem to emphasize culture-specific traits more strongly than classical music, and therefore could provide more conclusive evidence on language and music connections. Preliminary findings by Jekiel (2014) seem to support this intuition. Jekiel applied P&D's method to a small corpus of English and Polish speech utterances and musical themes. Jekiel's musical data included both folk songs and classical music themes by 19th national composers of both nationalities. Similar to Ramus (2002), Jekiel found significantly higher nPVI values for English utterances compared to Polish utterances. A similar pattern was found for English and Polish folk music, with a higher nPVI average for English folk songs. Yet for classical music, an opposite tendency was found with Polish averaging higher than English, similar to Huron & Ollen's (2003) findings for English and Polish classical music (see Patel, 2008 Appendix 2). Jekiel's findings remain inconclusive, however, due to his limited set of data.

A more detailed study of folk music by Nguyễn (2017) does not agree with Jekiel's findings. In her study, Nguyễn compared speech utterances and folk songs in English and Vietnamese. Vietnamese is a contour tone language lacking culminative word stress and was claimed to pattern with syllable-timed languages. According to Nguyễn, Vietnamese and English "represent two broadly contrastive prosodic types". Nguyễn's speech data were based on 20 English sentences from Nazzi et al. (1998), re-recorded by four Australian English (AuE) speakers, and 20 Vietnamese sentences recorded by four Southern Vietnamese speakers. Her musical data were based on 162 English folk songs (Australian, American and British) and 60 folk songs from three regions of Vietnam (Southern, Central and Northern). Following P&D, Nguyễn calculated nPVI musical values based on music notation. Unlike previous studies, she used a specially designed software to extract 7 different types of rhythmic measures from her

speech data,⁵ including vocalic nPVI and rPVI and the normalized and unnormalized standard deviation measures described in section 2.3. While nPVI values of AuE and Southern Vietnamese did not reveal a significant difference (51.73 vs. 50.07 respectively), an overall comparison of these various measurements showed that these languages are rhythmically distinct. Especially, a comparison of ΔC and $V\%$ values showed significantly greater durational variability for AuE, with higher ΔC and lower $\%V$ averages (more variation in consonantal intervals, smaller proportion of vocalic intervals). As opposed to that, Nguyễn found no significant distinction between English and Vietnamese folk music. If anything, Vietnamese folk songs had a slightly higher nPVI average than English folk songs. In addition, no significant regional/dialectal distinctions were found. Nguyễn found no significant differences in nPVI values between Australian, British and American folk songs and only a weak difference between Central Vietnamese folk songs and folk songs from Northern and Southern Vietnam. Despite these results, I believe that the basic intuition concerning speech and folk music should not be abandoned. Instead, I suspect that P&D's methodology proves less effective in the case of folk music and non-literate musical styles in general. By basing their analysis on notated music, P&D's method emphasizes the metrical aspects of musical rhythm and ignores rhythmic nuances of music performance, which could be more relevant for comparison with spoken language.

3.1.4. Metrical structure vs. rhythmic feel

By “rhythmic feel”, musicians refer to those aspects of musical rhythm which cannot be captured by the written score, namely non-metrical aspects of musical rhythm. A common type of rhythmic feel in contemporary music is known as “swing feel”. Swing feel is the central rhythmic characteristic of jazz music, and is characterized by noticeable alternation of long and short durations. Interestingly, although this alternation is clearly noticeable by ear, swing feel

⁵ "Correlatore", https://www.lfsag.unito.it/correlatore/index_en.html.

is most commonly notated by musicians in the form of *consecutive eighth notes*, as in the following phrase taken from the “Charlie Parker Omnibook” (Parker, 1978), a popular jazz practice book:

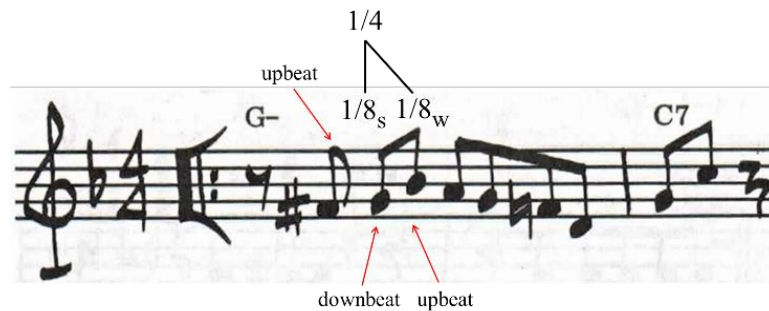


Figure 13: A notated transcription of a jazz phrase from saxophonist Charlie Parker’s composition “Scrapple from the Apple”, based on his own recording of this composition. Taken from the “Charlie Parker Omnibook for C Instruments” (Parker, 1978). The phrase is notated as consecutive eighth notes, beginning on an upbeat. Metrical $nPVI = 0$.

In consecutive eighth note rhythms such as in figure 13, eighth note pairs are analogous to binary prosodic feet, forming pairs of stressed and unstressed syllables. In music, the metrically prominent or strong eighth note is called a *downbeat* and the metrically weak eighth note is called an *upbeat*. Under this analogy, the prosodic foot level is thus equivalent to the metrical level of the quarter note. Because all notes in this phrase are notated with equal metrical values, by P&D’s method this phrase receives an **nPVI value of zero**. This nPVI value differs significantly from the nPVI we get by measuring the actual durations of notes as performed by Charlie Parker himself in the original recording of this phrase, as illustrated in figure 14 below:

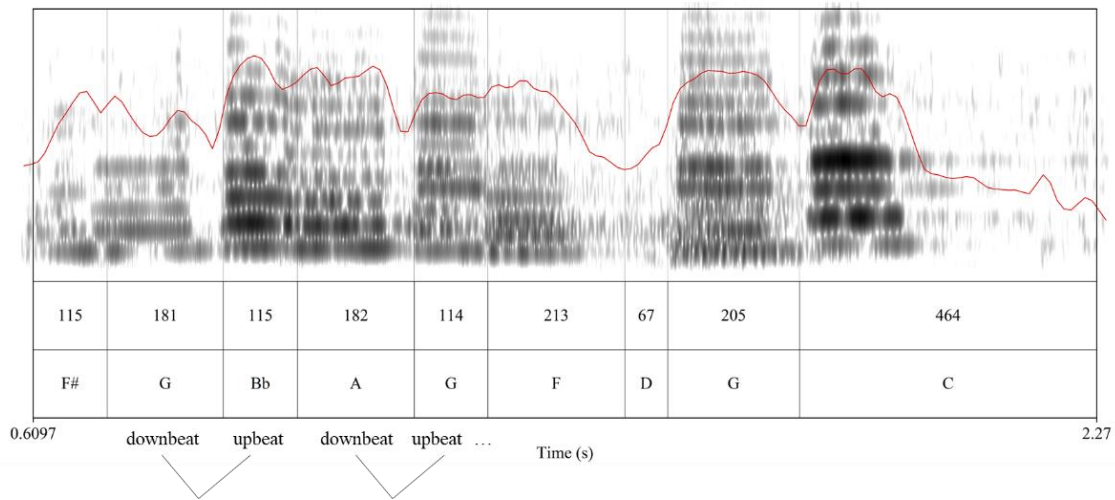
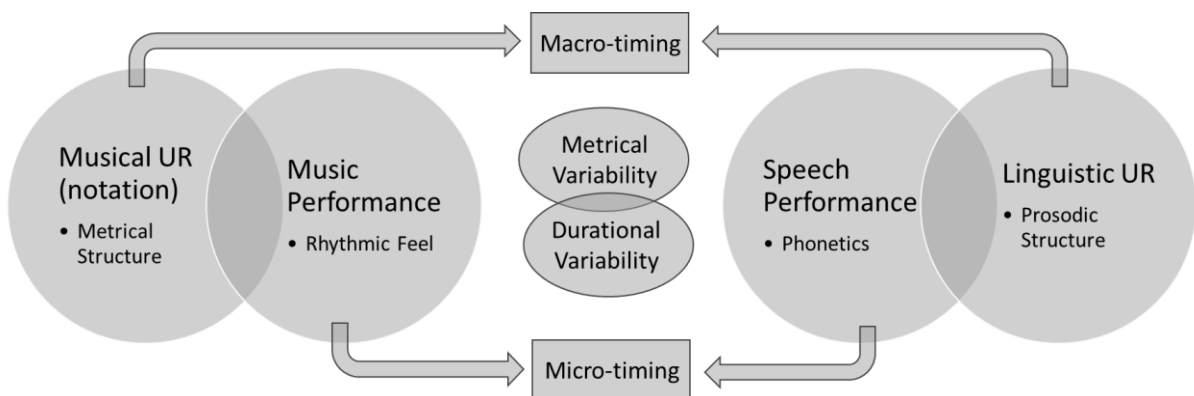


Figure 14: An acoustic analysis of the “Scrapple from the Apple” phrase in figure 13, from the original recording by Charlie Parker (“Jazz Masters”, EMI Jazz, 1999). Intensity contour marked in red, note durations marked in milliseconds. Acoustic nPVI = 65.4.

An acoustic analysis of Parker’s recorded performance of the phrase in figure 13, shows that on surface downbeats are mostly played longer than upbeats, resulting in a “stress-timed” long-short durational pattern. To a great extent, this is what makes this phrase swing. This variability on surface is reflected by a relatively high nPVI value of 65.4, similar to what we see in typical stress-timed languages, and in clear contrast to the zero nPVI value we get from the notated form of the phrase. This suggests that we should distinguish between two different uses of the nPVI, measuring two different dimensions of musical rhythm. I believe that a similar distinction can be made in speech. This is sketched in the diagram below:

(8) Durational vs. metrical variability:



The durational patterns of speech and music can be quantified in terms of their **durational variability**, measured by **acoustic nPVI** values. In music, these patterns reflect rhythmic nuances of performance (“rhythmic feel”) which may or may not overlap with metrical patterns of underlying musical structure. Datsieris et al. (2019) refer to such durational nuances as *micro-timing deviations*, which can be understood as the actual durations of notes in performance, compared to their ideal alignment on the abstract metrical grid. I believe that the concept of **micro-timing** or **micro-rhythm** can be adopted for similar processes in the phonetic-acoustic realization of phonological prosodic structure. Here too, the durational patterns of speech do not fully overlap with their corresponding underlying phonological representation. On a different rhythmic dimension, formal representations of higher-order metrical structures in both domains can be independently compared. Here I generalized these structures under the term **macro-timing** (or **macro-rhythm**).⁶ Lerdahl & Jackendoff (1983) laid-out a theoretical foundation for such a formal comparison between the domains, which was later revised by Katz & Pesetsky (2011). For our purpose, we should notice that the framework proposed by P&D provides an indirect comparison between **metrical variability** on the musical domain and **durational variability** on the linguistic domain. The former is a property of musical underlying representation measured in metrical nPVI values, and the latter is a property of linguistic performance measured in acoustic nPVI values. An alternative framework is proposed by more recent studies, which directly compare durational variability in speech and musical performance.

3.2. Durational variability in speech and musical performance

As P&D point out, the difficulty in comparing speech performance with musical performance is the selection of representative musical recordings for analysis. In classical music, traditions

⁶ The terms *micro-rhythm* and *macro-rhythm* have been used in a slightly different manner in phonological literature (see section 2.4.). To avoid confusion, I chose here to use terms *micro-timing* and *macro-timing* instead.

of performance have changed throughout the years and do not necessarily reflect the composer's original interpretation of the music. In this respect, the notated score is more objective than some specific recording of a piece. To overcome this difficulty, McGowan & Levitt (2011) and Carpenter & Levitt (2016) studied spoken utterances and musical phrases produced by musicians who are authentic performers of their style (e.g., Louis Armstrong and Herbie Hancock in jazz). This provides a synchronic framework for a direct comparison of speech and musical performance, which better controls for possible effects of phonological change and intercultural influences (as discussed in section 3.1.2 above).

3.2.1. McGowan & Levitt (2011)

McGowan & Levitt (2011) analyzed speech and musical data from three regions where different English dialects are spoken: (i) the Shetland Islands, Scotland (*Shetland*), (ii) County Donegal, Ireland (*Donegal*, a variant of *Ulster English*), (iii) Kentucky, USA (*Southern American English*). Scottish dialects, as well as Ulster English, are characterized by more uniformity in vowel durations compared to other English dialects, as a result of the Scottish Vowel Length Rule (lengthening or avoiding the reduction of vowels in certain phonological contexts). Accordingly, these dialects were found to have relatively low vocalic nPVI values. Pitch-peak delay in Ulster English (and other Scottish dialects) is supposed to further reduce contrast between stressed and unstressed syllables. As opposed to that, in the Shetland dialect, influence of Scandinavian syllable structure (in which long vowels are followed by short consonants and vice versa) contributes to slightly higher durational variability. Accordingly, Shetland was found to have significantly higher nPVI values than Ulster English. Compared to these dialects, Kentucky English incorporates Southern American characteristics such as lengthening and diphthongization of stressed vowels and on overall has a higher degree of durational variability.

Interestingly, these three regions also have three variants of a similar music style, the *reel*. The reel is a folk dance which is believed to have originated in Scotland and spread to other regions such as Ireland and the American Appalachia. Similar to jazz, reel playing too is characterized by some degree of long-short alternation in eighth note playing. It was observed that in each of these regions, reel eighth notes are played in slightly different degree of unevenness or swing (see references in McGowan & Levitt, 2011). Musicians in the southern Appalachian region (including Kentucky) are said to use the “dotted rhythm” with an eighth note ratio of roughly 3:1 between the downbeat and upbeat.⁷ In the Shetland variant of the reel, a “less ‘extreme’ dotted rhythm” was observed, with ratios of 4:3, 5:3 and sometime 2:1. In comparison, the Irish “rhythmic feel” is considered to be more even, and specifically in Donegal it was claim that “the uneven rhythms characteristic of Scottish music were leveled out”. Intuitively, this seems like a similar pattern to the differences in speech durational variability found in the English dialects of these regions: Kentucky > Shetland > Donegal. To test this empirically, McGowan & Levitt analyzed field recordings of fiddle players from these regions, including performances of reel music and spoken utterances by the same musicians. These recordings were made around the 1950’s before what is known as “the second revival” of Irish music and the spreading of Celtic musical influences during the 1960’s and 1970’s, which “undoubtedly has contributed to the undermining of regional identities”. Hence, McGowan & Levitt’s recordings represent a more traditional fiddle playing style, specific to each region, compared to later musicians in the same regions who had already absorb more intercultural influences.

Unlike previous research, speech data in McGowan & Levitt’s study is based on spontaneous speech utterances rather than conscious narration of constructed sentences. To

⁷ The dotted rhythm is notated by a dotted eighth note followed by a sixteenth note, representing a metrical value of 3/16 on the downbeat vs. 1/16 on the upbeat.

select suitable samples for analysis, utterances with fewer than four syllables, obvious pauses or hesitations and non-declarative intonation were excluded, resulting in a corpus of 20-30 utterances per speaker by three musicians in each region. In addition, 20 two-bar phrases from reel tunes played by each of these musicians were extracted from the recordings. Following Patel et al. (2006), McGowan & Levitt segmented speech samples to individual vowels rather than vocalic intervals. Note boundaries in the musical samples were marked between the onset of one note to the onset of the following note. Differences in speech nPVI values were found significant between Kentucky and Donegal English, and approaching significance between Kentucky and Shetland. Interestingly, differences in musical nPVI were found highly significant. On the overall, a similar pattern emerged for speech and music as can be seen in figure 15 below:

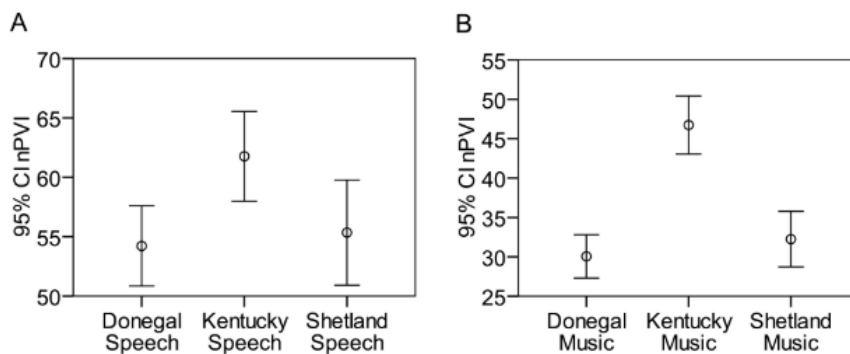


Figure 15: A comparison of acoustic nPVI averages for three English dialects – Donegal, Kentucky and Shetland, and their corresponding variants of reel music. From McGowan & Levitt (2011).

McGowan & Levitt conclude that this similar pattern supports their hypothesis that “speech and music would share rhythmic characteristics”.

Recall, that an opposite tendency was found in P&D’s and Huron & Ollen’s (2003) studies, where differences between musical values were considerably smaller than between speech values. When looking at a notated excerpt from McGowan & Levitt’s corpus we see that the

phrases are characterized by a relatively **uniform metrical structure**, composed mostly of consecutive eighth notes and some quarter notes:

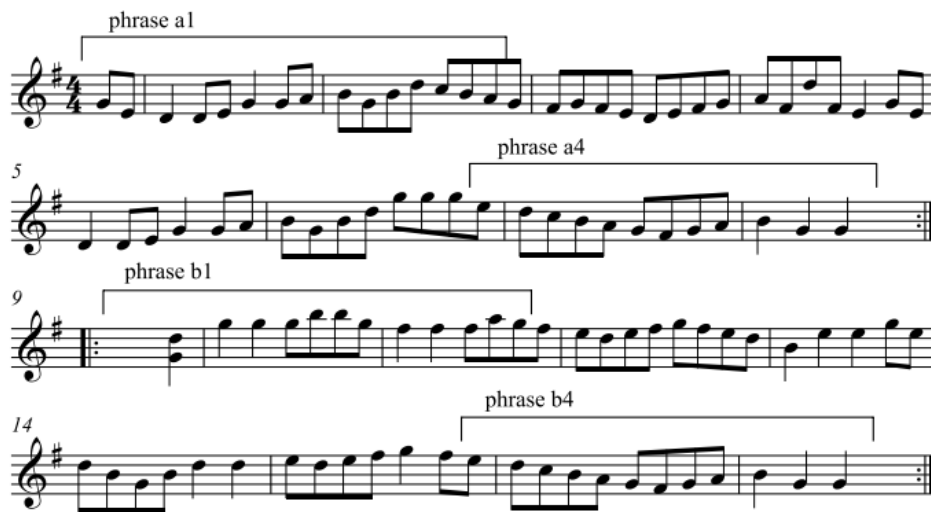


Figure 16: A notated example of a Shetland reel tune analyzed by McGowan & Levitt (2011). Brackets mark the parts of the tune for which acoustic nPVI values were calculated by McGowan & Levitt.

McGowan & Levitt note that this notated form “is a ‘skeleton’ of the tune that shows its structure but does not show the rhythmic nuances of the playing” and speculate that “the repetitive nature of the musical form of the reel serves to emphasize rhythmic patterns as compared to spontaneous speech”. I suspect that by “repetitive”, McGowan & Levitt refer to the relatively uniform metrical structure of consecutive eighth notes which emphasizes micro-timing nuances in performance, and these could be more relevant in a comparison with similar patterns in speech.

3.2.2. Carpenter & Levitt (2016)

Carpenter & Levitt (2016) adopted McGowan & Levitt’s methodology to study the music and speech of jazz musicians speaking American English and riddim musicians speaking Jamaican English. Jazz and riddim are distinct musical styles related to these dialects. Riddim is the musical backbone of Jamaican dancehall tunes, a contemporary style of electronic pop music. Music producers create riddim tracks as an instrumental accompaniment over which vocalists

and DJ's can create different versions of dancehall tunes, often with rap-like singing in Jamaican English. A successful riddim can be used for tens if not hundreds of different dancehall tunes. The riddim itself is mostly based on electronic beats and synthesized instruments, and includes short melodic “hooks” or motifs. In some cases, vocalists sing in sync with the rhythm of these melodies. Jazz music, which has already been discussed above, developed around the first half of the 20th century by African-American musicians. According to Carpenter & Levitt, it has been suggested that instrumental jazz is influenced by the speech patterns of its musicians. Thomas & Carter (2006) found a relatively low nPVI median for Jamaican English, reflecting “the fact that Caribbean Anglophone speech is often described as being more syllable-timed than other varieties of English”. In comparison, North Carolina English speakers of both African and European descent were found by Thomas & Carter to be “quite stress-timed overall, with no significant difference between them”. Carpenter & Levitt therefore assumed that there are “good reasons for predicting that both the rhythmic patterns of instrumental jazz and riddim music would reflect the different speech rhythms of their producers”.

To construct their study corpus, Carpenter & Levitt searched for recorded interviews of prominent jazz and riddim musicians. They then chose several riddims and jazz improvisations, performed / produced by each of these musicians. Like McGowan & Levitt, Carpenter & Levitt calculated speech and musical nPVI values based on measurements of vowel and note durations. In both speech and musical samples, they found significantly higher nPVI for American jazz musicians than for Jamaican riddim musicians. Like McGowan & Levitt (and contrary to P&D), Carpenter & Levitt found a smaller difference between speech values than between musical mean values:

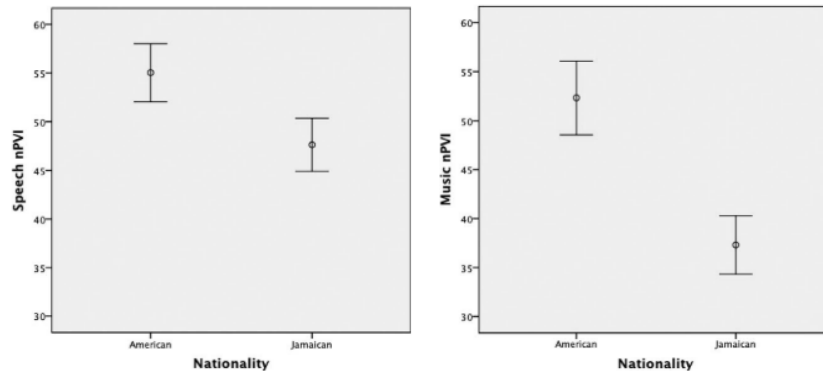


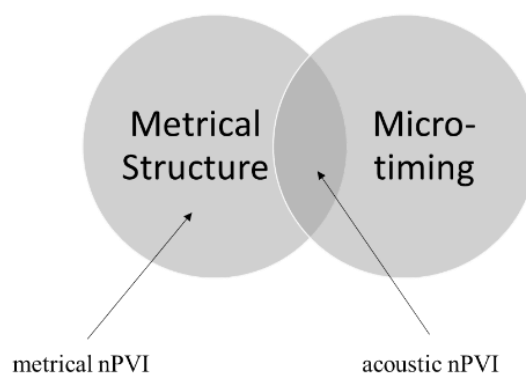
Figure 17: A comparison of acoustic nPVI averages for speech utterances and musical phrases produced by American jazz musicians and Jamaican riddim musicians. From Carpenter & Levitt (2016).

Carpenter & Levitt conclude that the use of spontaneous speech and recorded music “rather than read speech and music notation ... supports the more extensive findings in the literature showing that the linguistic rhythms of languages can be echoed in the music rhythms of their instrumental music”.

3.3. Durational variability vs. metrical uniformity

The **acoustic** nPVI value of musical phrases should be understood as the combined result of rhythmic variability in the underlying metrical level (“macro-timing”) and in real-time performance (“micro-timing”). That is because note durations are determined both by their metrical representation and the manner in which this representation is realized in performance:

(9) Acoustic vs. metrical nPVI:



It therefore follows that we can focus on the durational nuances of musical performance by reducing the effect of metrical variability in musical phrases. To illustrate this, consider again the Charlie Parker phrase from figure 13, rewritten below:

(10)

upbeat downbeat upbeat downbeat upbeat...

Metrical tier: 1/8 1/8 1/8 1/8 1/8 1/8 1/8 1/8 1/8 *metrical nPVI = 0*

Durational tier: 115 [181 115] [182 114] [213 67] [205 464] *acoustic nPVI = 65.4*

BUR: 1.57 1.59 3.17 0.44

By convention, this phrase is notated as a sequence of consecutive eighth notes, resulting in a **metrical** nPVI value of **zero**. This value reflects a metrically uniform structure, with no rhythmic variability in the underlying metrical structure of the phrase. However, as we have seen in figure 14, in the actual recording of the phrase by Charlie Parker, note durations alternate, with longer durations mostly on downbeats and shorter durations on upbeats. This alternation results in an **acoustic** nPVI value of **65.4**. We can therefore distinguish between rhythmic **uniformity** on the **metrical tier** of this phrase and rhythmic **variability** on its **durational tier**. Under this analysis, the acoustic nPVI value of this phrase only reflects variability on surface, capturing the rhythmic nuances of performance, independently of higher-order rhythmic structure.

Alternatively, one may argue that music notation in this case is simply irrelevant. Traditionally, swing eighth notes in jazz are described as a 2:1 ratio between the downbeat and upbeat, derived from a ternary division of the beat (triplets). Under this view, swing eighth notes are underlyingly unequal and durational variability on surface is in fact derived from underlying metrical variability. This suggests a simple model under which three types of rhythmic feel correspond to three underlying metrical categories:



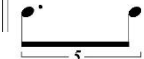


(11) a. even (“straight”) eighths = 1:1 downbeat-upbeat ratio

b. swing eighths = 2: 1 downbeat-upbeat ratio

c. heavy (“dotted”) swing eighths = 3:1 downbeat-upbeat ratio

However, this classification has not yet found any empirical support. Friberg & Sundström (2002) measured eighth note ratios of cymbal hits in recordings of four leading jazz drummers. They found no preference for the so-called classic swing ratio of 2:1. On the overall they found a “linear decrease in swing ratio with increasing tempo”, with swing ratios ranging from 3.5:1 on slower tempos (that is larger than the 3:1 ratio of the “dotted rhythm”) to almost 1:1 on faster tempo (even eighths). For descriptive purposes, Benadon (2006) divides the quarter note continuum to five categories based their on (down)beat to upbeat ratio, or BUR:

(12)

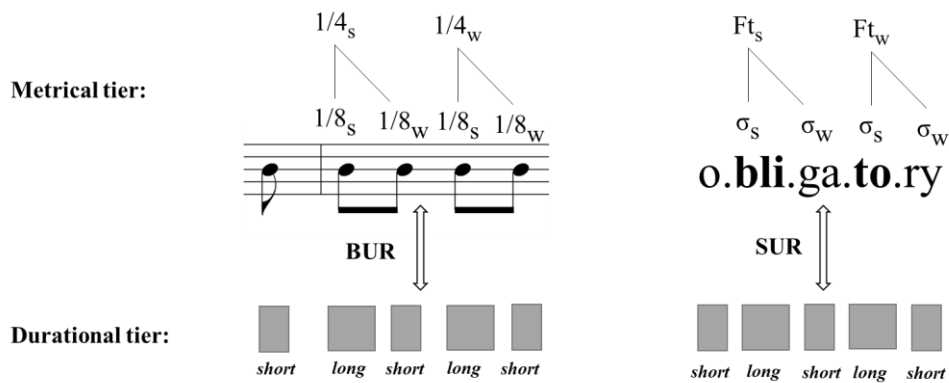
| | | | | |
|---|---|---|--|---|
| 1.0 (1:1) | 1.3 (4:3) | 1.5 (3:2) | 2.0 (2:1) | 3.0 (3:1) |
|  |  |  |  |  |

BUR values are calculated simply by dividing the duration of the downbeat by the following upbeat, as illustrated in example (10) above (unpaired notes, such as the first upbeat in the phrase are excluded). When observing the resulting ratios in 10, we see that none of these ratios approximates the so-called classic 2:1 swing ratio. BUR’s typically range from around a value of 1, which corresponds to (11a) (downbeats and upbeats are even), to a value of 3 which corresponds to (11c) (“dotted rhythm”, very uneven). In (10), the first two BUR’s are around 1.5, corresponding to a ratio of 3:2, the third BUR is larger than 3, that is larger than the “dotted” 3:1 ratio, and the last BUR is well below 1, indicating on an “opposite” swing ratio in which the downbeat is **shorter** than the **upbeat**. This is related to what I refer to here as *final lengthening*, a phenomenon which I came across in both speech and musical samples and is further discussed in chapter 4. With lack of other evidence, I will assume here that swing eighth

notes are **unspecified** for some durational ratio on surface, and form a **uniform metrical structure** with an nPVI value of zero, as represented by conventional notation.

BUR's are an additional tool for measuring durational variability in music. Unlike the nPVI which can only apply to a single rhythmic tier – either the metrical tier or the durational tier, the BUR takes both tiers into account. BUR's connect **long-short** durational relations on surface with **strong-weak** prominence relations in underlying structure (downbeats vs. upbeats). So far, only the nPVI has been used in comparison with speech. Can we use a similar rhythmic measure to the BUR for speech analysis? I believe we can. In music, BUR's are obtained from metrically uniform sequences of consecutive eighth notes. Comparable sequences in speech would be composed of consecutive binary feet with alternating **stressed-unstressed** (or unstressed-stressed) syllables:

(13)



Like downbeat-upbeat ratios, we could then calculate stressed-unstressed ratios. A comparable measure of Stressed-Unstressed Ratio in speech, would take into account both the durational patterns of syllables or vowels in performance, as well as their prosodic prominence on the metrical grid.

3.4. Chapter summary

This chapter reviewed two approaches for comparing speech and musical rhythm. Both approaches found evidence for correlated patterns of rhythmic variability between the domains.

In this chapter I suggested that P&D's approach can be described as an indirect comparison

between durational variability in speech, obtained from acoustic measurements, and metrical variability in musical structure, as represented by music notation. A second approach, proposed by McGowan & Levitt (2011) and Carpenter & Levitt (2016) directly compares durational variability in both domains by analysis of comparable acoustic data produced by individual musicians. To further focus on micro-timing nuances of real-time performance, I suggested that musical data can be restricted to metrically uniform sequences (consecutive eighth note phrases). In this type of sequences, BUR values of adjacent note pairs can be calculated in addition to nPVI values. This is illustrated in a small-scale study in chapter 4, comparing speech and musical data by musicians of two distinct styles – jazz and bluegrass.

4. Jazz and Bluegrass musicians as a case study

Like McGowan & Levitt (2011) and Carpenter & Levitt (2016), this study compares spontaneous speech and musical performance by individual musicians. In this study I chose to compare musicians of two styles which originated around the same period of time by communities speaking different variants of American English – jazz and bluegrass. Within the style of jazz, I compare East Coast African-American and West Coast European-American musicians, belonging to different schools of playing. Unlike previous studies, I use syllable durations rather than vocalic intervals as the comparable unit of musical notes (see discussion in section 2.4). In addition, I constrain the musical corpus to metrically uniform phrases for more focus on the durational nuances of real-time performance.

4.1. Relevant background on jazz and bluegrass

Some properties of jazz music make it especially interesting for comparison with language. Unlike classical music, which is first and foremost a literary tradition, jazz is an aural tradition, learned primarily by emulation of recordings by master musicians. In addition, jazz relies heavily on improvisation of highly sophisticated musical structures as means of interaction between musicians in the ensemble. This gives jazz a discursive dimension with possible similarities to language. Limb & Braun (2008) note that the process of improvisation is common to many aspects of human behavior “including adaptation to changing environments, problem solving and perhaps most importantly, the use of natural language, all of which are unscripted behaviors that capitalize on the generative capacity of the brain”. Donnay et al. (2014) note that like natural linguistic discourse, interactive generative musical performance “involves an exchange of ideas that is unpredictable, collaborative, and emergent”. In an fMRI study, Donnay et al. show that “interactive improvisation between two musicians is characterized by activation of perisylvian language areas linked to processing of syntactic elements in music”. Thirdly, jazz evolved in its formative period as a distinct musical style

within a homogenic ethnic community (African Americans). Fortunately, this period is fairly recent in historical terms, ranging from the 1920's to the 1960's, with some of the pioneer jazz musicians still alive today. This provides us with an abundance of musical recordings by master musicians in this period. Speech recordings by these musicians from interviews and public appearances are also available, although sometimes only from later years. Within this formative period of jazz, I find 1950's jazz recordings to be especially beneficial. First, advancement in recording technology makes these recordings significantly better in quality. Second, this period is characterized by the perfection of the 1940's bebop idiom and its evolution to a highly sophisticated form of expression. This period of the second-generation bebop musicians (e.g. John Coltrane, Sonny Rollins, Art Blakey, Horace Silver) is commonly known in jazz history as the "hard bop" idiom. While later periods in jazz are characterized by even greater complexity and sophistication, they are also characterized by increasing cross-cultural influences and the spreading of jazz worldwide. In comparison, the hard bop era represents more traditional jazz, predominated by African American musicians. Around the same time, another sub-genre of jazz developed from the bebop school, known as "cool jazz". Unlike the bebop and hard bop styles, the cool school was predominated by white musicians. The "cool" style of playing is often described as mellower, less energetic and more classically oriented, compared to a "tougher" playing style in the hard bop school. Owens (1995) describes some early cool jazz improvised solos as "similar to those played by most boppers of the time", except for "almost total lack of syncopation". Because some of its known figures at that time were active in the L.A. jazz scene (e.g. Dave Brubeck, Paul Desmond, Chet Baker), cool jazz was "accidentally construed as a regional style and dubbed 'West Coast Jazz'" (Gridley, 1990). As anti-thesis, the term "East Coast Jazz" has sometimes been used to refer to the hard bop style, whose leading figures centered around NYC. For descriptive purposes only, I will use here the term West Coast jazz to denote music by white jazz musicians from the cool tradition

who were actually born and raised in the West Coast region (California). Similarly, East Coast jazz will denote here music by black musicians of the hard bop school born in the East Coast region.

When comparing East Coast and West Coast jazz musicians, at least two types of classifications can be made. The first is on a basis of ethnic descent. As mentioned in chapter 3, Thomas & Carter (2006) found speech samples by African and European Americans (from North Carolina) to be similarly stress-timed. Interestingly, Thomas & Carter found evidence for “a former difference in rhythm between the two ethnicities” in an analysis of speech recordings by ex-slave African Americans and pre-Civil Southern European Americans. Thomas & Carter propose that African AE could have been less stress-timed in earlier stages, perhaps by creole influence, but even if so, no evidence for such a difference was found in contemporary AE speech.

A second classification of jazz musicians could be made by region. In the Nationwide Speech Project, Clopper & Pisoni (2006) provide a corpus of high-quality recordings by speakers representing the primary regional dialects of AE, based on a division by Labov et al. (2008).



Figure 18: The major dialects of American English, based on Labov et al. (2008). Taken from Clopper & Pisoni (2006).

Clopper & Smiljanic (2015) analyzed recordings from the Nationwide Speech Corpus of two read passages produced by 10 speakers (5 male, 5 female) of the six regional dialects in figure 18. A significant effect in vocalic nPVI measurements was found between Southern speakers to speakers from all other regions, with higher values for Southern speech. No significant difference was found between the Western and Northern regions, which could reflect East Coast vs. West Coast distinctions:

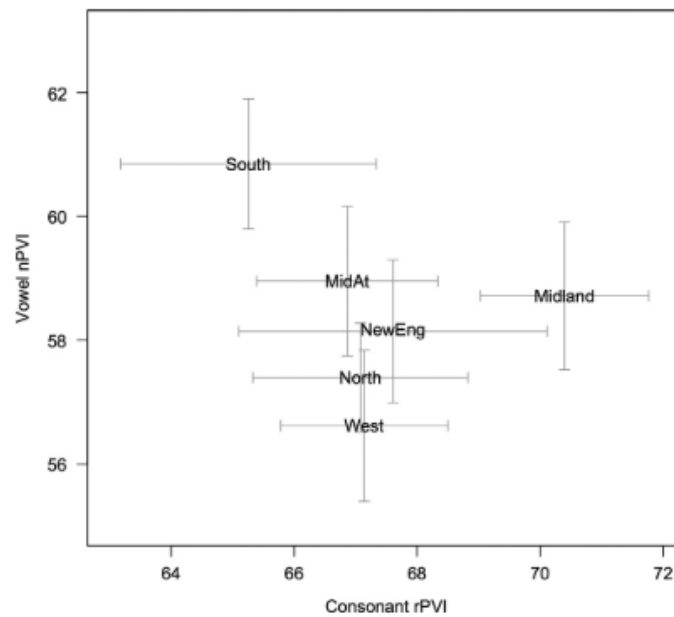


Figure 19: A comparison of mean vocalic nPVI and consonantal rPVI values for regional variants of American English. From Clopper & Smiljanic (2015).

These findings are consistent with McGowan & Levitt’s (2011) findings, showing greater durational variability in Kentucky English compared to Irish and Scottish English dialects. Recall from chapter 3, that Southern speech is characterized by greater contrast between stressed and unstressed vowels due to lengthening and diphthongization of stressed vowels. This phenomenon, commonly referred to as the “Southern drawl”, is associated with the stereotype of Southern speech being slower than Northern speech. In support of this stereotype, Clopper & Smiljanic also found a slower articulation rate and longer and more frequent pauses for Southern (and Midland) speakers than for New England (Northern) speakers. Within

Southern AE, Reed (2020) focused on the Appalachian region. He recorded 24 speakers from Hancock County in north East Tennessee and found a significantly higher vocalic nPVI for these speakers (63.35) compared to the average Southern nPVI in the Nationwide Speech Project corpus (60.85), for speakers who are all outside the Appalachian region. This evidence suggests that Southern AE is rhythmically distinct from other AE dialects.

A distinct musical style which is identified with the Southern region, and specifically the Appalachian region, is bluegrass. Bluegrass derives its name from a band formed by Bill Monroe, "the father of bluegrass" in the 1940's, called the Bluegrass Boys. While bluegrass has some Afro-American influences from jazz and blues music, its main origins are Celtic, from English and Scottish settlers in the Appalachian region. Bluegrass can be considered a sub-genre of country music, whose cultural center is Nashville, Tennessee. Like jazz, bluegrass music is also passed by aural tradition, and also involves musical improvisation, though to a lesser extent. At least in its earlier stages, bluegrass seems to be based primarily on memorized songs and dance tunes. As a relatively recent musical style, also bluegrass offers a considerable amount of musical recordings as well as speech recordings by its pioneer musicians for analysis.

I believe that jazz and bluegrass provide an interesting case study for the relations between speech and musical rhythm. These styles are identified with distinct ethnic communities speaking different regional variants of AE. Limited linguistic evidence suggests that bluegrass musicians, as speakers of Southern AE, should exhibit distinct durational patterns from jazz musicians, with possibly higher nPVI values. Should we also expect a similar pattern in jazz and bluegrass music? Not necessarily. Recall that "swing feel" is the most central rhythmic characteristic of jazz, which emphasizes long-short alternation of consecutive eighth notes. To the best of my knowledge, no systematic study of rhythmic feel in bluegrass playing has been made. Since swing is not central to bluegrass, I expect to find greater variability of eighth note

durations in jazz than in bluegrass. In addition, it would be interesting to see whether within the style of jazz any regional/ethnic differences between black East Coast musicians and white West Coast musicians can be found. In the following sections I present a small-scale study of these three groups of musicians – East Coast jazz, West Coast jazz and bluegrass musicians. The first stage of this study included the selection of musicians and recorded samples for analysis. This is described in section 4.2. Next, speech samples were phonetically transcribed in IPA and musical samples were transcribed in music notation. Both speech and musical samples were then acoustically analyzed in Praat (Boersma & Weenink, 2021) and measured for syllable and note durations. Based on these measurements nPVI values were calculated for each sample. The criteria and method of this analysis are described in section 4.3. The results of the analysis are presented in section 4.4 and further discussed in section 4.5.

4.2. Material selection

4.2.1. The musicians

As in McGowan & Levitt (2011) and Carpenter & Levitt (2016), speech and musical data in this study were obtained from recordings of the same individual musicians. Three pairs of musicians were selected from the three styles under study: (i) East Coast jazz (“hard bop”), (ii) West Coast jazz (“cool jazz”), and (iii) bluegrass. All musicians in the study play sustained pitch instruments, namely instruments capable of controlling the sustained portion of the tone. In jazz, I chose the saxophone as the most representative and influential instrument of the style. For bluegrass, I chose the violin (fiddle), since the saxophone is not a traditional instrument in this style. The reason for preferring sustained-pitch instruments over other instruments is that the acoustic durations of their tones can be more objectively measured. For instance, the banjo is a very typical bluegrass instrument, but as a plucked instrument only the attack of its tone is controlled by the player and its acoustic duration depends mostly on reverberation. By choosing instruments with a relatively similar acoustic envelope (similar attack and sustain patterns) I

assumed that nPVI differences between the musicians would mostly reflect rhythmic differences in their playing rather than timing differences resulting from the acoustic peculiarities of their instruments.

The musicians under study are considered influential figures in the formative period of their style (see appendix A). For East Coast jazz, I chose the two most influential saxophone players in 1950's jazz – John Coltrane (Hamlet, North Carolina) and Sonny Rollins (NYC). In West Coast jazz, I chose Paul Desmond (San Francisco) and Warne Marsh (Los Angeles) as prominent saxophonists of the cool jazz school. In bluegrass I chose two fiddlers from the original Bill Monroe band – Kenneth (Kenny) Baker (Burdine, Kentucky) and Robert “Chubby” Wise (Lake City, Florida).⁸ In principle, this comparison allows us to test two types of correlations: (i) **individual** – for any pair of musicians, regardless of their style and origin, can we find a similar difference in their speech and musical data? (ii) **stylistic** – can we find significant differences between jazz and bluegrass musicians as distinct groups? And within jazz, do we see distinct patterns for (black) East Coast and (white) West Coast musicians? Individual correlations are perhaps the most interesting, because they can reveal contrasts which are neutralized by averaging different musicians under a stylistic comparison. In this paper, however, I will address individual correlations only briefly, as this requires a larger set of data (more musicians, more samples per musicians), and focus mostly on stylistic correlations.

4.2.2. Recorded samples

Sources of all recorded samples in this study are listed in appendix B. The following paragraphs describe the main considerations which guided me in the selection of these samples.

⁸ Lake City is near Jacksonville, Florida, the largest city of the Southeastern U.S. region.

4.2.2.1. Diachronic considerations

In the selection of musical samples my aim was to capture jazz and bluegrass music in their formative period, avoiding more contemporary influences on these styles. In jazz, this was fairly simple, especially for the East Coast musicians, thanks to a vast discography of 1950's jazz. All jazz musical samples are taken from classic recordings made around the second half of the decade. In bluegrass, I was only able to find appropriate samples in later recordings (1970's to 1990's), but these recordings remain faithful to the traditional styles of the bluegrass musicians under study. In the selection of speech samples my preference was also for earlier recordings overlapping with the period in which the musical recordings were taken, yet this was only possible for some of the musicians.

4.2.2.2. Metrical uniformity

A central criterion for selecting musical samples was constraining the corpus to **metrically uniform** musical phrases. As proposed in section 3.3, the purpose of this is to focus on rhythmic nuances of real-time performance by limiting rhythmic variation on the underlying metrical level. All musical phrases in this corpus are composed of consecutive eighth notes. To illustrate this, compare the riddim phrase in figure 20, by Jamaican producer Don Corleon (from Carpenter & Levitt, 2016), with the jazz phrase in figure 21 by saxophonist John Coltrane:



Figure 20: A riddim musical phrase in notated form by Jamaican producer Don Corleon. The phrase is composed of mixed metrical values, marked below by me. Based on these values the metrical nPVI of this phrase equals 30.

Moment's Notice (1:05)

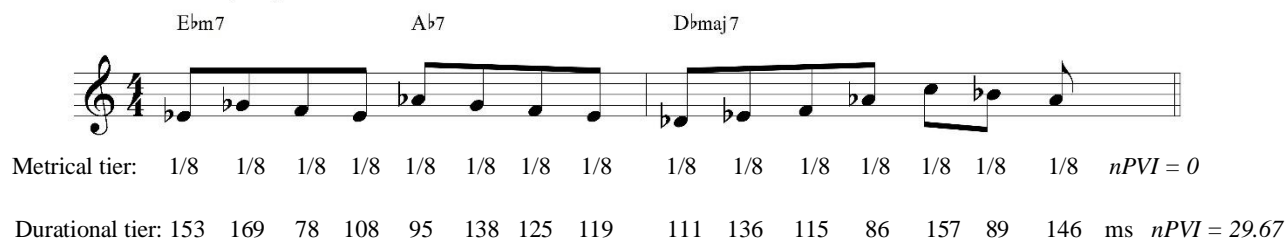
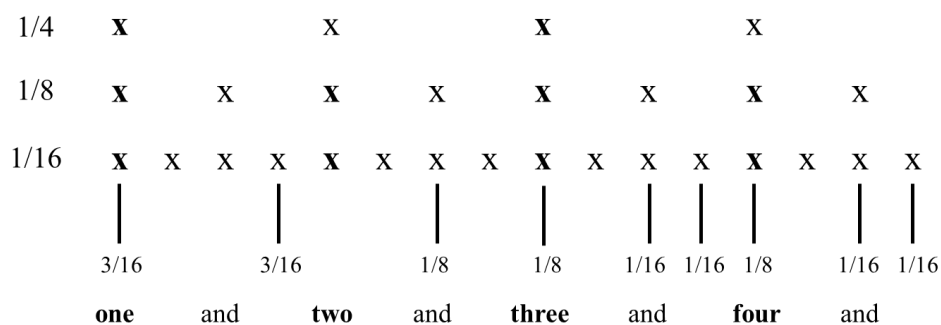


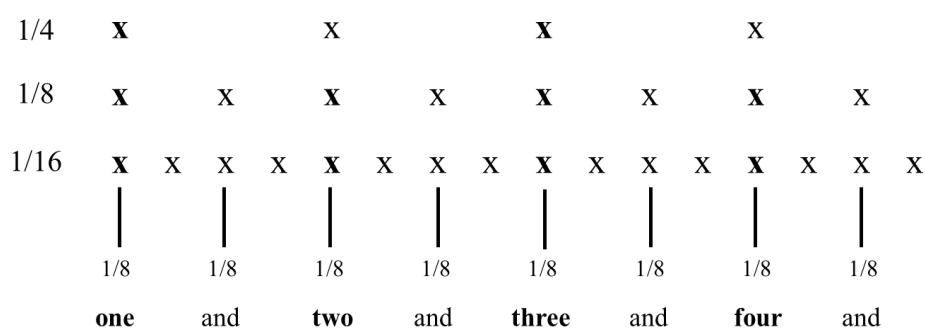
Figure 21: An improvised eighth note phrase in swing feel performed by saxophonist John Coltrane (see reference in appendix B-II). The phrase is composed of equal metrical values of eighth notes. Note durations based on acoustic measurements are marked below in milliseconds. Metrical nPVI = 0, acoustic nPVI = 29.67 (see appendix C-II).

In (14a,b) I sketched metrical grid representations of the Don Corleon phrase in figure 20 and (the first bar of) the Coltrane phrase in figure 21, respectively:

(14) a. **figure 20**: metrical nPVI = 30



b. **figure 21**: metrical nPVI = 0



We can see that the alignment pattern in (14a) is **not equidistant**, reflecting some degree of metrical variability for the riddim phrase in figure 20, with a metrical nPVI value of 30. That is because the phrase in figure 20 is composed of mixed metrical values (1/16, 1/8, 3/16). On average, Carpenter & Levitt found an **acoustic nPVI** value of **41.39** (SD = 12.84) for the Don

Corleon phrases in their corpus. This nPVI value is the combined result of rhythmic variability on both the metrical and the durational (acoustic) tier of his phrases. In comparison, the Coltrane phrase in figure 21 is composed of identical metrical values (consecutive eighth notes).⁹ Accordingly, the metrical alignment pattern of this phrase in (14b) is **equidistant** and **uniform**, with a **metrical nPVI** value of **zero**. When measuring the actual durations of notes in this phrase, as recorded by John Coltrane, we get an **acoustic nPVI** value of **29.67**. But in this case, variability on the durational tier does not overlap with variability on the metrical tier, which is completely uniform. On this basis, I assume that acoustic nPVI values in this type of musical phrases only reflect performance-related variability (micro-timing nuances), independently of higher-order rhythmic patterns.

Guided by this idea, I selected musical samples composed of consecutive eighth note rhythms. To do that, I listened to various tracks by the musicians under study, identified suitable phrases and transcribed these phrases in music notation (by ear). Phrases with rests or pauses, embellishments and tonal effects (trills, growling etc.) were excluded. The only exception to this rule were phrase final downbeats. Final downbeats are often notated as quarter notes regardless of their actual duration. For example, the final note in figure 21 is shorter than some of the other notes in the phrase, but because it falls on a downbeat musicians tend to notate it as a quarter note (ex. 15a), rather than as an eighth note followed by a rest (ex. 15b):

(15)

| | | | |
|-----|--------------|-----|--------------------------------|
| a. | quarter note | b. | eighth note + eighth note rest |
| | | | |
| 157 | 89 | 157 | 89 |
| | 146 ms | | 146 ms |

Since phrases with strictly consecutive eighth notes were difficult to find, excluding phrase endings such as (15a) was impractical. That is one reason why I decided to include cases such

⁹ Assuming long-short alternation in jazz eighth notes is **metrically unspecified**, as discussed in section 3.3.

as (15a) in my corpus. A more fundamental reason was that both speech and musical samples showed a tendency for phrase final lengthening. This is further discussed in section 4.3.1.2. For consistency, I notated both final downbeats and upbeats in my data as eighth notes.

4.2.2.3. Tempo, fluency and sample size

In section 3.3, we saw that eighth note long-short alternation is tempo sensitive. Eight note ratios tend to be larger in slower tempos and smaller in faster tempos. Around a tempo of 200 bpm Friberg & Sundström (2002) found that eighth notes approximated the so-called classic swing ratio of 2:1. Although the nPVI measure includes a tempo normalization component, I nevertheless preferred to avoid too fast or too slow tempos and focus on musical phrases around 200 bpm (“medium-up” tempo in jazz terminology). In speech samples, my preference was for fluent utterances with a steady speaking rate, including no pauses, hesitations or special intonation as much as this was possible. Ideally, in this type of studies samples should be matched in length for syllable and note number. In practice, this was not possible due to the limited sources of recordings by some of the musicians, mostly speech recordings. The table in (16) summarizes the main parameters related to sample size, tempo and duration:

(16) Main sample parameters:

| | Speech | Music |
|---------------------------|--------------------------|------------------------|
| Total samples | 30 | 30 |
| Segmented units | 311 syllables | 388 notes |
| Avg. sample length | 10.36 syl. per utterance | 12.93 notes per phrase |
| Avg. unit duration | syllable = 201.51 ms | note = 133.66 ms |
| Avg. tempo | 5.09 syl./second | 232.86 bpm |

On average, sample lengths are quite similar with 10.36 syllables per utterance compared to 12.93 notes per musical phrase. The average syllable duration in the corpus is 201.51 ms

relative to an average speech rate of 5.09 syllable per second. The average note duration is 133.66 ms relative to an average tempo of 232.86 bpm.

4.3. Segmentation

This section describes the methodology of segmentizing the speech and musical samples in the corpus to separate durational units – syllable durations in speech utterances and note durations in musical phrases.

4.3.1. Speech analysis

The analysis of speech samples included two stages: (i) syllabification of utterances by phonological criteria, (ii) measuring syllable durations by phonetic-acoustic analysis. These stages are described below.

4.3.1.1. Syllabification

Unlike previous studies, in this study I chose to segmentize speech samples to syllables rather than vocalic and consonantal intervals. This is based on the idea that the syllable is the basic rhythmic unit of speech and most comparable to musical tones (see discussion in section 2.4). Sound quality in my samples also made it difficult for me to accurately mark specific vowel and consonant boundaries, and marking syllable boundaries instead seemed more practical to me.

As opposed to speech segments, the syllable is a perceptual unit which is not discernible on the acoustic signal. My guiding principle here was to aim for **consistent** rather than **objective** method of syllabification. Automatic models of syllabification offer interesting insight for this purpose. Bartlett et al. (2009) review different approaches to automatic syllabification, from which I adopt here what they call the *categorical approach*. For descriptive purposes, this approach can be presented in the form of OT-like constraint rankings on English syllabification. I will illustrate this with specific examples from the corpus.

Onset Maximality

By default, English consonants are syllabified according to the principle of **onset maximality** (MAXONSET) which prefers consonants in onset position over coda position, excluding word final consonants. For instance:

(17) *different*: [di.frənt] >> *[dɪf.rənt]

In (17), MAXONSET prefers a complex CV.CCVCC structure with a [fr] cluster in the onset of the second syllable, over a simplified CVC.CVCC structure with a [f] coda in the first syllable.

SSP and the Legality Principle

MAXONSET is subject to the **Legality Principle** (LEGALITY), which filters impossible onsets and codas in English. LEGALITY is composed of: (i) the **Sonority Sequencing Principle** (SSP), and (ii) a set of English-specific conditions on syllabification. SSP requires a gradual rise and fall of sonority to and from the syllable nucleus. Bartlett et al. consider adjacent consonants as a sonorous cluster if these consonants differ by at least two levels on the sonority scale in (18):

(18) 0-Obstruents, 1-Nasals, 2-Liquids, 3-Glides, 4-Vowels

For example, in (19):

(19) *practicing*: [præk.tə.sɪŋ] >> *[præ.ktə.sɪŋ]

the [pr] cluster in the onset of the first syllable satisfies LEGALITY by a rise of two levels on the sonority scale (obstruent to liquid). In the second syllable, a complex [kt] onset with two adjacent obstruents is ruled out by LEGALITY as a non-sonorous cluster. In this case, the [k] must be associated to the coda of the first syllable despite a violation of MAXONSET. Here is another example:

(20) *interested*:

| /intrəstɪd/ | LEGALITY | MAXONSET |
|----------------------------|----------|----------|
| a. ɪ.ntrə.stɪd | *! | |
| b. ɪ n.trə.stɪd | | * |
| c. ɪn.trəs.tɪd | | ** |

Candidate (20a) satisfies MAXONSET by syllabifying [n] in the onset of the second syllable, but is ruled out by a LEGALITY violation of a non-sonorous /ntr/ cluster. In addition to SSP, LEGALITY includes some idiosyncratic conditions on English syllabification, such as the eligibility of [s] in non-sonorous clusters. This condition qualifies the [st] cluster in candidate (20b) as a possible English onset. Consequently, (20b) is favored by minimally violating MAXONSET with [n] in coda position of the first syllable instead of the onset of the second syllable. Although candidate (20c) with [s] in the coda of the second syllables is more sonorous, it is ruled out by an unjustified violation of MAXONSET.

Bartlett et al. list additional English-specific legality conditions from Kenstowicz (1994), prohibiting various types of complex onsets, such as a cluster with a voiced fricative (e.g., *[vr], *[zw]) or a cluster of a non-strident coronal followed by a lateral (e.g., *[tl], *[dl] as in (21):

(21) *fiddler*: [fɪd.lər] >> *[fɪ.dlər]

Although the [dl] onset in *[fɪ.dlər] satisfies SSP, it is filtered by LEGALITY as an impossible English onset. In this case a CVC.CVC structure is preferred over a CV.CVCC structure, despite a MAXONSET violation by the coda in the first syllable. Finally, I assumed that the velar nasal [ŋ] is restricted to coda position, as in (22) (assuming no ambisyllabicity, see below):

(22) *singing*: sɪŋ.ɪn >> *sɪ.ŋɪn

Word boundary

As a generalization, I avoided syllabifying over word boundary. This preference can be represented by a highly-ranked Prosodic Word Alignment Constraint (PWDCON, e.g. Selkirk, 1995) requiring an overlap between prosodic and lexical word boundaries:

(23) *wondered about*:

| / wʌndərd#əbaʊt/ | PWDCON | LEGALITY | MAXONSET |
|---------------------------------|--------|----------|----------|
| a. wʌ.ndə.rd#ə.baʊt | *! | ** | |
| b. wʌn.dər.d#ə.baʊt | *! | | ** |
| c. ɪ wʌn.dərd#ə.baʊt | | | *** |

Candidate (23a) maximally satisfies MAXONSET but is ruled out by syllabifying over word boundary (PWDCON) as well as by forming non-sonorous onsets (LEGALITY). Candidate (23b) satisfies LEGALITY over MAXONSET, but is still ruled out due to a violation of PWDCON. Candidate (23c) is favored despite multiple violations of MAXONSET, to avoid syllabification over word boundary.

In few cases where word boundaries were unclear, LEGALITY had to be favored over PWDCON. This typically involved cliticization and reduced forms, such as the cliticized *is* in (24):

(24) “*All conventional **harmony’s built** on the major and the minor scale*”:

a. [(hɑr.mə.ni)_{LexWd}-z]_{PWd} [(bɪlt)_{LexWd}]_{PWd}

b. *[(hɑr.mə.ni)_{LexWd}]_{PWd} [z-(bɪlt)_{LexWd}]_{PWd}

In both (24a) and (24b) the cliticization of *is* results in non-overlapping lexical word and prosodic word boundaries. In (24b), procliticization to the following word (**’s-built*) results in

an illegal onset with a voiced fricative. Encliticization of *is* to the previous word (*harmony's*) satisfies LEGALITY and is therefore favored.

Lastly, in the cases of /t/ flapping over word boundary I preferred an onset analysis of the flapped [ɾ], as in (25a):

(25) *What I:*

- a. onset analysis – [wɑ.rɑɪ]
- b. coda analysis – [wɑɾ.ɑɪ]
- c. ambisyllabic analysis – [wɑɾ.rɑɪ]

For alternative analyses of intervocalic consonants (coda, ambisyllabic) see Nesbit (2018). As argued above, my methodology of syllabification does not aim to find the most accurate analysis for each case, as this is impossible to do, but to adopt consistent criteria for the entire corpus.

4.3.1.2. Acoustic analysis of speech samples

The acoustic analysis of speech samples was performed in Praat and included the following steps:

- (i) marking syllable boundaries, based on the syllabification criteria above.
- (ii) marking syllable durations.
- (iii) calculating the nPVI value of each sample.

The marking of syllable boundaries was done by acoustic (visual) and perceptual (audible) cues. In some cases, and depending on sound quality, changes in formant structure and troughs in the amplitude contour corresponded to transition points between syllables quite clearly:

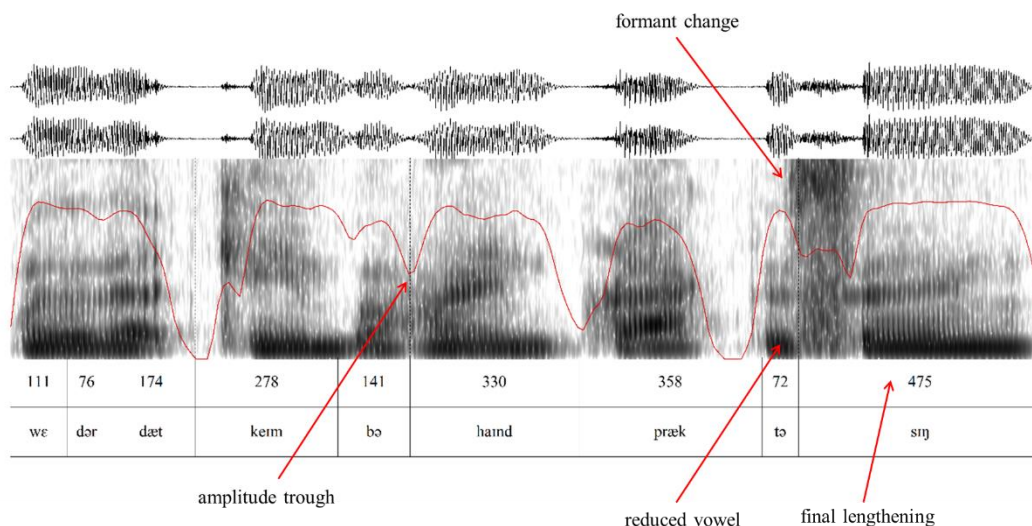


Figure 22: Waveform and spectrogram analysis of an utterance by jazz saxophonist Paul Desmond (see appendix C-I). Intensity contour marked in red, syllables durations are in milliseconds. Syllabic nPVI = 74.6.

To fine-tune syllable boundaries, I also relied on ear judgment. When a boundary was heard as crossing over some segment in the previous or following syllable, I readjusted its location until the two syllables sounded as separate as possible.

To calculate the nPVI value of an utterance, I marked the duration of each syllable as indicated by its boundary markings in Praat. As figure 22 illustrates, the (syllabic) nPVI value of an utterance is influenced by the relative differences in duration between adjacent syllables. In this example, sharp contrasts between adjacent syllables result in a high nPVI value of 74.6. One factor that generally influenced high nPVI values was **vowel reduction**. As a rule of thumb, syllables with two-digit durations (below 100 ms) greatly contributed to higher values. Another contributing factor is **phrase final lengthening**. Phrase final syllables tended to be longer than other syllables in the phrase. Compared to an average duration of 201.5 for all syllables in the corpus (311 syllables), the average duration of final syllables (30 final syllables) equals 386.1 ms. To control for phrase ending effects, Reed (2020) excluded the final feet of utterances from his measurements. However, Reed’s speech samples were obtained from a rich database of high-quality recordings, offering a much larger selection of samples for analysis.

Due to the limited selection of samples in my corpus, I had to include some relatively short utterances which cannot be shortened any further. Yet from a theoretical standpoint too, final lengthening could be considered an integral part of the utterance’s rhythmic pattern, and therefore should be taken into account in the overall analysis of the data. As mentioned in section 4.2.2.2 above, a similar phenomenon was also found in musical phrases. The average duration of final notes in the corpus (30 final notes) is 174.5, compared to an average duration of 133.6 for all notes in the corpus (388 notes total). For these reasons, I decided to include phrase final durations in my analysis.

To conclude this discussion, the following example illustrates the difference between a syllabic and a segmental analysis of speech samples:

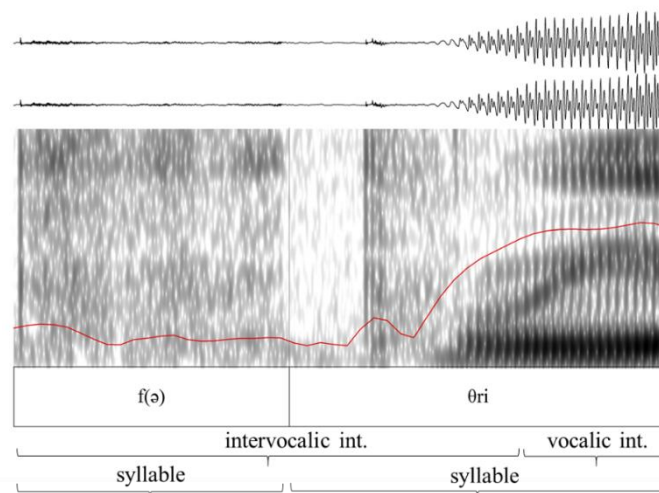


Figure 23: A two-word fragment (“for three”) from an utterance by bluegrass fiddler Kenny Baker (see appendix C-I).

In this short fragment, the preposition *for* is phonetically reduced to an [f] consonant. Audibly, only the friction of [f] is heard in this syllable. Visually, no vowel formants appear on the spectrogram. Under a segmental acoustic analysis, this [f] forms an intervocalic interval together with the adjacent [θr] cluster in *three*. For vocalic nPVI measurements, this interval would be excluded from the data. However, under a syllabic analysis the same [f] can be considered as a reduced syllable with a phonetically silent nucleus and should be measured in

succession with adjacent syllables. I believe that such an analysis is rhythmically more accurate.

4.3.2. Analysis of musical samples

The musical samples were analyzed in Praat in a similar procedure to that of the speech samples:

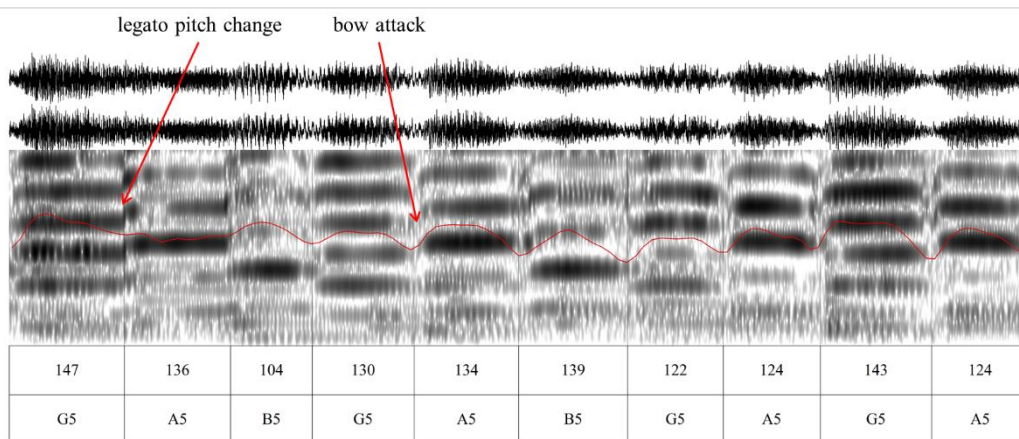


Figure 24: Acoustic analysis of an eighth note phrase performed by bluegrass fiddler Kenny Baker (“Sam’s Tune”, 0:53, see appendix C-II).

Because musical sound is more periodic than speech sound, pitch and formant structure in the musical samples were generally more consistent. In figure 24 we see that the sustained portion of the tone is quite easily discernible. We can also see quite clearly when pitches change to a different series of formants. The main difficulty in the analysis of the musical samples was in determining the precise transition point between the notes. In this respect, musical spectrograms were often less informative than speech spectrograms. One reason for this is that the musical phrases in the corpus contain a lot of background acoustic information from other instruments in the ensemble. This makes the overall intensity contour of the samples more uniform than in the speech samples, where a single speaker speaks alone. Another reason is that syllables have a more diverse internal structure than musical tones. SSP as a universal constraint on syllabic structure disfavors continuous sonorous sequences. Purely vocalic

speech sequences (e.g. V.V.V) are phonetically and phonologically marked. Another way to look at this is in terms of balance between periodic and non-periodic sound. Speech utterances require more balance between periodic and non-periodic sounds than musical phrases. In music, continuous periodicity or sonority is often favorable, especially in melodic sequences. Melodies are often played in *legato* articulation (in the musical sense), in which tones are maximally sustained and minimally interrupted by non-periodic attack noises. In figure 24 above, the first three notes (G5-A5-B5) are played in legato articulation. We can see this in the smooth transition between their formants and lack of decrease in intensity between the notes. In this case, we have to rely mostly on formant frequencies and ear judgement for marking note boundaries. The rest of the notes in figure 24 are separated by gentle bow strokes, known in music terminology as *detaché* (“detached”) or *non-legato* articulation. In the transition between these notes, we *can* see some decrease in periodicity and intensity as an indication of note boundary.

While the human articulatory system is essentially universal, musical articulation greatly depends on the mechanical and acoustic properties of different instruments. In the violin, sound is produced by arm motion. Notes on the violin are attacked by bow strokes and separated by lifting the bow from the string. The saxophone, as a wind instrument, uses more similar articulatory processes to that of the speech apparatus. In saxophone playing, sound energy is produced by pulmonic airstream causing a reed to vibrate against the player’s bottom lip. Similar to the production of stop consonants, notes on the saxophone are attacked by blocking the mouthpiece with the tongue, building air pressure and releasing the tongue. Figure 25 zooms-in on a fragment from a jazz saxophone phrase by Sonny Rollins:

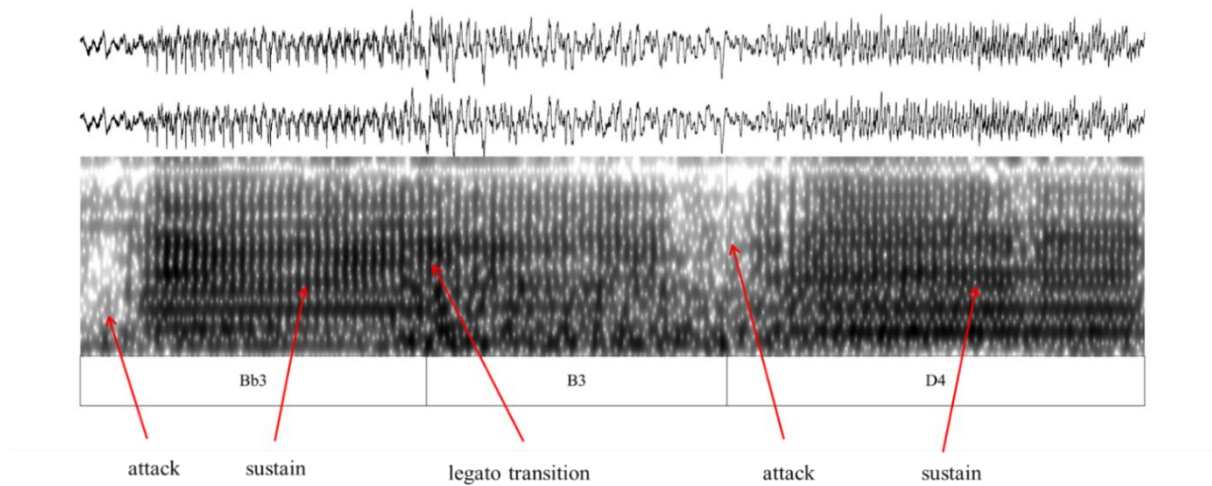


Figure 25: A three-note fragment from a jazz phrase performed by saxophonist Sonny Rollins (“But Not For Me”, 2:40, see appendix C-II).

The tonguing (attack) of the first note in this fragment (Bb3) is marked by a short non-periodic onset before the sustained part of the tone. The second note (B3) is slurred from the first note in legato. This makes it more difficult to determine the precise transition point between these notes. In the transition to the third note (D4) we see a short decay of B3 followed by the onset of D4. In the middle of this transition, we see a dip in the waveform which I interpret as the boundary point. As in the case of syllables, note boundaries could not be determined purely by acoustic/visual cues and inevitably involved ear judgement.

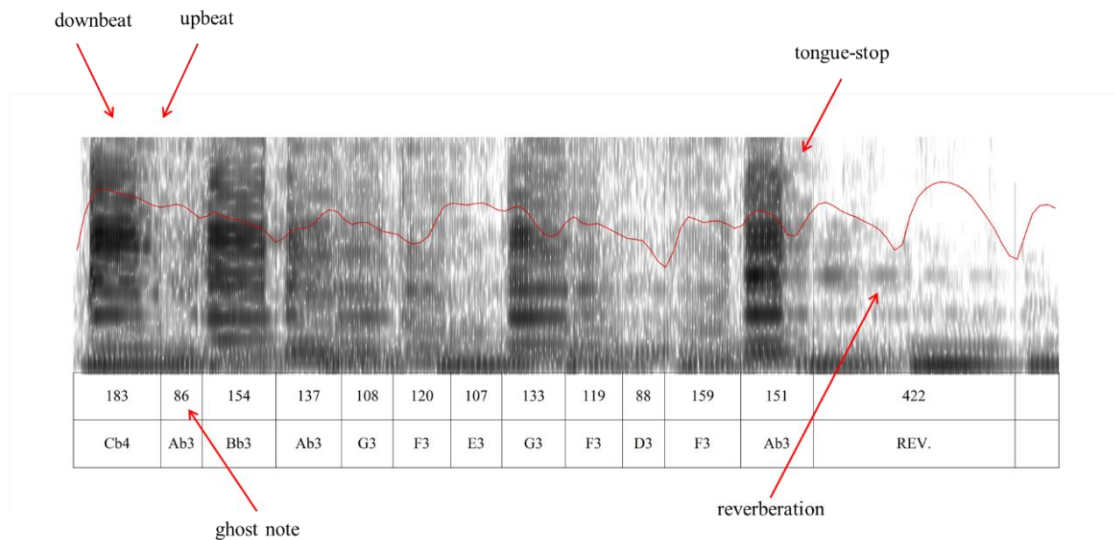


Figure 26: A eighth note jazz phrase by saxophonist Sonny Rollins (“Strode Rode”, 0:56, see appendix C-II). The final note resonates for additional 422 ms after its physical cut-off by the player.

Let us conclude with another phrase by Sonny Rollins. This phrase ends with the typical “bebop” or “doo-dat” articulation (F3-Ab3). The final note of the phrase (Ab3) ends with an abrupt blocking of the airstream by the tongue known as *tongue-stopping* (“dat”). This abrupt ending is seen by a slight drop in energy about 151 ms after its attack point. However, by acoustic reverberation this note keeps resonating in the recording more than twice as long. Both acoustically and perceptually, the note decays approximately 570 ms after its onset. Should this additional reverberation be measured as part of the total duration of this note? Here, I believe that this additional duration is rhythmically irrelevant and that the note boundary should correspond to the articulatory closure of the tone rather than to external factors such as reverberation. This example illustrates the type of considerations involved in the attempt to capture rhythmic phenomena by instrumental measurements. I believe that this equally applies to speech as it does to music. Both syllable and note durations in this study should be regarded as rough approximations, rather than absolute objective measurements.

Figure 26 also illustrates the musical phenomenon of *ghost notes*, which shows here great similarity to vowel reduction in speech. The term ghost notes describes notes that are played

notably shorter than their expected duration. Ghost notes also tend to be played with a weaker tone, sometimes lacking consistent pitch. We see this, for example, in the second note of the phrase (Ab3), which is well below 100 ms and with much weaker formants than the adjacent notes. Ghost notes increase the rhythmic contrast in the phrase and have a percussive effect which makes them an important rhythmic device in jazz and related styles. Similar to what we saw in the speech samples, such “phonetically reduced” notes significantly increase nPVI values, especially when alternating with adjacent prominent notes.

Finally, in each phrase I calculated the BUR value for each pair of downbeat-upbeat eighth notes (see section 3.3). BUR is calculated by dividing the duration of the downbeat eighth note by the duration of the following upbeat eighth note. Phrase initial upbeat notes (anacrusis) and phrase final downbeat notes were excluded. Notice that for the first pair of eighth notes in figure 26 (Cb4-Ab3) we get a relatively high BUR of 2.12 ($183/86$), because of the 86 ms ghost note on the upbeat. This BUR represents an uneven swing feel with a swing ratio of more than 2:1 between the downbeat eighth and the upbeat eighth. Compare this to the next pair of eighth notes in the phrase (Bb3-Ab3) which are played almost evenly, with a BUR approximating 1 ($154/137=1.12$).

4.4. Results

Let us return to P&D’s claim that “spoken prosody leaves an imprint on the music of a culture”. How can the prosodic patterns of a language can be “imprinted” on the music of its speakers? Clearly, this would have to involve musicians from that culture, who are native speakers of that culture’s language. The focus of this study has been on performance-related rhythmic nuances in music and speech. On this rhythmic dimension, articulatory and motoric processes of speech timing could also surface in the playing habits of musicians, resulting in similar patterns of durational variability. The framework laid out here allows to test this hypothesis. As in McGowan & Levitt (2011) and Carpenter & Levitt (2016), this study analyzes synchronic

speech and musical data produced by the same individual musicians. We can therefore hypothesize that **for any pair of musicians**, regardless of their style and cultural background, we expect correlated patterns of durational variability in music and in speech. This was tested by Carpenter & Levitt, who compared individual speech and music nPVI averages for the musicians in their corpus:

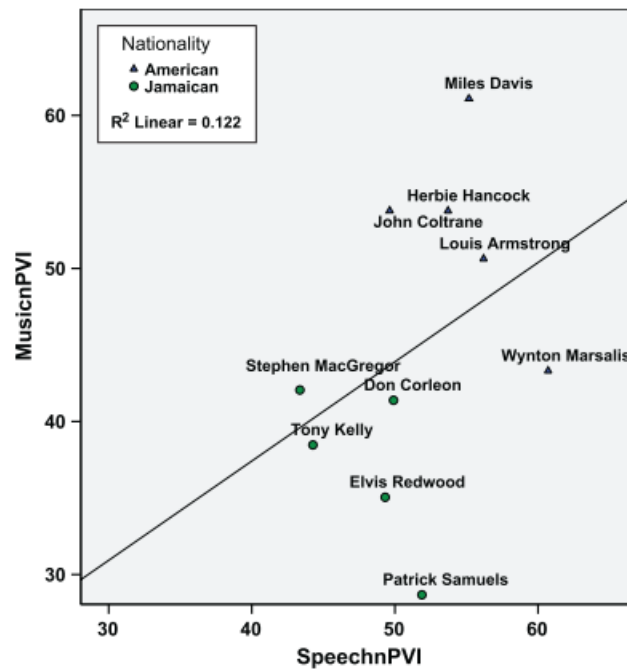


Figure 27: Comparison of speech and musical nPVI averages for individual musicians in Carpenter & Levitt (2016).

In this comparison, Carpenter & Levitt could not find individual significant differences between musicians. A similar comparison for the musicians in my corpus is given in figure 28 below:

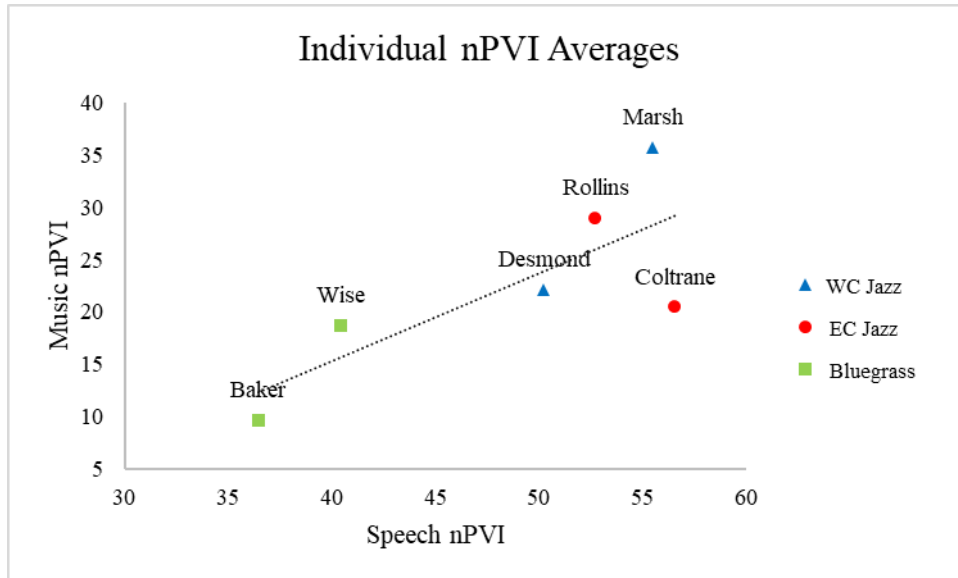


Figure 28: A comparison of individual speech and music nPVI averages for bluegrass musicians (Baker, Wise), East coast jazz musicians (Coltrane, Rollins) and West Coast jazz musicians (Desmond, Marsh).

Here too, no correlation is found between speech and musical data on the individual level. Such a comparison requires a much larger corpus with more musicians and more samples per musician. Hopefully, the current set of data will be expanded in the future to shed more light on this this interesting aspect of speech and music similarity. What we do see in figure 28 is the separate grouping of jazz and bluegrass musicians by style, but not quite as predicted. The table in (26) compares the average and median speech nPVI values for each group of musicians:

(26)

| | Average Speech nPVI (SD) | Median Speech nPVI |
|-----------------|--------------------------|--------------------|
| East Coast jazz | 54.62 (9.78) | 56.54 |
| West Coast jazz | 52.86 (10.75) | 50.02 |
| Bluegrass | 38.43 (7.24) | 37.12 |

Previous research suggests that Southern AE is rhythmically distinct from other variants of AE, with significantly **higher** nPVI values, especially in the Appalachian region. Among other AE variants (e.g., West and North), no significant differences in nPVI values were found so far,

nor between African and European AE speakers (see section 4.1). This pattern maintains in the data collected here, but with a different trend. Southern bluegrass musicians were found distinct from East Coast and West Coast jazz musicians with significantly **lower** speech nPVI values. A Kruksal-Wallis test found a significant between-groups effect on nPVI ($H(2)=11.74$, $p=0.004$). Post hoc comparisons (Mann-Whitney tests using a Bonferroni-adjusted α level of $0.05/3=0.017$) show that the difference between the mean ranks of East coast jazz musicians, 20.10 (median = 55.56), and bluegrass musicians, 8.00 (median = 37.13), is statistically significant ($U=11.00$, $p=0.003$) and the effect size is medium ($r = -0.66$); and the difference between the mean ranks of West Coast jazz musicians, 18.40 (median = 50.03), and bluegrass musicians is also statistically significant ($U=14.00$, $p=0.007$) and the effect size is also medium ($r=-0.61$). No significant difference was found between East Coast and West coast jazz musicians ($p=0.60$). Interestingly, the same pattern was also found in the musical phrases produced by these musicians:

(27)

| | Average Music nPVI (SD) | Median Music nPVI |
|-----------------|--------------------------------|--------------------------|
| East Coast jazz | 24.67 (9.35) | 24.81 |
| West Coast jazz | 28.84 (10.61) | 31.13 |
| Bluegrass | 14.11 (8.15) | 12.06 |

Significantly lower nPVI values were found for musical phrases produced by bluegrass musicians, compared to phrases by both East Coast and West Coast jazz musicians (Kruksal-Wallis test, $H(2)=9.68$, $p = 0.008$). Post hoc comparisons (Mann-Whitney tests using a Bonferroni-adjusted α level of $0.05/3=0.017$), show that the difference between the mean ranks of East Coast jazz musicians, 17.06 (median = 24.82), and bluegrass musicians, 8.06 (median = 12.7), is statistically significant ($U=16.00$, $p=0.01$) and the effect size is medium ($r=-0.57$);

the difference between the mean ranks of West Coast jazz musicians, 20.30 (median = 31.14), and bluegrass musicians is also statistically significant ($U=15.00$, $p=0.008$) and the effect size is also medium ($r=-0.59$). Here too, no significant difference was found between East Coast and West coast jazz musicians ($p=0.33$). Figure 29 compares the overall jazz nPVI averages (East Coast and West Coast) with the bluegrass averages in music and in speech:

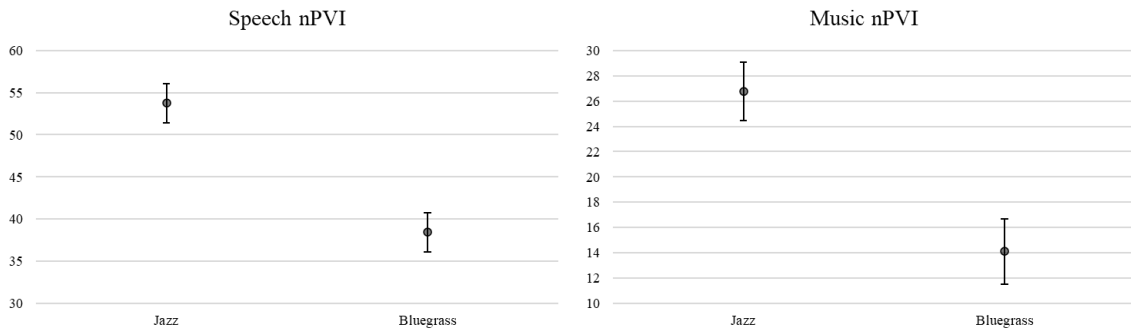


Figure 29: A comparison of nPVI averages for speech utterances and musical phrases produced by four jazz musicians (2 East Coast, 2 West Coast) and 2 bluegrass musicians. Jazz speech nPVI = 53.7 (10.3), bluegrass speech nPVI = 38.4 (7.2), jazz music nPVI = 26.7 (10.2), bluegrass music nPVI = 14.1 (8.1).

As in previous studies, we see the same pattern of difference in speech and music. In speech, jazz musicians scored higher (median = 54.12, mean rank = 19.25) than bluegrass musicians (median = 37.13, mean rank = 8.00). Mann-Whitney U-value was found to be statistically significant ($U=25.00$, $z = -3.30$, $p = 0.001$) and the effect size was medium ($r=-0.60$). Also in music, jazz musicians scored higher (median = 27.72, mean rank = 18.92) than bluegrass musicians (median = 12.06, mean rank = 8.60). Mann-Whitney U-value was found to be statistically significant ($U=31.00$, $z=-3.04$, $p=0.002$), and the effect size was medium ($r=-0.55$). This is compatible with the general idea that languages/dialects and corresponding musical styles share common rhythmic characteristics. However, the specific patterning in this set of data is not compatible with the initial prediction that bluegrass musicians, speaking Southern AE, should exhibit a higher degree of durational variability in speech. This issue is further discussed in section 4.5 below.

In chapter 3, I proposed that beat-upbeat ratios in musical performance (BUR) could also contribute to the comparison of speech and musical rhythm. Unlike the nPVI, the BUR measure takes into account also metrical prominence relations between adjacent downbeat and upbeat notes. I proposed that a comparable phonological measure would calculate durational relations between strong (stressed) and weak (unstressed) syllables. Constructing such a set of data is beyond the scope of this paper. In the current study, I only tested this idea by comparing BUR measurements with speech nPVI measurements. An advantage of the BUR in a small-scale study such as this is that it takes note pairs as its unit of measurement rather than musical phrases. This provides a larger set of data for analysis (186 note pairs compared to 30 musical phrases in the entire corpus). Because BUR measurements ignore unpairable notes in phrase edges (phrase initial upbeats and phrase final downbeats), I found it more useful to average BUR values over the entire set of samples in each style. Figure 30 compares the average BUR for the jazz and bluegrass samples with their speech nPVI averages:

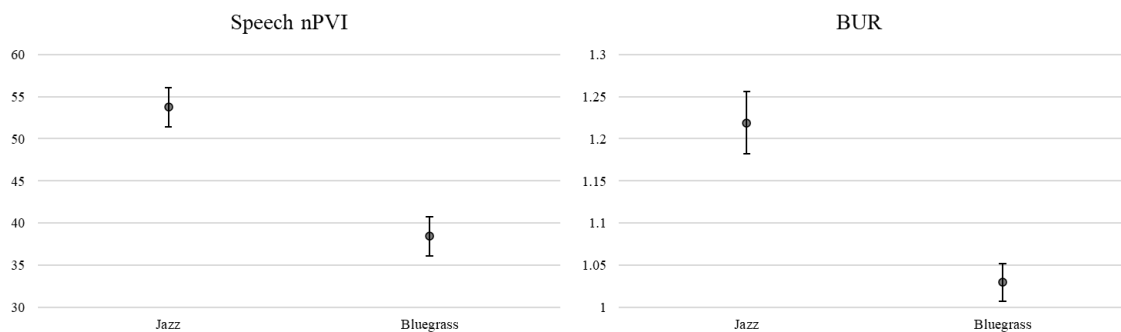


Figure 30: A comparison of speech nPVI (repeated from figure 29) and BUR averages of jazz and bluegrass musical phrases. Average jazz BUR = 1.21 (0.4), average bluegrass BUR = 1.02 (0.18).

A BUR average ranging from roughly 1 to 1.2 indicates that both jazz and bluegrass phrases in the corpus are played quite evenly. An average BUR around 1.2 seems especially low for jazz phrases, which are all played in swing feel. However, this finding is compatible with Benadon's (2006) findings of jazz eighth notes being played much more evenly than is commonly believed. Still, compared to bluegrass eighth notes which were practically even with an average

BUR around 1, jazz eighth notes were found significantly less even. Jazz musicians scored higher on BUR (median = 1.14, mean rank = 102.33) than Bluegrass musicians (median = 1.04, mean rank = 77.45). Mann-Whitney test produced statistically significant results ($U=2900.50$, $z=-3.02$, $p=0.003$) and the effect size was medium ($r=-0.55$). In addition, higher standard deviation in jazz BUR's compared to bluegrass BUR's (0.4 vs. 0.18) indicates that jazz eighth notes ratios in the data are more variable than bluegrass eighth notes.

Clearly, BUR values pattern with music nPVI values, because the two are measures of musical durational variability. It is not surprising, then, to find that BUR values also pattern with speech nPVI's in this corpus, as illustrated in figure 30. There is, however, a fundamental difference between these measures. The BUR includes a structural component, grouping note pairs by **metrical prominence** (downbeat-upbeat). The nPVI, as opposed to that, is indifferent to prominence relations and only takes into account surface durational relations between adjacent elements, in our case syllables and musical tones. We therefore do not expect a one-to-one correspondence between the measures. When comparing East Coast jazz and West Coast jazz BUR's separately, a different pattern emerges than before. West Coast jazz BUR values were found significantly different from both East Coast jazz and bluegrass BUR's. No significant difference in BUR values was found between East Coast jazz and bluegrass BUR values. A Kruskal-Wallis test found a significant between-groups effect on BUR, ($H(2)=18.66$, $p < 0.001$). Post hoc comparisons (Mann-Whitney tests using a Bonferroni-adjusted α level of $0.05/3=0.017$) show that the difference between the mean ranks of **East Coast** jazz musicians, 86.35 (median = 1.04), and **West Coast** jazz musicians, 117.80 (median = 1.23), is statistically significant ($U=1261.00$, $p=0.005$) and the effect size is small ($r=-0.26$); and the difference between the mean ranks of **West Coast** jazz musicians and **bluegrass** musicians, 77.45 (median = 1.04), is also statistically significant ($U=1146.50$, $p < 0.001$) and the effect size is also small ($r=-0.38$). No significant difference was found between **East Coast** jazz musicians

and **bluegrass** musicians, ($p=0.520$). The table in (28) compares the average and median BUR values of these three styles:

(28)

| | Average BUR (SD) | Median BUR |
|-----------------|-------------------------|-------------------|
| East Coast jazz | 1.12 (0.38) | 1.04 |
| West Coast jazz | 1.3 (0.4) | 1.22 |
| Bluegrass | 1.02 (0.18) | 1.04 |

According to these data, West Coast jazz musicians play less evenly than both East Coast jazz musicians and bluegrass musicians. However, we see that East Coast and West Coast jazz have similarly higher BUR standard deviations than bluegrass. This suggests that although East Coast jazz phrases were played more evenly than West Coast jazz phrases, jazz phrases in overall are characterized by more rhythmic variability than bluegrass.

4.5. General discussion

In this study, I took a slightly different route from previous works. On the linguistic domain, I measured syllable durations rather than vowel durations in comparison with note durations. This was based on the view that the syllable is the basic rhythmic unit of speech, and as such it is most comparable to musical tones. This choice remains consistent with previous results. On average, jazz musicians produced speech utterances and musical phrases with significantly greater durational variability than bluegrass musicians. This supports the idea that syllabic measurements can be used as a tool for comparing speech and musical rhythm. It remains to be seen whether syllable durations are in fact a better rhythmic measure than the commonly used method of vocalic and intervocalic durations. On the musical domain, my focus was the phenomenon of long-short alternation in eighth notes playing, also known as swing feel. As originally observed by McGowan & Levitt (2011), the ratio of unevenness in consecutive

eight notes is a possible point of similarity between speech and musical rhythm. I argued here that this unevenness is a property of music performance (“rhythmic feel”), which should be studied independently of the underlying metrical properties of musical phrases. To do this, I constrained my musical data to phrases composed of strings of consecutive eighth notes. I proposed that such phrases represent a metrically uniform structure with no variability on the metrical rhythmic tier. Previous studies analyzed musical phrases composed of mixed metrical values. Consequently, in these studies nPVI measurements incorporate some degree of variability on the metrical tier. One could argue that similarly in speech, nPVI values are the combined outcome of variability on both the underlying prosodic level (syllable and foot structure) and the phonetic-acoustic surface, and that the two should not be distinguished. In section 3.1.4, I proposed that metrical variability in music and in speech can be studied independently by comparing formal representations of prosodic/metrical structures in phonological and musical theories. A comprehensive investigation of speech and musical rhythm should therefore pursue these three possibilities: (i) an independent comparison of the metrical properties of both domains, (ii) a comparison of performance-related durational patterns in these domains, and (iii) a comparison of their overall durational patterns on surface.

Because swing feel is a central characteristic of jazz and not central to bluegrass, I assumed that jazz phrases should exhibit greater durational variability than bluegrass phrases. This is reflected in the data, with significantly higher nPVI average for jazz phrases compared to bluegrass phrases. One feature of swing feel is the use of **ghost notes**, the equivalents of reduced syllables in speech. We saw that ghost notes tend to push phrase nPVI values higher. Based on my personal impression, the use of ghost notes is less common in bluegrass than in jazz, but the current set of data is too small to test this. Just as an idea, I compared and found that the average shortest note in the jazz phrases is slightly shorter than in the bluegrass phrases (99.05 ms vs. 103.4 ms). However, this does not take into account differences in tempo between

the phrases in both styles. Ideally, all phrases in the corpus should have been matched for tempo. In my data, however, bluegrass phrases were slightly faster, with an average bpm of 246.7 compared to 225.95 bpm in jazz. Therefore, if tempos were matched, we might have not even found this small difference. More crucially, higher tempos reduce the overall contrast between long and short notes in absolute values. Hence, one could argue that lower nPVI values in bluegrass phrases are simply a result of higher tempos. This is less likely, first because the difference in tempos is not dramatic (around 20 bpm) and mainly because nPVI measurements are normalized for tempo. Note that a similar claim can be made for speech samples. The average speech rate for the two bluegrass musicians in the corpus is 5.75 syllables per second, compared to 4.76 syllables per second for the jazz musicians. This by itself is purely accidental, and mostly affected by one of the musicians (Chubby Wise) who is an extremely fast talker with some utterances reaching more than 8 syllable per second (!). Clearly, this is not representative. In comparison, Clopper & Smiljanic (2015) found an average speech rate of 5.40 syllables per second for Southern speakers. However, if nPVI values were shifted by speech rate we would expect speakers with faster speech rate to have lower nPVI values, and this is clearly not the case. For instance, bluegrass fiddler Kenny Baker has the lowest speech nPVI average in the corpus (36.45) despite a relatively slow speech rate in the data (4.56 syllables per second compared to an average rate of 5.09 for all musicians, see section 4.2.2.3). I therefore see no clear connection between nPVI values and tempo. Without good reason to assume otherwise, I therefore assume that nPVI values reflect relative durational relations independently of tempo. With that said, perfect matching in tempo would be preferable in future studies.

Another factor that seems to affect nPVI values in music and speech is **final lengthening**. On average, the longest syllables in utterances produced by bluegrass musicians were notably shorter than longest syllables produced by jazz musicians (340.5 ms vs 470.7 ms). Similarly,

the longest notes in bluegrass phrases were shorter than in phrases by jazz musicians (177.5 ms vs. 205.35 ms). Again, this could be affected by the average slower speech rate and musical tempo in the jazz samples and/or be purely accidental. More data are needed on this issue. Phrase final syllables and notes could play a perceptual role by marking phrase boundary and affect the overall pacing of the phrase. Reed (2020) excluded final feet from the data to focus on mid-phrase durational contrast. Here I chose to include final durations in the nPVI calculation. Both options should be compared over a larger set of data, to check whether final lengthening has a significant effect on speech and music nPVI data.

While greater variability in jazz musical phrases was expected, previous evidence suggested that an opposite tendency should be found in utterances by jazz musicians. In speech, I expected to find greater higher nPVI's in utterances by bluegrass musicians, as speaker of Southern AE. These predictions contradict the general hypothesis that similar patterns of variability should be found in speech and musical data by the same musicians. Surprisingly, the findings in my study are compatible with this hypothesis by showing greater variability in both speech utterances and musical phrases by jazz musicians. To explain these contradictory findings, the current corpus should be expanded to include more samples by more musicians of each style.

Finally, I believe that more quantitative measures are needed for comparing speech and musical data. Out of various measures that have been used for the study of speech rhythm (see section 2.3), only the nPVI has been used in comparison with music. Measures of vocalic and intervocalic intervals (e.g., %V, ΔC) are specific to speech and cannot be applied directly to musical notes. In section 3.3., I proposed that the musical BUR measure can provide more rhythmic data for comparison with speech. An interesting property of the BUR is that it takes into account both the surface acoustic durations of notes as well as their metrical prominence relations (downbeat vs. upbeat). Arguably, a comparable SUR (Stressed-Unstressed Ratio) measure can be theoretically devised for speech analysis, by dividing durations of adjacent

stressed and unstressed syllables/vowels. For this purpose, speech samples should be composed of metrically uniform strings of alternating strong-weak (or weak-strong) syllables, similar to the strings of consecutive eighth notes in my musical corpus. Such strings seem difficult to find in naturally occurring utterances, but perhaps can be artificially constructed in a controlled study. Alternatively, instead of analyzing full utterances and phrases, random pairs of stressed-unstressed syllables and downbeat-upbeat notes can be analyzed within comparable corpora of speech and musical data

5. Conclusion

This thesis is motivated by the speculation that there is some similarity in the manner in which musicians speak their native language and play their instrument. On the rhythmic domain, this idea can be explored by comparing durational nuances of speech and musical performance, generalized here under a common notion of micro-timing. It has been previously suggested that language and music could interface on a more underlying rhythmic dimension, which I labeled here as macro-timing. This rhythmic dimension involves the structural organization of sound in metrical grid representations, which could be a unique cognitive capacity for language and music (Jackendoff, 2009). In this paper, I suggested that rhythmic similarities could also exist in micro-timing patterns of speech and music performance, independently of their underlying metrical representation. To explore this possibility, I measured note durations in metrically uniform musical phrases from authentic recordings by jazz and bluegrass musicians, and compared these to durations of syllables in spontaneous speech utterances by the same musicians. Similar to previous studies, I found a correlated pattern of durational variability between these distinct musical styles. While the specifics of this comparison require further investigation over a larger set of data, a more general question is raised here about the nature of such a possible connection between the domains: **how can similar durational patterns be shared in speech and music performance?** I speculate that such similarity could emerge from a common interface of both domains with the articulatory-perceptual system. Both language and music involve highly complex processes of converting hierarchical metrical structure to linear signals of sound. These processes form an acquired rhythmic knowledge, which may not be grammatically encoded but could still be accessible in real-time performance. Musicians could possibly use this knowledge in a similar manner when speaking and playing their instruments.

References

- Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh University Press.
- Barlow, H., & Morgenstern, S. (1983). *A Dictionary of Musical Themes* (2nd ed.). Faber & Faber.
- Bartlett, S., Kondrak, G., & Cherry, C. (2009). On the Syllabification of Phonemes. *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 308–316.
- Beckman, M. E. (1992). Evidence for Speech Rhythms across Languages. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech Perception, Production and Linguistic Structure* (pp. 457–463). OHM Publishing Co.
- Benadon, F. (2006). Slicing the Beat: Jazz Eighth-Notes as Expressive Microrhythm. *Ethnomusicology*, 50(1), 73–98.
- Bertinetto, P. M. (1989). Reflections on the Dichotomy “Stress” vs. “Syllable-Timing.” *Revue de Phonétique Appliquée*, 91–93, 99–130.
- Boersma, P., & Weenink, D. (2021). *Praat: doing phonetics by computer [Computer program]* (6.1.43). www.praat.org
- Carpenter, A. C., & Levitt, A. G. (2016). Rhythm in the Speech and Music of Jazz and Riddim Musicians. *Music Perception*, 34(1), 94–103.
- Classe, A. (1939). *The Rhythm of English Prose*. Basil Blackwell.
- Clopper, C. G., & Pisoni, D. B. (2006). The Nationwide Speech Project: A new corpus of American English dialects. *Speech Communication*, 48(6), 633–644.
- Clopper, C. G., & Smiljanic, R. (2015). Regional Variation in Temporal Organization in American English. *Journal of Phonetics*, 49, 1–15.
- Couper-Kuhlen, E. (1993). *English Speech Rhythm: Form and Function in Everyday Verbal Interaction*. John Benjamins.
- Dasher, R., & Bolinger, D. (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association*, 12(2), 58–71.
- Dauer, R. M. (1983). Stress-Timing and Syllable-Timing Reanalyzed. *Journal of Phonetics*, 11(1), 51–62.
- Dauer, R. M. (1987). Phonetic and Phonological Components of Language Rhythm. *Proceedings of the 11th International Congress of Phonetic Sciences*, 5, 447–450.

- Dellwo, V. (2006). Rhythm and Speech Rate: A Variation Coefficient for ΔC . In P. Karnowski & I. Szigeti (Eds.), *Language and Language-Processing* (pp. 231–241). Peter Lang.
- Donnay, G. F., Rankin, S. K., Lopez-Gonzalez, M., Jiradejvong, P., & Limb, C. J. (2014). Neural substrates of interactive musical improvisation: An fMRI study of “trading fours” in jazz. *PLOS ONE*, 9(2).
- Friberg, A., & Sundström, A. (2002). Swing Ratios and Ensemble Timing in Jazz Performance: Evidence for a Common Rhythmic Pattern. *Music Perception*, 19(3), 333–349.
- Grabe, E., & Low, E. L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In C. Gussenhoven & N. Warner (Ed.), *Laboratory Phonology 7* (pp. 515–546). Mouton de Gruyter.
- Gridley, M. (1990). Clarifying Labels: Cool Jazz, West Coast and Hard Bop. *Journal of Popular Music Studies*, 2(2), 8–16.
- Heffner, C. C., & Slevc, L. R. (2015). Prosodic Structure as a Parallel to Musical Structure. *Frontiers in Psychology*, 6(1962).
- Huron, D. (1994). *The Humdrum Toolkit: Reference Manual*. Center for Computer Assisted Research in the Humanities.
- Huron, D., & Ollen, J. (2003). Agogic Contrast in French and English Themes: Further Support for Patel and Daniele (2003). *Music Perception*, 21(2), 267–271.
- Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception*, 26(3), 195–204.
- Jekiel, M. (2014). Comparing rhythm in speech and music: The case of English and Polish. *Yearbook of the Poznan Linguistic Meeting*, 1(1), 55–71.
- Jun, S.-A. (2014). Prosodic Typology: by Prominence Type, Word Prosody, and Macro-Rhythm. In *Prosodic Typology II* (pp. 520–539). Oxford University Press.
- Katz, J., & Pesetsky, D. (2011). *The Identity Thesis for Language and Music*. https://doi.org/https://www.researchgate.net/publication/279236407_The_Identity_Thesis_for_Language_and_Music
- Kenstowicz, M. (1994). *Phonology in Generative Grammar*. Blackwell.
- Labov, W., Ash, S., & Boberg, C. (2008). *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Walter de Gruyter.
- Ladefoged, P. (1967). *Three Areas of Experimental Phonetics*. Oxford University Press.
- Ladefoged, P. (1975). *A Course in Phonetics*. Harcourt Brace Jovanovich.
- Lehiste, I. (1977). Isochrony Reconsidered. *Journal of Phonetics*, 5, 253–263.

- Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press.
- Limb, C. J., & Braun, A. R. (2008). Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation. *PLoS ONE*, 3(2).
- Lloyd James, A. (1940). *Speech Signals in Telephony*. Sir I. Pitman & Sons.
- Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English. *Language and Speech*, 43(4), 377–401.
- McGowan, R. W., & Levitt, A. G. (2011). A Comparison of Rhythm in English Dialects and Music. *Music Perception*, 28(3), 307–313. <https://doi.org/10.1525/MP.2011.28.3.307>
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language Discrimination by Newborns: Toward an Understanding of the Role of Rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756–766.
- Nesbitt, M. (2018). Acoustic Correlates to Ambisyllabic Representations in American English. *University of Pennsylvania Working Papers in Linguistics*, 24(1), 16.
- Nespor, M. (1990). On the Rhythm Parameter in Phonology. In I. M. Roca (Ed.), *Logical Issues in Language Acquisition* (pp. 157–175). Foris.
- Nguyễn, T.-A. (2017). Comparing Rhythm of Speech and Folk Music: The Case of Vietnamese versus Australian, American and British English. *Australian Journal of Linguistics*, 37(4), 486–501.
- Owens, T. (1995). *Bebop: The Music and its Players*. Oxford University Press.
- Parker, C. (1978). *Charlie Parker Omnibook: For C Instruments*. Atlantic Music Corp.
- Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford University Press.
- Patel, A. D., & Daniele, J. R. (2003a). An Empirical Comparison of Rhythm in Language and Music. *Cognition*, 87(1), B35–B45.
- Patel, A. D., & Daniele, J. R. (2003b). Stress-Timed vs. Syllable-Timed Music? A Comment on Huron and Ollen (2003). *Music Perception*, 21(2), 273–276.
- Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the Rhythm and Melody of Speech and Music: The Case of British English and French. *The Journal of the Acoustical Society of America*, 119(5), 3034–3047.
- Pike, K. L. (1945). *The Intonation of American English*. University of Michigan Press.
- Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In *Proceedings of Speech Prosody 2002, Aix-en-Provence* (pp. 115–120). Aix-en-Provence: Laboratoire Parole et Langage.

- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of Linguistic Rhythm in the Speech Signal. *Cognition*, 73, 265–292.
- Reed, P. E. (2020). Prosodic Variation and Rootedness in Appalachian English. *University of Pennsylvania Working Papers in Linguistics*, 25(2), 13.
- Roach, P. (1982). On the Distinction between “Stress-Timed” and “Syllable-Timed” Languages. In D. Crystal (Ed.), *Linguistic Controversies* (pp. 73–79). Arnold.
- Schreuder, M., & Gilbers, D. (2004). The Influence of Speech Rate on the Perception of Rhythm Patterns. In M. Schreuder, D. Gilbers, & N. Knevel (Eds.), *On the Boundaries of Phonology and Phonetics* (pp. 183–201). University of Groningen.
- Selkirk, E. (1995). The Prosodic Structure of Function Words. *Papers in Optimality Theory*, 18, 439–469.
- Stetson, R. H. (1951). *Motor Phonetics: A Study of Speech Movements in Action* (2nd ed.). North Holland Publishing Co.
- Thomas, E. R., & Carter, P. M. (2006). Prosodic Rhythm and African American English. *English World-Wide*, 27(3), 331–355.
- Wagner, P., & Dellwo, V. (2004). Introducing YARD (Yet Another Rhythm Determination) and Re-Introducing Isochrony to Rhythm Research. *Proceedings of Speech Prosody 2004, Nara, Japan*.
- White, L., & Mattys, S. L. (2007). Calibrating Rhythm: First Language and Second language Studies. *Journal of Phonetics*, 35(4), 501–522.

Appendix A – The musicians

East Coast jazz saxophone players

- *John Coltrane* (1926, Hamlet, North Carolina – 1967, Huntington, New York)
- *Sonny Rollins* (1930, New York City, New York –)

West Coast jazz saxophone players

- *Paul Desmond* (1924, San Francisco, California – 1977, New York City, New York)
- *Warne Marsh* (1927, Los Angeles, California – 1987, Los Angeles California)

Bluegrass fiddle players

- *Kenneth (Kenny) Clayton Baker* (1926, Burdine, Kentucky – 2011, Gallatin, Tennessee)
- *Robert Russel ("Chubby") Wise* (1915, Lake City, Florida – 1996, Bowie, Maryland)

Appendix B – Sources of recordings

I. Speech recordings

John Coltrane

"John Coltrane Interview by Carl-Erik Lindgren" (Stockholm, 22/3/1960) – *Miles Davis & John Coltrane, The Final Tour: The Bootleg Series Vol. 6*, Legacy (2018).

Sonny Rollins

(i) "Jazz Casual": Sonny Rollins with Jim Hall, *NET* (23/3/1962).

<https://www.youtube.com/watch?v=a5dR0cHAYBQ>

(ii) Interview with Sonny Rollins, *Antenne 2* (1980).

<https://www.youtube.com/watch?v=YAhqUrEUHMo>

Paul Desmond

Paul Desmond interviews Charlie Parker (1954).

<https://bobreynoldsmusic.com/paul-desmond-charlie-parker/>

Warne Marsh

"Logiske Linjer", directed by Jan Horne, *NRK* (Trondheim, 1984).

<https://www.youtube.com/watch?v=-IzkHFIFrMk>

Kenny Baker

Interview with Kenny Baker by Josh Graves (24/6/2005), *Bluegrass Music Hall of Fame and Museum Oral History Project, Louie B. Nunn Center for Oral History, University of Kentucky Libraries*.

<https://kentuckyoralhistory.org/ark:/16417/xt7gth8bk46k>

Chubby Wise

(i) "'Orange Blossom Special' & Background (Oklahoma Public TV)"

<https://www.youtube.com/watch?v=WeFe7RGsbXk>

(ii) "Chubby Wise Obituary on TNN 1996"

<https://www.youtube.com/watch?v=fh6tM-d0NZQ>

II. Music recordings

Tracks are ordered by recording dates. Album release dates are given when recording dates are unknown.

John Coltrane

- *Lush Life*, Prestige (7188): "I Hear a Rhapsody" (5/1957).

- *Blue Trane*, Blue Note (1577): "Moment's Notice" (9/1957).

- *Kenny Burrell & John Coltrane*, New Jazz (8276): "Freight Trane" (03/1958).

Sonny Rollins

- *Bag's Groove*, Prestige 7109: "But Not for Me (Take 2)" (06/1954).

- *Saxophone Colossus*, Prestige 7079: "Strode Rode" (06/1956).

- *The Sound of Sonny*, Riverside 12-241: "Just in Time" (06/1957).

- *Sonny Rollins and the Contemporary Leaders*, Contemporary Records S7564: "I've Told Every Little Star" (10/1958).

Paul Desmond

- *Brubeck Time*, Culombia CL 622: "Jeepers Creepers", "Why Do I Love You?", "Stompin' for Mili", "Brother Can You Spare a Dime" (10/1954).

Warne Marsh

- *Lee Konitz and Warne Marsh*, Atlantic 1217: "Donna Lee" (06/1955)

- *Warne Marsh*, Atlantic 1291: "Yardbird Suite", "It's Alright with Me", "Excerpt" (01/1958).

Kenny Baker

- *A Baker's Dozen*, County Records 730: "Johnny The Blacksmith", "Sam's Tune" (09/1970).

Chubby Wise

- *Precious Memories*, Stoneway Records STY-112: "Do Lord, Remember Me", "This World Is Not My Home" (released on 1971).

- *Chubby Wise Plays Hank Williams*, Stoneway Records STY-169: "I Saw the Light" (released on 1977).

- *An American Original: The '94 Sessions*, Pinecastle Records PRC-1041: "Little Lisa Jane" (late 1994).

Appendix C – Data Analysis

I. Speech data

John Coltrane

1:05: "I'm trying so many things at one time you see"

aim traɪŋ so mæ.ni θɪŋz ə(d) wʌn taɪm jə si

1:38: "To take the one single line through 'em"

tə teɪk ðə wʌn sɪŋ.g(ə)l laɪn θru əm

2:15: "To produce a more beautiful sound"

tu prə.dus ə mɔr bju.rə.f(ə)l saʊn(d)

2:26: "That's what I mean by beautiful"

ðæt(s) wə (r)aɪ min baɪ bju.rɪ.f(ə)l

3:03: "It was a good recording"

ɪʔ wəz ə gʊd rɪ.kɔr.dɪŋ

| Utterance | Syl. # | Rate * | nPVI |
|-----------|--------|--------|-------|
| 1:05 | 11 | 4.24 | 65.75 |
| 1:38 | 9 | 4.76 | 55.82 |
| 2:15 | 9 | 5.10 | 60.26 |
| 2:26 | 8 | 6.38 | 39.31 |
| 3:03 | 7 | 4.82 | 61.58 |

Sonny Rollins

- Video #1:

7:50: "But within that form of course it's very free"

bʌ(d) wɪð.ɪn ðæt fɔ:m ʌv kɔ:s ɪts vɛ.ri fri

9:50: "Play it in different keys"

pleɪ ɪt ən dɪ.frənt ki:z

10:25: "I was quite interested in Coleman Hawkins"

aɪ wəz kwɑɪt ɪn.trə.stɪd ɪn kəʊl.mən hɔ:kɪnz

- Video #2:

0:09: "It's even more important than music"

ɪts i:vən mɔ:r əm.pɔ:r.dʌ(t) dən mju:zɪk

0:34: " Because I get a great deal of strength from meditation"

bɪ.kɔ:z aɪ geɪ.rə greɪt dil ə(v) streŋθ frʌ(m) mɛ.də.teɪ.ʃən

* Speech rate is measured by the number of syllables per second in the utterance.

| Utterance | Syl. # | Rate | nPVI |
|-----------|--------|------|-------|
| Video #1 | | | |
| 7:50 | 11 | 4.31 | 52.41 |
| 9:50 | 6 | 3.44 | 68.82 |
| 10:25 | 11 | 3.41 | 57.26 |
| Video #2 | | | |
| 0:09 | 10 | 5.24 | 40.21 |
| 0:34 | 14 | 5.2 | 44.81 |

Paul Desmond

0:51: "Yeah, how to play any horn"

jæ hæʊ rə pleɪ ɛ.ni hɔːn

1:51: "And you always do have a story to tell"

ən ju əl.wɪz du hæv ə stɔːri tu tel

2:13: "I always wondered about that too"

aɪ əl.wɪz wʌn.dərd ə.baʊt ðæt tu

2:17: "whether that came behind practicing"

wɛ.ðər ðət keɪm bə.haɪnd præk.tɪsɪŋ

2:49: "that's what I wondered"

ðæts wɒt aɪ wʌn.dərd

| Utterance | Syl. # | Rate | nPVI |
|-----------|--------|------|-------|
| 0:51 | 7 | 5.61 | 37.42 |
| 1:51 | 11 | 5.28 | 40.52 |
| 2:13 | 9 | 5.54 | 51.39 |
| 2:17 | 9 | 4.46 | 74.67 |
| 2:49 | 5 | 5.03 | 47.12 |

Warne Marsh

- 0:34: "You should know by memory"

ju ʃʊd noʊ baɪ məm.ri

- 0:41: "And tell me what it was"

ən tɛl mi wʌ.rət wəz

- 7:28: "All conventional harmony's built on the major and the minor scale"

ɔl kən.vən.ʃnəl hɑr.mə.niz bɪlt ʌn də meɪ.dʒər ən də maɪ.nər skeɪl

- 10:06: "Training your ear is a matter of singing the pitch"

treɪ.nɪŋ jər ɪr ɪz ə mə.tər əv sɪŋ.ɪŋ də pɪtʃ

- 19:10: "Or you rewrite the harmony your own way"

ɔr ju ri.raɪt ðə hɑr.mə.ni jər oʊn weɪ

| Utterance | Syl. # | Rate | nPVI |
|-----------|--------|------|-------|
| 0:34 | 6 | 5.07 | 47.69 |
| 0:41 | 6 | 3.7 | 48.66 |
| 7:28 | 17 | 5.31 | 65.58 |
| 10:06 | 13 | 4.43 | 57.18 |
| 19:10 | 11 | 3.95 | 58.41 |

Kenny Baker

- 5:08: "If I don't get that itinerary, I'm not going to Japan with you"

ɪf aɪ dɔŋ ɡet ðæt ə.tɪ.nə.re.ri a(ɪ)m nɒt ɡɔ.ɪŋ tə dʒə.pæn wɪð ju

- 5:18: "I worked with Bill two more weeks"

aɪ wɜrkt wɪð bɪl tu mɔr wɪks

- 6:58: "I didn't feel bad at all doin' that"

aɪ dɪ.dənt fi:l bæd æt.ɔl du.əŋ dət

- 8:14: "Bill n' me never had no problem"

bɪl ən mi ne.vər hæd noʊ prɒ.bləm

- 9:55: "They worked on that hand for 3 hours and 45 minutes"

ðei wɜrkt ɔn dæt hæn(d) f(ə) θri aʊrɪz ən fɔr.rə faɪv mi.nəts

| Utterance | Syl. # | Rate | nPVI |
|-----------|--------|------|-------|
| 5:08 | 19 | 5.17 | 28.84 |
| 5:18 | 7 | 3.76 | 33 |
| 6:58 | 10 | 4.4 | 48.96 |
| 8:14 | 9 | 5.26 | 32.93 |
| 9:55 | 14 | 4.19 | 38.55 |

Chubby Wise

- Video #1:

0:58: "I heard Bill say that he gon' loose his fiddler"

aɪ hɜrd bɪl seɪ dæt hi gɔn lʊz ɪz fɪd.lər

1:11: "I'm a fiddle player from Florida and I want that job"

a mə fɪ.dl pleɪ.ər frəm flɔ.rə.də ən aɪ wɔn dæt dʒɔb

1:27: "One of my favorites is footprints in the snow"

wʌn ə(v) maɪ feɪ.vrɪts ɪz fʊt.prɪnts ɪn də snəʊ

- Video #2:

1:00: "And I had to go check on my cab and go to work"

ən aɪ həd tə ɡoʊ tʃek ɔn maɪ kæb ən ɡoʊ.rə wɜrk

1:07: " You wanna go and get a copyright on it"

jə wɑ.nə ɡoʊ ən ɡe.rə kɑ.pə raɪ(d) ɔn ɪt

| Utterance | Syl. # | Rate | nPVI |
|-----------|--------|------|-------|
| Video #1 | | | |
| 0:58 | 11 | 6.13 | 29.5 |
| 1:11 | 15 | 6.59 | 46.74 |
| 1:27 | 11 | 5.51 | 48.08 |

| | | | |
|----------|----|------|-------|
| Video #2 | | | |
| 1:00 | 13 | 8.03 | 42.07 |
| 1:07 | 12 | 8.45 | 35.69 |

II. Musical data

John Coltrane

I Hear a Rhapsody (2:13)

A m7(b5) D 7(b9)

Musical notation for 'I Hear a Rhapsody' in 4/4 time, featuring a melodic line with a key signature of two flats (Bb and Eb). The notation includes a treble clef, a 4/4 time signature, and a key signature of two flats. The melody consists of quarter and eighth notes. Above the staff, the chords A m7(b5) and D 7(b9) are indicated.

Moment's Notice (1:05)

Ebm7 Ab7 Dbmaj7

Musical notation for 'Moment's Notice' in 4/4 time, featuring a melodic line with a key signature of three flats (Bb, Eb, and Ab). The notation includes a treble clef, a 4/4 time signature, and a key signature of three flats. The melody consists of quarter and eighth notes. Above the staff, the chords Ebm7, Ab7, and Dbmaj7 are indicated.

Moment's Notice (2:00)

Dm7(b5) G 7(b9) Cm7 Bbm7

Musical notation for 'Moment's Notice' in 4/4 time, featuring a melodic line with a key signature of three flats (Bb, Eb, and Ab). The notation includes a treble clef, a 4/4 time signature, and a key signature of three flats. The melody consists of quarter and eighth notes. Above the staff, the chords Dm7(b5), G 7(b9), Cm7, and Bbm7 are indicated.

Freight Trane (0:33)

Bbm7 Eb7

Musical notation for 'Freight Trane' in 4/4 time, featuring a melodic line with a key signature of three flats (Bb, Eb, and Ab). The notation includes a treble clef, a 4/4 time signature, and a key signature of three flats. The melody consists of quarter and eighth notes. Above the staff, the chords Bbm7 and Eb7 are indicated.

Freight Trane (2:00)

Bbm7 Eb7

Musical notation for 'Freight Trane' in 4/4 time, featuring a melodic line with a key signature of three flats (Bb, Eb, and Ab). The notation includes a treble clef, a 4/4 time signature, and a key signature of three flats. The melody consists of quarter and eighth notes. Above the staff, the chords Bbm7 and Eb7 are indicated.

| Phrase | Note # | bpm | nPVI |
|--------------------------|---------------|------------|-------------|
| I Hear a Rhapsody (2:13) | 9 | 206 | 13.83 |
| Moment's Notice (1:05) | 15 | 246 | 29.67 |
| Moment's Notice (2:00) | 18 | 246 | 14.37 |
| Freight Trane (0:33) | 10 | 231 | 23.59 |
| Freight Trane (2:00) | 11 | 233 | 20.56 |

Paul Desmond

Jeepers Creepers (0:48)

Cm7 F7 Bbmaj7

Musical notation for the first staff of 'Jeepers Creepers (0:48)'. It is in 4/4 time with a key signature of two flats (Bb and Eb). The melody consists of quarter notes: C4, D4, Eb4, F4, G4, Ab4, Bb4, C5, Bb4, Ab4, G4, F4, Eb4, D4, C4. Chord changes are indicated above the staff: Cm7 (under C4), F7 (under F4), and Bbmaj7 (under Bb4).

Jeepers Creepers (2:08)

Gm7 C7 F7

Musical notation for the second staff of 'Jeepers Creepers (2:08)'. It is in 4/4 time with a key signature of two flats. The melody starts with a quarter rest, followed by quarter notes: Bb3, Ab3, G3, F3, Eb3, D3, C3, Bb3, Ab3, G3, F3, Eb3, D3, C3. Chord changes are indicated above the staff: Gm7 (under G3), C7 (under C3), and F7 (under F3).

Why Do I Love You? (0:54)

Bbm7 Eb7

Musical notation for the third staff of 'Why Do I Love You? (0:54)'. It is in 4/4 time with a key signature of two flats. The melody consists of quarter notes: Bb3, Ab3, G3, F3, Eb3, D3, C3, Bb3, Ab3, G3, F3, Eb3, D3, C3. Chord changes are indicated above the staff: Bbm7 (under Bb3) and Eb7 (under Eb3).

Stompin' for Mili (0:22)

Bb Fm7

Musical notation for the fourth staff of 'Stompin' for Mili (0:22)'. It is in 4/4 time with a key signature of two flats. The melody consists of quarter notes: Bb3, Ab3, G3, F3, Eb3, D3, C3, Bb3, Ab3, G3, F3, Eb3, D3, C3. Chord changes are indicated above the staff: Bb (under Bb3) and Fm7 (under F3).

Brother Can You Spare a Dime (3:49)

Fm7 Bb7 Ebmaj7 Abmaj7

Musical notation for the fifth staff of 'Brother Can You Spare a Dime (3:49)'. It is in 4/4 time with a key signature of two flats. The melody consists of quarter notes: Bb3, Ab3, G3, F3, Eb3, D3, C3, Bb3, Ab3, G3, F3, Eb3, D3, C3. Chord changes are indicated above the staff: Fm7 (under F3), Bb7 (under Bb3), Ebmaj7 (under Eb3), and Abmaj7 (under Ab3).

| Phrase | Note # | bpm | nPVI |
|-------------------------------------|--------|-----|-------|
| Jeepers Creepers (0:48) | 10 | 222 | 13.27 |
| Jeepers Creepers (2:08) | 14 | 221 | 17.87 |
| Why Do I Love You? (0:54) | 13 | 249 | 40.02 |
| Stompin' for Mili (0:22) | 8 | 235 | 11.83 |
| Brother Can You Spare a Dime (3:49) | 15 | 159 | 27.28 |

Warne Marsh

Donna Lee (5:34)

Gm7(b5) C7(b9)

Musical notation for Donna Lee (5:34) in 4/4 time, key of B-flat major. The melody consists of quarter notes: B-flat, A, G, F, E, D, C, B-flat. Chords Gm7(b5) and C7(b9) are indicated above the staff.

Yardbird Suite (3:49)

Dm7 G7 C

Musical notation for Yardbird Suite (3:49) in 4/4 time, key of D major. The melody consists of quarter notes: D, E, F#, G, A, B, C, D. Chords Dm7, G7, and C are indicated above the staff.

It's Alright With Me (3:08)

Cm Dm7(b5) G7(b9)

Musical notation for It's Alright With Me (3:08) in 4/4 time, key of C minor. The melody consists of quarter notes: C, D, E-flat, F, G, A-flat, B-flat, C. Chords Cm, Dm7(b5), and G7(b9) are indicated above the staff.

It's Alright With Me (6:26)

F#m7 B7

Musical notation for It's Alright With Me (6:26) in 4/4 time, key of F# minor. The melody consists of quarter notes: F#, G, A, B, C, D, E, F#. Chords F#m7 and B7 are indicated above the staff.

Excerpt (1:44)

G

Musical notation for Excerpt (1:44) in 4/4 time, key of G major. The melody consists of quarter notes: G, A, B, C, D, E, F#, G. Chord G is indicated above the staff.

| Phrase | Note # | bpm | nPVI |
|-----------------------------|--------|-----|-------|
| Donna Lee (5:34) | 12 | 249 | 30.79 |
| Yardbird Suite (3:49) | 10 | 210 | 41.58 |
| It's Alright with Me (3:08) | 17 | 255 | 31.47 |
| It's Alright with Me (6:26) | 13 | 246 | 41.46 |
| Excerpt (1:44) | 23 | 222 | 32.82 |

Kenny Baker

Johnny The Blacksmith (1:06, 1:28)

A D E A

Sam's Tune (0:20)

G

Sam's Tune (0:53)

G

Sam's Tune (1:26)

G

| Phrase | Note # | bpm | nPVI |
|------------------------------|--------|-----|-------|
| Johnny the Blacksmith (1:06) | 16 | 276 | 8.28 |
| Johnny the Blacksmith (1:28) | 16 | 275 | 6.51 |
| Sam's Tune (0:20) | 10 | 229 | 7.56 |
| Sam's Tune (0:53) | 10 | 232 | 11.83 |
| Sam's Tune (1:26) | 10 | 233 | 13.71 |

Chubby Wise

Do Lord, Remember Me (2:31)

A



This World Is Not My Home (1:25)

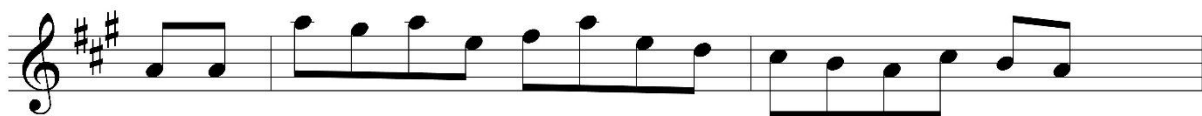
E7

A



I Saw The Light (12:54)

A



Little Lisa Jane (1:37, 1:41)

A



| Phrase | Note # | bpm | nPVI |
|----------------------------------|--------|-----|-------|
| Do Lord, Remember Me (2:31) | 7 | 202 | 23.17 |
| This World Is Not My Home (1:25) | 20 | 231 | 33.85 |
| I Saw The Light (0:42) | 16 | 267 | 7.45 |
| Little Lisa Jane (1:37) | 15 | 260 | 16.42 |
| Little Lisa Jane (1:41) | 15 | 262 | 12.29 |

BUR values by musicians:

| Coltrane | Rollins | Desmond | Marsh | Baker | Wise |
|----------|---------|---------|-------|-------|------|
| 0.77 | 1.86 | 0.89 | 1.08 | 0.86 | 1.71 |
| 1.51 | 0.72 | 1.44 | 2.04 | 1.11 | 0.86 |
| 1.14 | 0.76 | 1.02 | 1.21 | 1.12 | 1.02 |
| 1.05 | 0.91 | 0.79 | 1.22 | 1.05 | 0.74 |
| 0.87 | 2.13 | 0.93 | 1.46 | 1.11 | 0.71 |
| 0.98 | 1.12 | 1.11 | 1.68 | 1.06 | 0.78 |
| 1.03 | 0.90 | 1.28 | 0.80 | 0.96 | 0.97 |
| 1.14 | 0.80 | 1.14 | 2.03 | 0.95 | 0.72 |
| 0.70 | 1.35 | 1.20 | 1.43 | 1.09 | 0.62 |
| 0.91 | 1.05 | 1.21 | 0.62 | 1.04 | 1.39 |
| 0.72 | 1.23 | 1.11 | 1.37 | 1.08 | 1.03 |
| 0.69 | 1.34 | 1.70 | 1.15 | 1.10 | 0.58 |
| 1.05 | 1.97 | 1.80 | 1.23 | 1.16 | 0.54 |
| 0.82 | 2.52 | 1.05 | 1.39 | 1.01 | 1.05 |
| 1.34 | 0.96 | 1.10 | 1.08 | 0.98 | 1.04 |
| 1.76 | 1.48 | 1.62 | 2.03 | 0.94 | 1.12 |
| 1.36 | 1.04 | 1.44 | 1.44 | 1.31 | 0.99 |
| 1.04 | 0.99 | 1.41 | 0.70 | 1.05 | 0.93 |
| 0.97 | 1.13 | 1.06 | 2.76 | 0.99 | 1.19 |

| | | | | | |
|------|------|------|------|------|------|
| 1.02 | 1.35 | 1.04 | 2.02 | 1.02 | 0.99 |
| 1.03 | 0.80 | 1.03 | 0.93 | 1.06 | 1.05 |
| 0.79 | 1.07 | 1.18 | 0.64 | 1.08 | 0.98 |
| 0.95 | 0.99 | 1.23 | 1.37 | 0.80 | 1.23 |
| 0.78 | 1.14 | 1.06 | 1.42 | 0.96 | 1.23 |
| 0.56 | 1.37 | 1.34 | 1.38 | 0.98 | 0.96 |
| 1.20 | 1.70 | 1.36 | 0.84 | 1.15 | 0.96 |
| 1.27 | 0.71 | 1.67 | 1.04 | 1.17 | 1.22 |
| 1.44 | | 1.31 | 1.26 | 1.21 | 1.10 |
| 1.21 | | | 2.35 | 1.08 | 1.22 |
| 0.72 | | | 1.28 | 1.18 | 1.05 |
| | | | 0.89 | 1.09 | 0.96 |
| | | | 0.89 | | 1.26 |
| | | | 1.91 | | 1.00 |
| | | | 1.71 | | 1.06 |
| | | | 0.91 | | 0.95 |

תוכן עניינים

| | |
|----|---|
| 1 | מבוא |
| 6 | 2. חקר קצב הדיבור |
| 6 | 2.1. טיפולוגיה ריתמית והסיווג לשפות קצובות-טעם וקצובות-הברה |
| 8 | 2.2. התכונות הפונולוגיות של קבוצות ריתמיות |
| 10 | 2.3. מדדים אקוסטיים לקצב הדיבור |
| 10 | 2.3.1. רצפי תנועות ועיצורים |
| 12 | 2.3.2. גיוון משך ומדד ה-nPVI |
| 18 | 2.4. ההברה כיחידה ריתמית |
| 23 | 2.5. סיכום הפרק |
| 24 | 3. המחקר ההשוואתי של קצב לשוני ומוסיקלי |
| 24 | 3.1. ה-nPVI כמדד לגיוון מטרי |
| 24 | 3.1.1. Patel ו-Daniele (2003א) |
| 27 | 3.1.2. גורמים משפיעים |
| 29 | 3.1.3. מוסיקה עממית ומוסיקה קלאסית |
| 30 | 3.1.4. מבנה מטרי ותחושת קצב |
| 33 | 3.2. גיוון משך בביצוע לשוני ומוסיקלי |
| 34 | 3.2.1. McGowan ו-Levitt (2011) |
| 37 | 3.2.2. Carpenter ו-Levitt (2016) |
| 39 | 3.3. גיוון משך לעומת אחידות מטרי |
| 42 | 3.4. סיכום הפרק |
| 44 | 4. נגני ג'אז ובלוגראס כמקרה מבחן |
| 44 | 4.1. רקע רלוונטי על ג'אז ובלוגראס |
| 49 | 4.2. איסוף החומרים למחקר |
| 49 | 4.2.1. המוסיקאים |
| 50 | 4.2.2. קטעי ההקלטה |

| | |
|---------|---|
| 51..... | 4.2.2.1. שיקולים דיאכרוניים..... |
| 51..... | 4.2.2.2. אחידות מטריית..... |
| 54..... | 4.2.2.3. טמפו, שטף ואורך..... |
| 55..... | 4.3. סגמנטציה..... |
| 55..... | 4.3.1. ניתוח מבעי הדיבור..... |
| 55..... | 4.3.1.1. חלוקה להברות..... |
| 59..... | 4.3.1.2. ניתוח אקוסטי של מבעי הדיבור..... |
| 62..... | 4.3.2. ניתוח הפראזות המוסיקליות..... |
| 66..... | 4.4. תוצאות..... |
| 73..... | 4.5. דיון כללי..... |
| 78..... | 5. סיכום..... |
| 79..... | מקורות..... |
| 83..... | נספח א' - המוסיקאים..... |
| 83..... | נספח ב' - מקורות קטעי ההקלטה..... |
| 83..... | I. הקלטות הדיבור..... |
| 84..... | II. הקלטות המוסיקה..... |
| 85..... | נספח ג' - ניתוח הנתונים..... |
| 85..... | I. נתונים לשוניים..... |
| 90..... | II. נתונים מוסיקליים..... |

תקציר

עבודה זו בוחנת את הקשרים שבין קצב הדיבור לקצב מוסיקלי ע"י השוואה בין תכונות ריתמיות דומות במבעים לשוניים ופראזות מוסיקליות המופקים ע"י מוסיקאים. בעקבות Patel ו-Daniele (2003א), מחקרים השוואתיים בין שפה ומוסיקה השתמשו במדד ה-normalized Pairwise Variability Index (nPVI) לכימות של מידת הגיוון הריתמי בשפות וסגנונות מוסיקליים שונים. מדידת ערכי nPVI בדיבור מראה כי שפות כגון אנגלית והולנדית, שסווגו באופן מסורתי כשפות "קצובות-טעם" (stress-timed), מתאפיינות בגיוון רב יותר (ערכי nPVI גבוהים יותר) של משכי תנועות לאורך מבעים, מאשר בשפות שסווגו כשפות "קצובות-הברה" (syllable-timed), כגון צרפתית וספרדית (Grabe ו-Low, 2002; Ramus, 2002). Patel ו-Daniele השתמשו ב-nPVI כדי למדוד את מידת הגיוון הריתמי של משכי תווים במנגינות מאת מלחינים דוברי אנגלית וצרפתית, ומצאו דפוס דומה עם ערכי nPVI גבוהים יותר במנגינות מאת מלחינים דוברי אנגלית. דפוס דומה נמצא גם במחקרים הבוחנים ביצוע ספונטני של מבעי דיבור ופראזות מוסיקליות ע"י מוסיקאים, הנבדלים בדיאלקטים שונים של אנגלית אותם הם דוברים ובסגנונות מוסיקליים שונים בהם הם מתמחים (McGowan ו-Levitt, 2011; Carpenter ו-Levitt, 2016).

בעבודה זו אני בוחן בצורה דומה דפוסים ריתמיים בדיבור ובנגינה של מוסיקאים משני סגנונות שונים במוסיקה האמריקאית – ג'אז ובלוגראס. בשונה ממחקרים קודמים, חישבתי ערכי nPVI לשוניים לפי משכי הברות ולא לפי משכי תנועות, מתוך תפישה שההברה היא יחידת קצב הדיבור הבסיסית והמקבילה ביותר לצלילים מוסיקליים (Patel, 2008). לשם כך נעזרתי בקריטריונים פונולוגיים המשמשים במודלים של חלוקה אוטומטית להברות (Bartlett ועמיתים, 2009). בתחום המוסיקלי אני מציע שיש להבחין בין דפוסים ריתמיים הנגזרים ממבני העומק המטריים של פראזות מוסיקליות, וניתנים לייצוג בכתב תווים, לבין ניואנסים של משכי צליל בזמן ביצוע. בעבודה זו התמקדתי בדפוסים הריתמיים מהסוג השני. לצורך זה, הגבלתי את הקורפוס המוסיקלי לפראזות המורכבות מערכים מטריים זהים של תווי שמיניות, בהן הגיוון הריתמי היחיד נובע מהבדלים בין אורכי הצלילים המופקים בפועל ע"י המוסיקאים. אני מציע כי מתודה דומה ניתנת ליישום גם בתחום הלשוני. בדומה לממצאים קודמים, התוצאות הראשיות של המחקר מראות דפוס תואם של ערכי nPVI גבוהים יותר אצל נגני ג'אז לעומת נגני בלוגראס, הן בדיבור והן במוסיקה. העבודה מרחיבה

את הדיון בקשרים האפשריים בין קצב לשוני לקצב מוסיקלי. היא מדגימה כיצד קשרים כאלו יכולים להתקיים בשתי מערכות ידע שונות – מבני עומק מטריים וניואנסים של משכי צלילים בשלב הביצוע, ומציעה מתודות ומדדים נוספים לחקר הקשרים הללו.

אוניברסיטת תל אביב

הפקולטה למדעי הרוח ע"ש לסטר וסאלי אנטין

החוג לבלשנות

דמיון ריתמי בין שפה ומוסיקה:

נגני ג'אז ובלוגראס כמקרה מבחן

חיבור זה הוגש כעבודת גמר לתואר "מוסמך אוניברסיטה" -

M.A. באוניברסיטת תל-אביב

על ידי

אודי ורזגר

תחת הדרכתו של

ד"ר אוון גרי-כהן

יולי, 2021