Patterns in Speech and Music:

The technical rationale of slope variability.

Universiteit Leiden

2021

Panajotis Papa

Master of Arts : Thesis submitted in partial fulfillment of the requirements for the degree of THEORETICAL AND EXPERIMENTAL LINGUISTICS, Leiden University.

Supervisor: Professor Yiya Chen, Leiden University, Faculty of Humanities, Leiden University Centre for Linguistics

Dr. D.j. van de Velde, Leiden University, Faculty of Humanities, Leiden University Centre for Linguistics

Date of Submission: 17/11/2021

Contents

Equations and important terms 5

Abstract 7

1 Introduction 8

2 Literature Review and Research Question 9

3 Background Knowledge 13

3.1 Rhythm 13

3.2 Melody 14

3.3 Slope Variability 15

4 Methods 16

4.1 nPVI 16

4.2 Prosogram 18

5 Materials 20

5.1 Speech 20

5.2 Music 21

5.3 Rhythm coding for English and French 22

5.3.1 Speech Rhythm for English and French 23

5.3.2 Music Rhythm for English and French 24

5.4 Melody coding 25

5.4.1 Speech melody 25

5.4.2 Music melody 27

5.5 Slope Variability 28

5.5.1 Speech Slope 29

5.5.2 Music Slope 29

6 Results 30

6.1 Rhythm 30

6.1.1 Speech 30
6.1.2 Music Rhythm 35
6.2 Melody 38
6.2.1 Speech 38
6.2.2 Music 38
6.2.3 Slope Results 40
7 Discussion and Conclusion 45
7.1 Discussion on Rhythm 45
7.2 Discussion on Melody 46
7.3 Discussion in Slope 47
7.4 Conclusion 48
8 Acknowledgments 50
A Appendix: ENGLISH AND FRENCH SENTENCES 51
B Appendix: COMPOSERS AND MUSICAL THEMES 52
C Appendix: SLOPE CALCULATION: AN EXAMPLE 53
References 54

List of Figures

Figure 1: *An example of nPVI calculation* *18*
Figure 2: *Oscillogram and Spectrogram* *19*
Figure 3: *Illustration of the prosogram* *20*
Figure 4: *Hanning window* *23*
Figure 5: *Gaussian window* *24*
Figure 6: *An example of durational coding* *25*
Figure 7: *Illustrates an example of pitch coding* *27*

Figure 8: *English speech* ... *31*
Figure 9: *French speech* ... *32*
Figure 10: *Boxplots and Histograms for English and French nPVI* ... *33*
Figure 11: *Boxplots and histograms for English and French CV* ... *34*
Figure 12: *MonteCarlo Speech* ... *35*
Figure 13: *nPVI and CV plot for English Music* ... *36*
Figure 14: *nPVI and CV plot for French Music* ... *37*
Figure 15: *MonteCarlo Music* ... *37*
Figure 16: *Pitch Interval Variability* ... *39*
Figure 17: *nPVI values for English and French* ... *40*
Figure 18: *Boxplots and Histograms of Slope in English and French Speech* ... *42*
Figure 19: *Boxplots and Histograms of Slope in English and French Music* ... *43*
Figure 20: *Slope Interval Variability* ... *43*
Figure 21: *A combination of CV slope, CV pitch interval, nPVI* ... *44*

List of Tables

Table 1: *Basic descriptive statistics on the sentences studied* ... *21*
Table 2: *nPVI and CV values for Speech* ... *30*
Table 3: *nPVI and CV values for Music* ... *35*
Table 4: *Pitch height and Pitch intervals CV values for Speech* ... *38*
Table 5: *Pitch height and Pitch intervals CV values for Music* ... *39*
Table 6: *CV(s.e) Values for English/French Speech and Music Slope* ... *41*

Equations and important terms

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

Grabe and Low (2002) define the "normalized Pairwise Variability Index" (nPVI) for a rhythm with adjacent inter-onset intervals (IOI). Where ***m*** is the number of adjacent vocalic intervals in an utterance, and ***dk*** is the duration of the ***kth*** interval.

2) Monte carlo

Monte Carlo is a computational algorithm that rely on randomly repeated sampling to obtain numerical results. The basic idea of this method is based on the concept of using randomness to solve problems that might be deterministic in principle. Monte Carlo methods are often used in three problem classes: numerical integration, generating draws from a probability distribution and optimization.

3) Pitch interval:

In musical set theory, a pitch interval (PI or ip) is the number of semitones that separates one pitch from another, upward or downward.

4) Pitch height:

The term that describes the perceptual 'highness' or 'lowness' of a pitch and it is related to frequency.

5) CV:

It shows the extent of variability in relation to the mean of the population. The coefficient of variation (CV) is defined as the ratio of the standard deviation σ to the mean μ. CV = σ/μ

6) Glissandos threshold

$$G = 0.32/T^2$$

To determine whether a given vowel should be assigned a level tone or a glide, a glide threshold of $0.32/T^2$ semitones/s was used, where **T** is the duration of a vowel in **s**.

7) Musical notation:

In music theory, musical notation is a series of symbols and markings that inform musicians how to perform a composition.

Abstract

Kirkpatrick claimed that French keyboard music sounded like French. Hall (1953) talked about a resemblance of Elgar's Music to the intonation of British speech. Both scholars studied the similarity of instrumental music and speech, a topic, that has remained controversial. This thesis investigates the robustness of the metrics used for the two experiments in rhythm and melody to provide information about their linguistic origins. To this end, this thesis replicates the two experiments of Patel et al. (2006) regarding the rhythm and melody of British English and French speech and instrumental music. In addition, a third parameter labelled as "slope" was tested to determine whether melody, when taking into consideration the durational property of the melodic interval, provides the same results as the two previous replicated experiments. We replicated the findings for rhythm and melody reported in Patel et al. (2006). The slope variability parameter, however, showed a reverse pattern raising critical questions about the validity of the metrics as a qualified method to portray statistical evidence of national characteristics in speech and music.

1 Introduction

Language and music make our species unique. These two terms are important and play an essential role in how societies are organized. It is a vital humxn need to use, produce, and perceive these multiplex and meaningful sounds (Patel, 2007). The trend of comparing language and music is not new. Philosophers and scientists tried to understand, examine, and define the relationship between language and music (Plato, Darwin, Rousseau and many more). In Ancient Greece, Mouses was the goddess of music, who later became the goddess of poetry as well, portraying the close relationship of language and music. Rousseau (1781) assumed that a common root for language and music existed and that language separated from this root out of humxn necessity for societal organisation. He also maintained that the first languages were sung. Later, Darwin (1871) said that "primeval man, or rather some early progenitor of man, probably first used his voice in producing true musical cadences, that is in singing, as do some of the gibbon-apes at the present day", illustrating how language and music have been the root of our species way of producing speech. He also claimed that language came first. In a later study, Besson and Schon (2001) emphasize, "in both language and music, emotional excitement is expressed through fast, accelerating, and high-registered sound patterns." In a similar vein, Jackendoff (2009) stated that language and music could be combined in the form of songs in every culture. The formers highlight the similarity of sound patterns in language and music when expressing emotions. The latter went further by pointing out the "fact" that language and music can be combined and form songs, something which he claimed is universal. It is essential to examine both similarities and differences these two domains (language and music) have. While language is not interchangeable with speech, it can hold out as spoken or signed. In both cases, spoken or signed, language comprises of basic but rather meaningless patterns of phonemes or handshapes. In speech, for example, when phonemes are joined, they can form basic units such as words. These units constitute the lexicon. When connected, these words can create phrases that can be used to form larger structures. At this point, it is important to underline the unique skill of humxn to acquire and produce language. Additionally, humxns are attracted to music and use it in several contexts and conditions (entertainment, religious or cultural ceremonies, art and many more). While it is challenging to use a definition for music, it is vital to see "musicality" as a biological outcome from numerous "musics", each of which can be described as a cultural phenomenon derived from that very biology (Honing, 2010). Music, like language, is dynamic and evolves. Based on Darwins' point of view, music provides a historical record of the foundation of humxn communication. It is not easy to determine the properties or functions language and music use. Language uses lexicon, grammar and semantics to form meaningful sequences. Music is more abstract and has no need to produce expressively meaningful sentences. Both language and music have rhythm and melody and we, as humxns, tend to intermingle language and music to strengthen our memory (Jackendoff, 2009). Many people use music in sentences when they want to remember words (Fennell et al., 2021). Music, like language,has a vital role in defining culture. In 1900 in Greece, for instance, rembetika (occasionally transliterated as rembetiko or rebetico) -which can be described as the urban songs of the Greeks, especially the poorest- is a kind of music which, with its lyrics and sounds, depicts cultural characteristics. Rebetiko was added to the UNESCO Intangible Cultural Heritage Lists in 2017.

Several studies in music and language acquisition in infants investigated whether music is a by-product of our language, finding similarities and differences (Trehub, 2003; Koelsch et al., 2003). This connection can be exemplified by songs for toddlers that tend to be fast, high with exaggerated rhythmic accents, or lullabies that are slow and low. Nevertheless, language does not typically have rhythm structures that can be used to actively compare it with music. Several approaches involving linguistics, cognitive psychology and cognitive neuroscience used to draw parallels between the two. Thus, by comparing results obtained from various disciplines, will provide insights of the existence or absence of cross-domain traits.

2 Literature Review and Research Question

Patel, Iversen and Rosenberg (2006) investigated the claim that a composer's music reflects prosodic patterns in her or his native language. By the same token, Kirkpatrick and Hall talked about the convergence of music and language by analysing instrumental music. It is of great interest to investigate whether speech patterns are mirrored in instrumental music, which, as Patel et al. (2006) describe, is much more controversial. Patel et al. (2006) in their paper compared rhythm and melody in speech and music from England and France. They concluded that music reflects patterns of duration and pitch interval variability in speech.

This thesis will replicate both experiments on rhythm (ratio of durations) and melody (pitch height and pitch interval) of the same data set, using additional and updated tools to test if the same results can be observed. For the experiment on rhythm, the vowel boundaries will be marked from scratch using one additional tool for the analysis (see section 5.3.1). The same approach will be used for music rhythm as in Patel et al. (2006). Furthermore, one program (Praat) will be used for both rhythm and melody experiments instead of using two different ones (SIGNAL and Praat) as in Patel et al. (2006) (see section 5.3.1). In the experiment on melody, we will use the updated version of the Prosogram program (3.00f instead of 1.3.6 used in Patel et al, (2006)) for both language and music. This thesis will test the technical rationale of the slope. It will investigate the parameter of their slope variability (rhythm and melody). Patel et al. (2006), studied variability in consecutive note intervals in melodies, more specifically on the basis of pitch height and pitch interval. The approach by Patel et al. (2006) is not taking into account the durational property of the melodic interval. This thesis, therefore, repeats the study by Patel et al. (2006), but instead of using pitch height and pitch interval, it treats slope as the dependent variable. The results will be compared to those by Patel et al. (2006).

As Burns (1999) argues, comparing pitch-interval sizes alone is an artificial task with no musical or ecological validity. Based on this approach, we assume that pitch interval is one factor to understand melody, but it has its limitations raising some questions on the approach performed by Patel et al. (2006). The study by Monahan et al. (1985) formulated the idea to test this extra technical rationale of slope variability. In their paper, they tested pitch and durations as two determinants of what they called “musical space”. They tested four melodies which were played in four rhythmic patterns. They found that in order to impute to these melodies a given perceptual space, several dimensions were needed. Most importantly, they found that the major dimensions were rhythmic. As both determinants (pitch and durations) were needed to categorize melodies, we decided to take this approach one step further and test this extra technical rationale (slope variability) that was not investigated by Patel et al. (2006). Taking into account the durations in addition to the pitch interval will update our view on consecutive variability of the tested languages. On the one hand, on the occasion that the results based on note intervals in melodies, as done by Patel et al. (2006), show a trend for higher English -than French- variability and the results based on slope variability, as done in the current study, show the same trend, then the methodology by Patel et al. (2006) will be probably robust. It will also mean that the durations of the pitch interval do not provide any additional information in terms of variability which pitch interval cannot describe alone. On the other hand, if the latter results show a different trend then this will raise questions on methodology and question the capacity of pitch interval as a tool to describe the variability in melody. In other words, if taking into account the durations of the pitch interval results have an opposite/different trend, this will show that the technical rationale that includes durations of the pitch interval has an effect on variability in these two tested languages. It will also challenge the robustness of the methodology by Patel et al. (2006).

This thesis will pose new questions and contribute to the discussion on comparing language and music in terms of their melody. The aim of this thesis is to show whether and how the durations of the pitch interval affect variability. This will be investigated by comparing the results of slope variability with the ones of pitch interval.

In speech there are consonants between vowels. The complexity of the consonant cluster differs between English and French. In French, longer consonants occur at the end of the stimuli or sentence but also consonants in French can be shorter or longer within clusters based on the difficulty of articulation (O'Shaughnessy, 1980).

We expect that the variability in speech as found by Patel et al. (2006), to change in the opposite direction when measuring slope variability instead of pitch interval, resulting to larger variability in French speech than in English. We believe that a change in absolute numbers will be visible due to consonants in-between vowels (which we are measuring for this experiments) that are affecting the interval of the measured vowels timewise. The reason to believe that French will have a larger variability than English speech is based on the findings by O'Shaugnessy (1980) when he found that in French, consonant clusters in final position (after vowel) are longer with a range of 25% to 75%. In English, Kluender et al. (1988) discussed that "a language community may tend to shorten vowels before phonemically long or geminate consonants and to lengthen them in front of phonemically short consonants". With this in mind we hypothesise that English speech will have smaller variability than French when testing their slope variability. Additionally, we hypothesize that when measuring the slope variability in musical themes for both English and French music, the same trend will appear as in the findings by Patel et al. (2006) when measuring pitch intervals. This hypothesis is based on the fact that in musical themes there are no consonants, or in simple terms, musical notes that are intervening in the measurements. To make it simple, in speech we measure vowels but consonants exist in the stimuli whereas in music we measure all the musical notes of each musical theme.

The overview of this thesis is as follows: Section 3 introduces the background knowledge and presents the core issues and information which this study investigates. It also provides a profound explanation about the methods and approaches this study puts forward. Section 3 is split into three subsections. In detail, Section 3.1 presents rhythm and provides an insight into how rhythm is accounted for speech and music whereas Section 3.2 deals with the melody part. Section 3.2 also explains the difficulty of defining melody in one term for both tested domains (speech and music). Next, in Section 3.3, the slope variability approach is introduced. It also addresses with the main hypothesis of this thesis, that if ratio of durations (rhythm) and pitch interval (melody) have the same trend as observed in the study by Patel et. al. (2006), it is worth testing if a combination (slope variability) of rhythmic and prosodic variability provides the same trend. The results of the slope variability analysis are two fold. They can either confirm the robustness of the results of the two previous experiments or they can question the metrics' ability to provide information about the linguistic origin of the two tested languages. Section 4.1 covers with informative details about the nPVI equation, which is employed for quantifying the rhythmic variability in speech and music. In turn, Section 4.2 proposes the Prosogram program used to transform the pitch stimuli of speech to music like pitch levels for better comparability to music, as used in Patel et al. (2006). In Section 5, materials are described similarly with the coding (5.3, 5.4 and 5.5) parts of the experiments that this study performed. Subsections 5.3 and 5.4 focus on replicating the results of the experiments done by Patel et al. (2006) that are needed to compare with the ones of the main study (slope) that is introduced and explained in subsection 5.5. Section 6 includes the analysis and describes the results for the experiments. Finally, Section 7 is divided into four subsections. Subsection 7.1 deals with a discussion on rhythm, the nPVI metric and the future aspects of this quantitative method. Next, subsection 7.2 discusses melody when subsection 7.3 focuses on slope analysis. Finally, subsection 7.4 summarises this study's conclusions.

3 Background Knowledge

3.1 Rhythm

The term “rhythm” (Greek rhythmos) is used in several contexts. It is essential to highlight that there is still no universally accepted definition of rhythm. Notwithstanding language and music, “rhythm” is used to describe dancing, paintings, oscillations in the brain and even rhythmic calls of certain animals. The similarity of these contexts is that rhythm implies periodicity. In simple terms, periodicity is a pattern that regularly repeats in time. Patel (2007) in his book “Music, Language and the Brain”, offers an essential distinction between these two terms, demonstrating that although all periodic patterns are rhythmic, not all rhythmic patterns are periodic. Hence, rhythm should not be aligned to periodicity. Rhythm plays an indispensable role in how we understand speech and as listeners, in order to comprehend spoken language, we must perceive the temporal organization of phonemes, syllables, words, and phrases from an ongoing speech stream (Kotz and Schwartze, 2010; Patel, 2011). A distinction is important. Although, rhythm in poetry has been studied since Ancient Greek texts, the study of rhythm in spoken language is a relatively recent subject of research in linguistics. In that light, three approaches are mainly considered when studying rhythm in speech (typological, theoretical and perceptual). This study is primarily interested in the typological approach, which is focused on understanding similar or different rhythmic patterns across languages and this study follows Patel’s (2007: 96) approach that rhythm is “the systematic patterning of sound in terms of timing, accent, and grouping”. Lloyd James (1940) addressed two categories of temporal regularity that are evident in speech. He referred to English as “machine gun” and French as “Morse code” styles. Pike (1945) renamed these two categories as syllable-timed and stress timed speech, respectively. It must be remarked that both could be found within the speech of an individual, and Cummins (2012) provides an empirical example from Martin Luther King’s “I have a dream” speech. Additionally, a third category was added with mora-timing of

Japanese to be included as one, which has been studied in several publications. Thus rhythm, can be understood as systematic in patterns of accent or timing, even if there is no regular beat. It is similarly challenging to describe rhythm in music using one term, as components of accent, metre and tempo are considered by many researchers. This study understands musical rhythm as "a regularly timed beat, a perceptually isochronous pulse to which one can synchronize with periodic movements such as taps or footfalls" (Patel, 2007).

3.2 Melody

Similarly to rhythm, the term melody has been used widely in both language and music. There is no definition that can accurately describe both speech and music melody. Ringer (2001) describes melody as "pitched sounds arranged in musical time in accordance with given cultural conventions and constraints", shifting the focus on musical time, thus not accurately describing speech melody. This study uses the definition of melody by Patel (2007: 182) who maintain that it is "an organized sequence of pitches that conveys a wide variety of information to a listener". However, it is essential to highlight that there are important differences between these two domains. An important difference is that musical melodies are structured within or around a stable set of intervals whereas linguistic melodies are not. In prosody, it is common to refer to pitch properties of speech as speech melody (Bolinger, 1989; Ladd, 2008). Prosody is a set of speech parameters that apply across individual speech sounds (e.g., at the level of the syllable or sentence), including fundamental frequency, rhythm, and stress. The most critical issue in testing or even comparing melody in speech and music is representing speech melodies. One way to do so is to use the raw f0 contours of sentences. Another way is to use the ToBI (Tone and Break Index) system of prosodic notation. ToBI builds on assumptions made by the Autosegmental metrical theory of intonation when phonologically events are marked with H for High and L for Low (Pierrehumbert, 1980; Ladd, 1997). In this study, the prosogram program is used, providing a representation of intonation as perceived by listeners. Moreover, described in the study by Patel et al. (2006), it follows perceptual principles regarding the stylization of f0. The ecological validity of this method has been tested in several experiments in linguistics, phonetics, and psychology (Patel et al., 2007; Sanchez et al., 2015; Ballier et al., 2021). This thesis explains the prosogram program and its use while it explores guidelines and principles on formulating the "image" of the way intonation looks in a listeners' brain. Findings by Magdics (1963) explain that even if speech and music have a common origin, they are developed separately as two forms that explain nature and purpose with dissimilar expressions. As in the study by Patel et al. (2006), both pitch height and pitch interval were calculated for this thesis to compare speech and music.

3.3 Slope Variability

Different researchers have studied duration and pitch as characteristics of both language and music (Aalto et al., 2013; Thompson, 1994). In the study by Patel et al. (2006), rhythm as ratio of durations and melody as pitch heights and pitch intervals were studied to investigate and compare language and music. Additionally, both rhythm and melody for speech and music are tested to investigate if their juxtaposition can provide patterns that depicts national origin. Both rhythm and melody, as described in the previous sections, can not be fully separated into two categories given that they "work" together to "create" a sentence or a musical theme respectively. Cumming (2010) investigated the interdependence of f0 and duration as cues and the influence of dynamic f0 on the perception of duration. He found that a rising f0 and an increased duration are interdependent cues to rhythmic groups. Moreover, Burns (1999) argued that, as previously explained, comparing pitch intervals alone does not produce definitive answers. Alto et al. (2013) supported that "the fundamental frequency of a complex sound modulates the perceived duration of a sound." All these previous works raise a critical question regarding the present study's effort to compare speech and music using the technical rationale of the slope variability. In Patel et al. (2006), the ratio of durations and pitch interval and pitch height were understood as characteristics for quantifying rhythm and melody. This thesis will test whether their combination (slope variability) can produce the same trend as in the study by Patel et al. (2006).

Specifically, each pitch interval was divided by the duration of the initial vowel/note within the tested pair for speech and music (see Section 3.5 and Appendix C for an example of this computation). The importance of this experiment is two-fold. On the one hand, if the same result is observed, the metrics used to investigate the national origin in the two replicated experiments as in Patel et al. (2006) will prove to be robust enough to provide this information and the extra technical rationale (slope variability) will have verified the validity of this approach. On the other hand, if the results are not aligned with the findings of the two previous replicated experiments, then essential questions are raised. A different trend will challenge the pitch interval's variability effect as a technical tool on comparing languages. Moreover, a non-aligned or even a reverse trend is not evidence of absence of features that can attribute national origins. Yet, it does challenge the metrics' ability to categorize languages, raising questions about their capability of providing such information.

4 Methods

4.1 nPVI

The normalized pairwise variability or nPVI (Low and Grabe, 2002) is a measure of the variability of successive syllabic durations in spoken language based on vowel length. In other words, it measures the contrast between successive rhythmic events. The difference with variability is that in nPVI, the consecutive order of elements matters while in variability it does not. nPVI was used by Low and Grabe as a tool to measure rhythmic variability of linguistic utterances (Low and Grabe, 2002). They argued that the nPVI metric could be used as a tool to identify differences between languages but also place them in groups, such as stress-timed (English, German) and syllable-timed languages (French, Italian). Moreover, the nPVI was used in music research, quantifying variation between nations/cultures, eras, and composers (Daniele and Patel, 2015; Hanson, 2017; McGowan Levitt, 2011; Patel and Daniele, 2003; Patel and Daniele, 2013; Sadakata, Desain, Honing, Patel, Iversen, 2004; VanHandel and Song, 2010). Additionally, nPVI in this study, as in Patel et al. (2006), was used as a measure of durational contrast between neighbouring vowels. Expecting to have a high nPVI when the contrast between neighbouring durations of vowels was large and low nPVI when the contrast is small while it remains sensitive to duration ratio. To compute the nPVI, we repeat three steps in a group of pairs of durations, as shown in Figure 1. At first, we take the absolute difference of a pair. Then, we take the average of the two numbers, and lastly, we divide the absolute value with the average value of the same pair. The computation is performed for all tested pairs in a sequence, and it will result in nPVI scores for all the tested pairs. Lastly, the mean of these values is calculated and then is multiplied by 100. The sum is the nPVI value for this sequence. It is essential to underline that nPVI differs from variability while it is sensitive to the order of durations when this is not the case for variability (e.g. standard deviation).

Thus, the equation for nPVI is written as follows:

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

The nPVI metric was used in the study by Patel et al. (2006) to compare English and French speech, while this equation can be applied to any sequence of numbers. nPVI was used to music notation because it represents music as a sequence of numbers. It was used to examine whether vowels in English express higher durational variability than vowels in the French language and whether a similarity applies to music. As in the study by Patel et al. (2006), a measure of overall variability for each sentence and musical theme is computed to investigate the relationship between variability and nPVI. In addition, as Patel et al's study (2006), the Monte Carlo method was used to quantify the likelihood of observing an nPVI difference of a given magnitude between two languages or two musics given existing differences in variability.

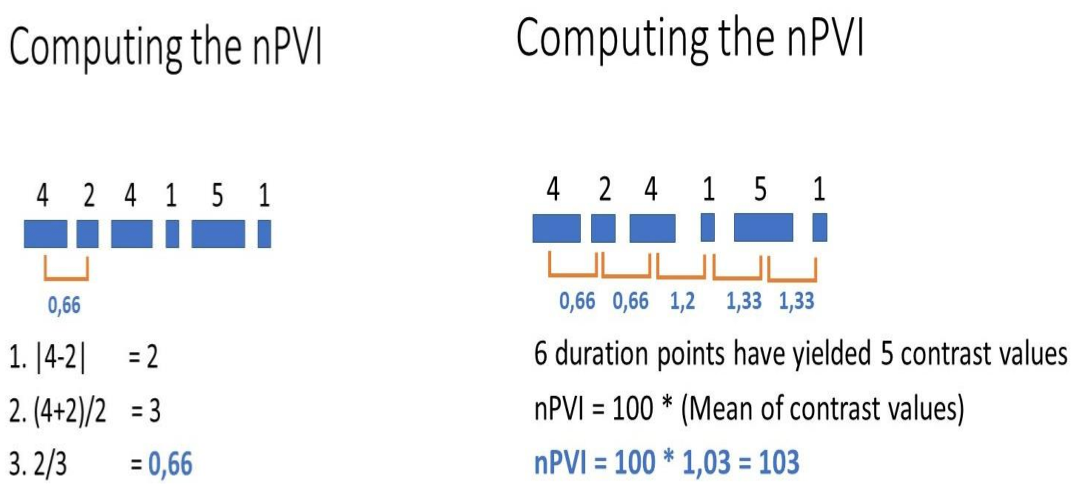

Figure 1: *An example of nPVI calculation*

4.2 Prosogram

While for rhythm the nPVI metric can be used, for melody, a big issue occurs when comparing speech and music regarding the manner for representing speech melodies (raw f0 contours, autosegmental metrical theories). In this study, prosogram was used similarly to the study by Patel et al. (2006). The prosogram program computes a representation of speech prosody's acoustic and perceived parameters, with a focus on the pitch. Analogously, the oscillogram and spectrogram (see Figure 2) show the evolution of the waveform and spectrum in time, a Prosogram does so for speech prosody.

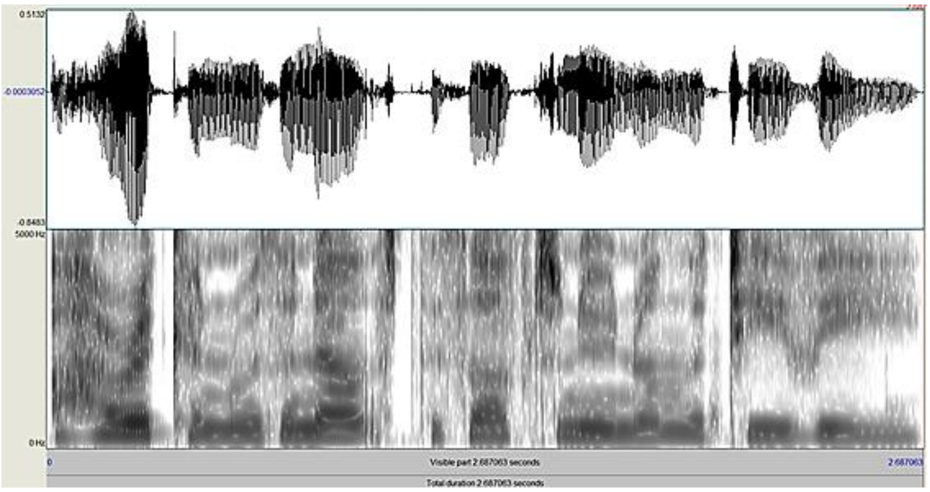

Figure 2: *Oscillogram (top) and Spectrogram (bottom) of the sentence "A hurricane was announced this afternoon on the TV."*

A stylized pitch contour is obtained by simulating how various perceptual thresholds shape the perception of fundamental frequency variations for the average listener. Moreover, quantitative data on prosodic variables are obtained. Prosogram is implemented as a script (a program) for the Praat software for acoustic analysis of speech (Boersma Weenink 2020). Praat was used for the speech segmentation as described in Section 3.3.1. Specifically, the prosogram uses four perceptual transformations for pitch perception in speech, as described by Patel et al. (2006). First is the segregation of the f0 contour into syllable-sized units (House, 1990). Second is the threshold for detecting pitch movement within a syllable (the "glissando threshold"). Third is the threshold for detecting a change in the slope of a pitch movement (the "differential glissando threshold"). Fourth is the temporal integration of f0 (the process of combining acoustic cues over time for periodicity detection within a syllable). The result of these transforms is a sequence of discrete tonal segments, as seen in Figure 3. More information about the details used for the computation in Section 3.4.1.

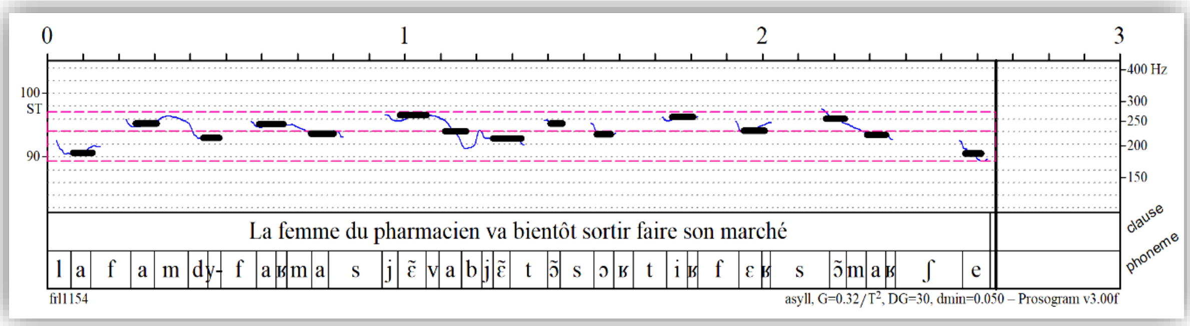

Figure 3: *Illustration of the prosogram, using the French sentence "La femme du pharmacien va bient^ot sortir faire son march´e" shows the original f0 contour(blue lines) and the prosogram (black thick lines).*

It is worth noting that the representation produced by the prosogram consists mainly of level pitches (music like) and that the prosogram can provide quantitative data for pitch height and pitch interval. For the pitch height, it measures the spread of pitches about a mean pitch and for the pitch interval measures whether steps between successive pitches tend to be more uniform or more variable in size.

5 Materials

5.1 Speech

The speech materials used for this experiment are the same as those in Patelet al. (2006) which will be referred to as "the previous study" until mentioned otherwise. They consist of 40 sentences, taken from the database of Nazzi et al. (1998) offered for this experiment by Patel after personal communication. Specifically, they consist of 20 sentences in English and 20 sentences in the French language. For that matter, four female speakers per language were recorded, who read five different sentences in a quiet room. These recordings were digitized with the sampling rate of 16000 Hz. These are short newslike utterances (see Appendix A) and have been used in several studies of speech rhythm (e.g., Nazzi et al., 1998; Ramus et al., 1999; Ramus, 2002). Table 1 consists of basic descriptive statistics on the sentences studied which have been recalculated for the replication and the one additional experiment.

	Duration(s) Mean	Speech rate (syll/s) Mean	No. Vowels Per Sentence Mean	Total Vowels
English Speech n=20	2.8	5.8	15.7	314
French Speech n=20	2.8	6.1	17.3	346

Table 1*: Basic descriptive statistics on the sentences studied*

5.2 Music

Musical data, derived from musical themes were taken from a musicological sourcebook for instrumental music (A dictionary of Musical Themes, Barlow and Morgenstern, 1983). The musical themes are from 20th century for English and French composers born in the 1800s and died in the 1900s. This period is in close proximity to today which is preferable since the speech analysis is based on living speakers, and languages do change phonologically over time (Ohala, 1993; Labov, 1963; Kiparsky, 1995;). This period encompasses a great mixture of post-Romantic styles and modernist classical music including late romantic, expressionist, impressionist and neoclassical styles of composition. The era, as mentioned earlier, was described as the era where many composers decided to forgo or consider the frequently used, until then, specific values of the common practice period, such as traditional tonality, structure, instrumentation and melody. Morgan (1984), in his paper “Secret Languages: The Roots of Musical Modernism” refers to this era as a period during which “*a crisis developed in musical language as shattering as that in the language to literature [...] not by chance, this system began to be theoretically codified at just the time instrumental music began to break away from its vocal-linguistic heritage. It was as if music, suddenly removed from the semantic and syntactic foundation previously supplied by language, had to discover its grammar.*”

Selection criteria introduced by Patel and Daniele (2003) serve as the guiding principles. Themes with “song” or similar terms in their title (i.e. chant, choral and barcarole) were deducted. In that light, titles with an external rhythmic agenda such as waltzes dances were excluded, apart from pieces where the composer was consciously seeking a particular style, such as children’s music. Each theme has at least 12 notes and no internal pauses/rests, thus providing a good sample for nPVI calculation without any grace notes or fermatas which introduce durational uncertainties. A total of 316 out of 318 musical themes passed the criteria verifying the review by Patel et al. (2006). One hundred thirty-six were English and 180 French themes (Appendix B. includes a list of composers and themes derived from the dictionary as mentioned above).

5.3 Rhythm coding for English and French

Rhythm coding performed for both English and French speech and music, as in Patel et al. (2006) was used to calculate the duration of utterances and music notation. By rhythm coding we refer to the procedure of extracting the durational data that we used for the measurements. For speech this was done with speech segmentation (see Section 3.3.1) and for music the measurement was made directly from music notation (see Section 3.3.2). The results of this coding were used for both the replication experiments in addition to investigating the slope variability.

5.3.1 Speech Rhythm for English and French

Vowel boundaries were marked in English and French, intending to investigate the linguistic nPVI. In order to calculate the nPVI, the duration of each vowel in an utterance was calculated as in Patel et al. (2006). Wide-band speech spectrograms were generated with Praat running on a personal computer using Linux software. In the initial experiment, SIGNAL was used for the marking of vowel boundaries. Praat program was used for this experiment instead of SIGNAL because the same program was used for melodic coding and analysis. Additionally, SIGNAL could not operate on Linux software. Wide-band spectrograms were computed using Hanning and Gaussian windows when using only the Hanning one in the previous study. For the Hanning window, the frequency resolution is 125 Hz and the time resolution 8ms. For the Gaussian window, the window length settings were: 0.005 sec., the maximum frequency: 8000Hz, the time step: 0.002 sec., the Frequency step: 50 Hz. The Gausian window was used as well as it is superior and gives no sidelobes in the spectrograms. Moreover, it analyzes a factor of 2 slower than the other window shapes because the analysis is being performed on twice as many samples per frame. The phonetic segmentation was based on the waveform and both spectrograms while using the interactive playback. Vowel onset and offset were defined using standard criteria as per Peterson and Lehiste (1960). The results report vowel measurements only. An example of both windows is on Figures 4 and 5.

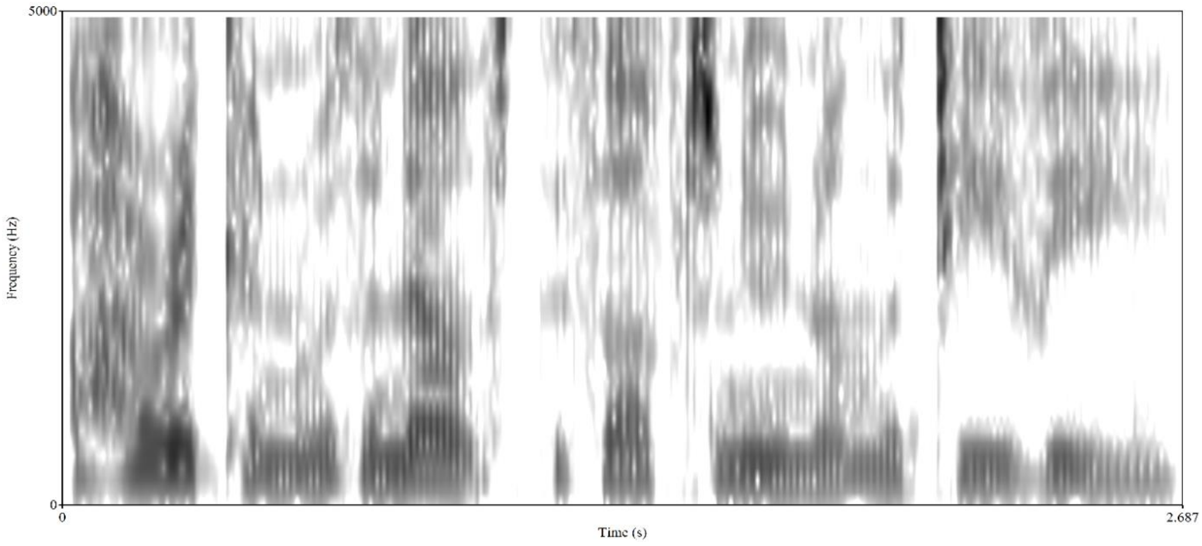

Figure 4: *Hanning window of the sentence "A hurricane was announced this afternoon on the TV."*

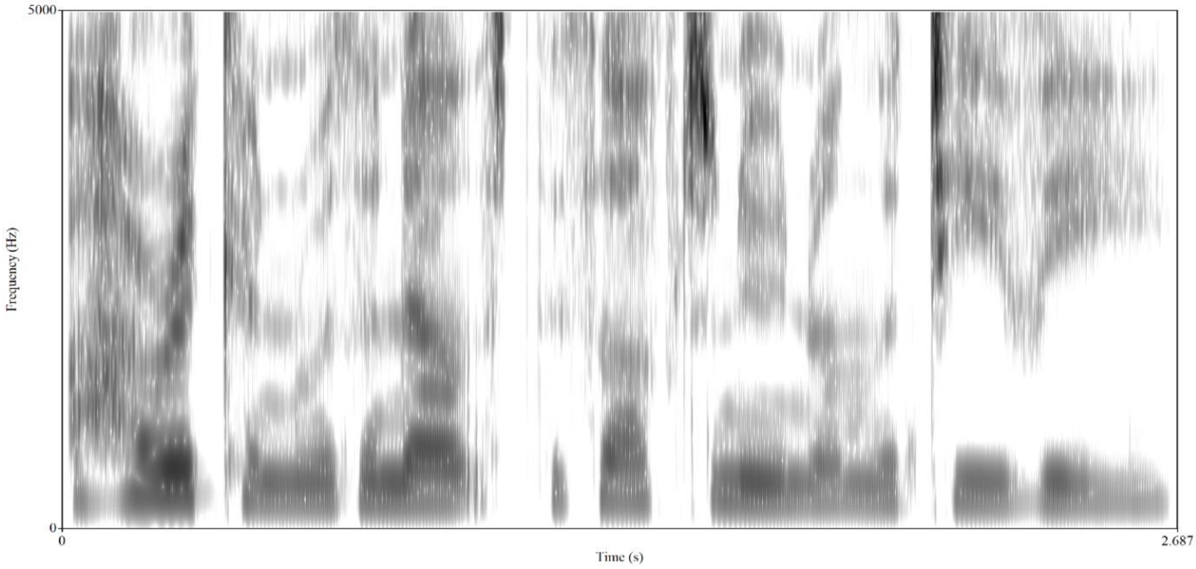

Figure 5: *Gaussian window of the sentence "A hurricane was announced this afternoon on the TV."*

5.3.2 Music Rhythm for English and French

Durational coding was made directly from music notation instead of acoustic recordings. Furthermore, there is the contrast of speech measurements and musical coding which is based on music notation. On the one hand, analyzing music from acoustic recordings raises its own problems, as already mentioned by Patel and Daniele (2003). While most acoustic recordings come from musicals, analyzing music must overcome the contrast of different performances. Moreover, it is difficult to get safe data when the recordings are from musicals in which musicians play more than one instrument. On the other hand, music notation is aligned with the advantage of safer metrics in addition to a more realistic view of the composer's intentions. Notes were assigned durations according to the time signature, with the basic beat assigned to a duration of 1 for the quarter note. An example is shown in Figure 6. Furthermore, a duration of 4 was assigned for the whole note, a duration of 2 was assigned for the half note, a duration of 1 for the quarter note, 1/2 for the eighth note and a duration of 1/4 for the sixteenth note. nPVI concerns about the ratio of durations so any relative measure which preserves the relative durations will conclude the same nPVI for the same sequence.

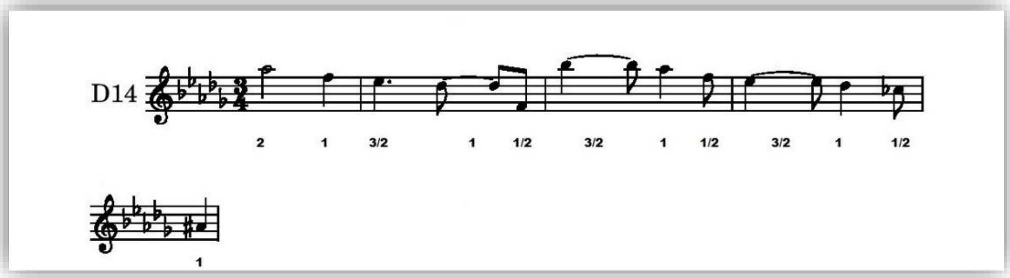

Figure 6: *An example of durational coding for Debussy's musical theme. Each note is assigned a duration according to the time signature.*

5.4 Melody coding

Melody coding performed for both English and French speech and music, as in Patel et al. (2006), was used to calculate the pitch height and the pitch interval of vowels and music notation.

5.4.1 Speech melody

Prosograms were computed for all 40 sentences using the updated prosogram version 3.00f as instantiated in Praat (Mertens, 2004a; 2004b). To compute the f0 measurements, the autocorrelation pitch detection algorithm used was provided by Praat and used by Prosogram.

In detail, the Prosogram script automatically selects the frequency range of lower and upper frequencies used for f0 detection. The automatic pitch range is selected by entering a 0 (zero) for the lower frequency. The pitch range of 30 ST is larger than commonly used by speakers, though it is useful to turn to speech recordings containing multiple speakers with different pitch ranges. The same parameters were used as in Patel et al. (2006), combining the default parameters and pre-set ones. To exemplify, the minimum and maximum pitch were set to 60Hz and 450Hz respectively. The frame period of 0.005s was used which is recommended for a high temporal resolution. To determine whether a given vowel would be assigned a level tone or a glide the (Glissando's) threshold of

$$G = 0.32/T^2$$

was used. These thresholds are better suited for continuous speech, where the stimulus is heard once and there are no systematic pauses after syllables. Furthermore, the choice of this Glissando threshold, as described by Patel, is based on perceptual research on the threshold for detecting pitch movement in speech, while prosogram output is compared to humxn transcriptions of intonation.

Another key fact is that, only level tones were used in the quantification of pitch variability in speech. Similarly, the coefficient of variation (CV) was used to compute the variability in pitch height and pitch interval. To examine the pitch height variation and interval variability the same process was performed as in the study by Patel.

Coefficient of variation (CV) is a standardized measure of dispersion of a probability frequency distribution. It can be described as:

$$CV = \frac{\sigma}{\mu}$$

Where σ is the population standard deviation and μ the population mean. The standard deviation is a measure of the amount of variation or dispersion of a set of values.

For the former, each level tone was assigned a semitone distance from the mean pitch of all level-tones in the sentence. In turn, the CV of these pitch distances was computed. For the latter, adjacent level tones in a sentence were assigned a pitch interval in semitones,

$$st = 12\, log2\, (f2/f1)$$

where (f1) and (f2) represent the initial and final tone of the pair, respectively. The CV of these intervals was then quantified. To elude ill-defined CV's, measurements of pitch distances and pitch intervals used absolute values. Since the CV is dimensionless, one could assess pitch in speech and music in different units. The precise choice of units for speech is not expected to affect the results outlined here.

5.4.2 Music melody

Musical themes were coded as sequences of pitch values where each represents a given pitch semitone distance from A440 (Figure 7). Measures of pitch height and interval variability were then computed in precisely the same manner as for speech. As long as the relative positions of tones are preserved, any pitch referent measurement will lead to the same results.

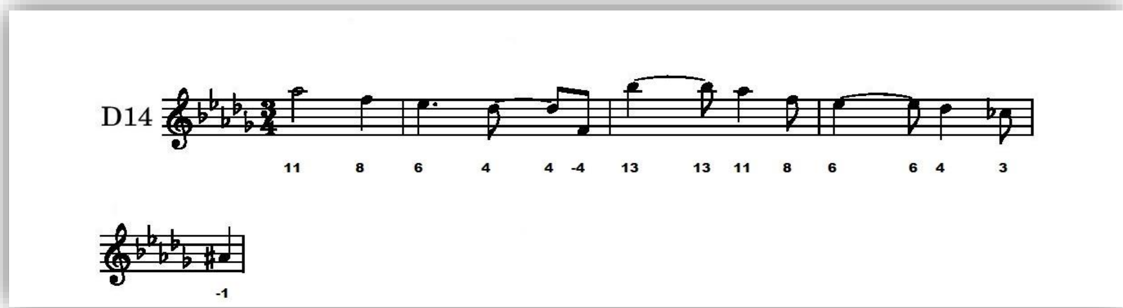

Figure 7: *Illustrates an example of pitch coding for Debussy's musical theme. Each note is assigned a pitch value based on its semitone distance from A4. The coefficient of variation CV of pitch intervals in D14 is 1.35.*

5.5 Slope Variability

The slope variability coding was performed for both speech utterances and musical themes. In contrast to the melody analysis where the level tone distance between the consecutive notes was tested for pitch height and pitch interval, the technical rationale of slope variability notes the durations of each pitch interval. Specifically, slope tests whether duration accounts for changes of pitch interval. The pitch interval is divided by the duration of the previous note or vowel (see Appendix C). Gussenhoven and Zhou (2013), reported that the perceived duration of vowels correlates with the height of level pitch. By the same token, many researchers have examined pitch-interval selectivity (in musical contexts). The ratio of durations for both speech (duration ratio of vowels) and musical themes (duration ratio of notes) was calculated. Some of the duration intervals in both sentences and musical themes summed an interval of zero (0). Pitch interval can not be divided by a zero-duration interval which will result in ill defined results. Subsequently, the pitch interval of the note pair was divided by the duration of the first note of the same note pair (see Appendix C). For speech and music, each pitch interval was divided by the duration of the initial vowel/note within the tested pair. This provided the slope variability of each pair which was calculated as per standard deviation divided by the mean. This was performed for all sentences and musical themes and the variability of the slope within a sentence/musical theme was computed by the CV. Specifically, in order to test if ratio of durations when measuring pitch interval has an effect on variability, we collected all the raw duration points (vowels in milliseconds / notes in beats) and divided them by the pitch interval points (semitones) which were used for the pitch interval approach. Then the Coefficient of Variation (CV) method was used to measure their variability.

5.5.1 Speech Slope

Pitch data points for speech were driven from prosogram. Hence, if the vowel was devoiced, its intensity would be too low or if Praat produced an erroneous f0 value, the prosogram would not assign a tonal element. Such emissions were rare: 5% of English vowels were assigned glides, while on French vowels 2%. The pitch data points that were not assigned a level tone were excluded.

5.5.2 Music Slope

For the musical slope coding, this experiment uses the same data set as in the Duration Coding. The last data point from each musical theme is erased as it does not give additional information when divided with the data points from the pitch interval. Pitch interval points in every musical theme will end with one data point less than those in duration if this deduction is not performed and the last duration point does not serve any reason in the calculation of the slope. An example is shown in Figure 6, when Debussy's musical theme has 12 pitch intervals and 13 durations. The whole data set went under the same computation.

6 Results

6.1 Rhythm

6.1.1 Speech

Table 2. shows the nPVI and CV values for speech compared to the initial study. Reported p-values in the following tables were computed using the Mann-Whitney U-test, except for p-values associated with the Monte Carlo analysis as in the study by Patel et al. (2006). The Mann-Whitney U test is used to compare whether there is a difference in the nPVI variable for each tested language. The same was performed to determine whether there is a difference in the CV. It compares whether the distribution of the dependent variable is the same for the two groups and therefore from the same population. Table 2 shows that English and French sentences have a highly significant difference in durational contrastiveness (nPVI) as well as in durational variability (CV). Results with the remark of **"Patel"** are those reported in the initial study for better comparability. The differences in absolute numbers as opposed to the initial study can be due to the different segmentation but as the same procedure was performed for all sentences this does not change the results. It is important to highlight here that the difference is significant.

	nPVI (s.e)	CV (s.e)	nPVI (s.e) *Patel*	CV (s.e) *Patel*
English Speech n=20	**60.4 (4.0)**	**0.59 (0.02)**	*55.0 (3.0)*	*0.55 (0.03)*
French Speech n=20	**36.3 (2.0)**	**0.35 (0.02)**	*35.9 (1.8)*	*0.36 (0.02)*
p	**< 0.001**	**< 0.001**	*< 0.001*	*< 0.001*

Table 2*: nPVI and CV values for Speech*

English speech has higher nPVI and CV scores than French speech. However, the difference in nPVI and CV scores for English speech compared to the “Patel” results is slightly higher than the ones in French speech. The difference that the replicated study has with the “Patel” results for English speech is 5.6 for nPVI and 0.04 for CV, whereas for French speech the difference is higher for nPVI by 0.04 and slightly lower by 0.01 for CV when compared with the “Patel” results.

Additionally, regressions of CV on nPVI in each domain are shown in Figure 8. Next, in black and white are the results by Patel et al. (2006) for comparability. The linear regressions reveal that within each domain higher CV is predictive of higher nPVI.

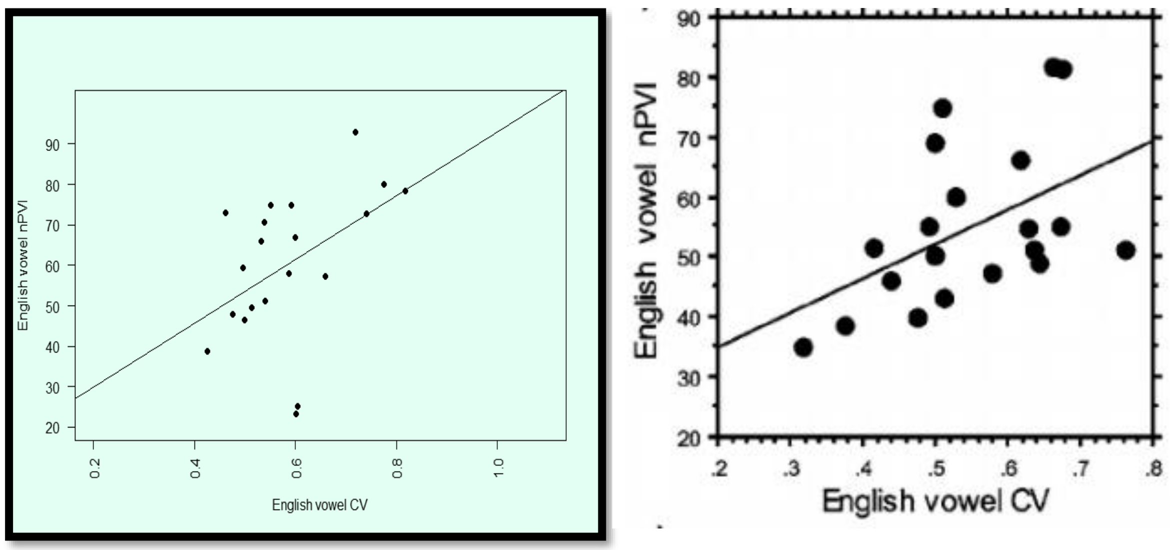

Figure 8: *English speech*

$nPVI = 14 + 79 \cdot CV, \quad R^2 = 0.17, df = 18, p = 0.04$

In English speech the results of the linear regression model show the expected nPVI, when CV is constant (14). For a unit increase in CV, the nPVI is increased by 79. Small *p* values show that the effect in question is significant. R^2 adjusted, the proportion of variance, shows that the CV in English speech can explain 17% of the variability of nPVI. Moreover, the *F* statistic tests the null hypothesis that there is no relationship between nPVI and CV. The further that statistic is from 1, the higher the chance is that null hypothesis will be rejected. In this case it is equal to 4.96 and the p value < 0.05. This shows that there is a relationship between nPVI and CV, visualized by the linear regression in Figure 8.

In French speech, as illustrated in Figure 9, for a unit increase in CV, the nPVI is increased by 57.5. Small p values show that the effect in question is significant. R^2 adjusted, the proportion of variance, shows that CV in French speech can explain 51% of the variability of nPVI. Likewise, the F statistic is equal to 20.8 and the p value < 0.05. This shows that there is a relationship between nPVI and CV, visualized by the linear regression in Figure 8.

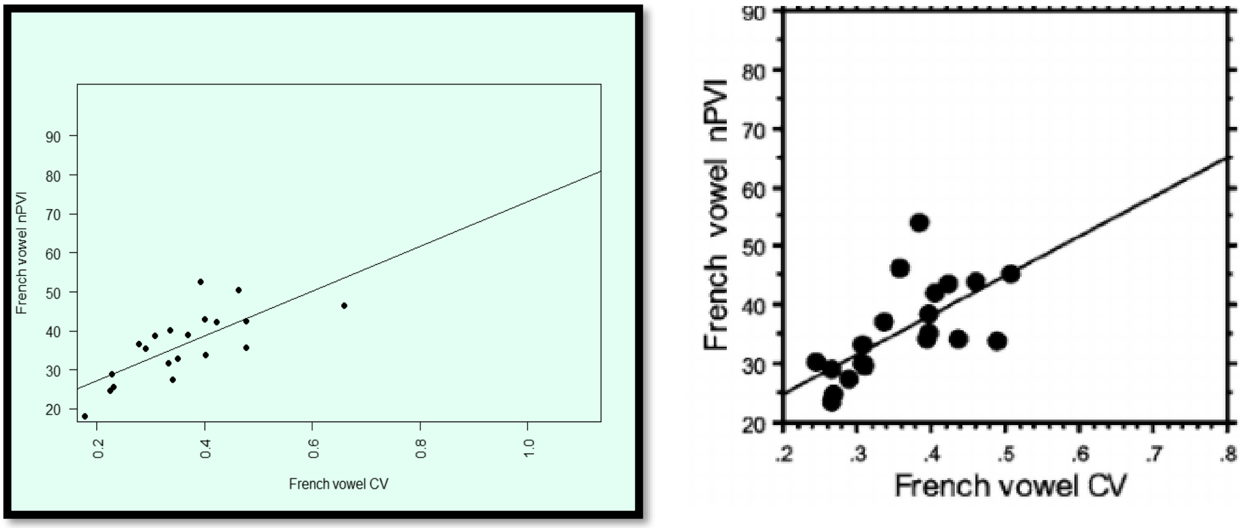

Figure 9: *French speech:*

$$\boldsymbol{nPVI = 15.8 + 57.5 \cdot CV,\ R^2 = 0.51, df = 18, p < 0.001}$$

This array indicates a relationship between nPVI and CV in French speech. Figure 10 illustrates boxplots and histograms for both English and French nPVI which provide visualization of the distributional characteristics. English nPVI values seem to be slightly negatively skewed while French nPVI are quite symmetrical.

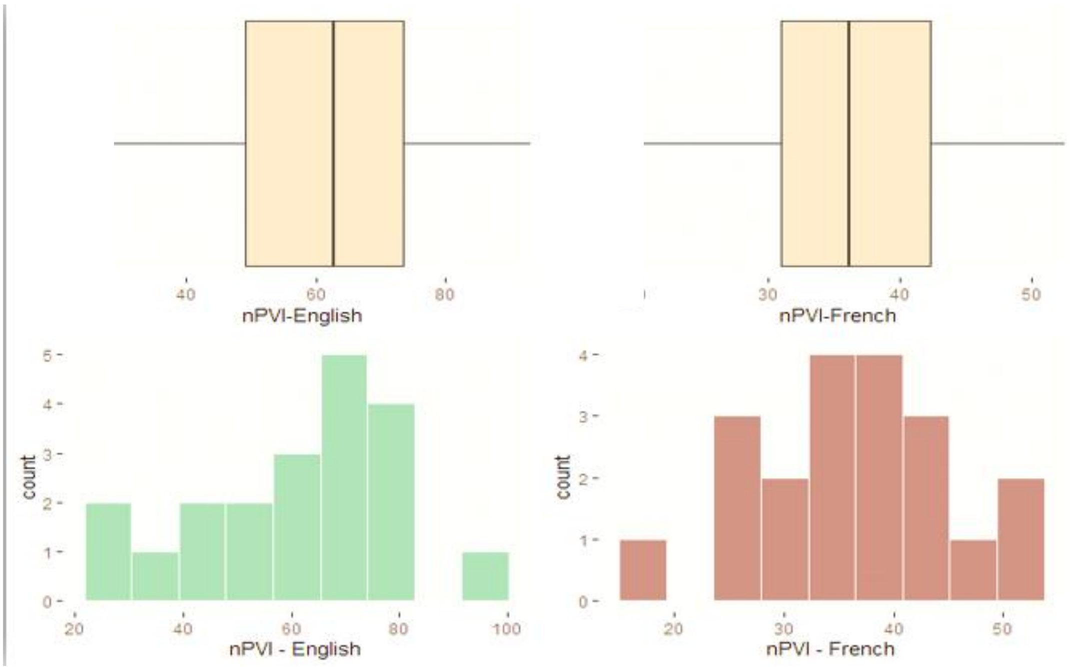

Figure 10: *Boxplots and Histograms for English and French nPVI. The y axis in the histograms show how many times the nPVI score was observed on the x axis.*

Furthermore, a visualization of boxplots and histograms for both English and French CV is shown in Figure 11. In the boxplot of English CV an outlier is observed, and the median is between $0.5 - 0.6$. The scatter of CV values in English seems to be larger than the one in French. None of the CVs appear to be extremely positively or negatively skewed, but a slight negative skewness can be observed in CV English. The histograms, present in detail what is being shown in the boxplot. The CV French has only one extreme value, while in English the amount of extreme values raises to four. Both suggest normalized data. At the same time, English CV seems to be slightly negatively skewed. The reason for that, supported by the histograms, is that four values above 0.7 were picked by the boxplot as outliers.

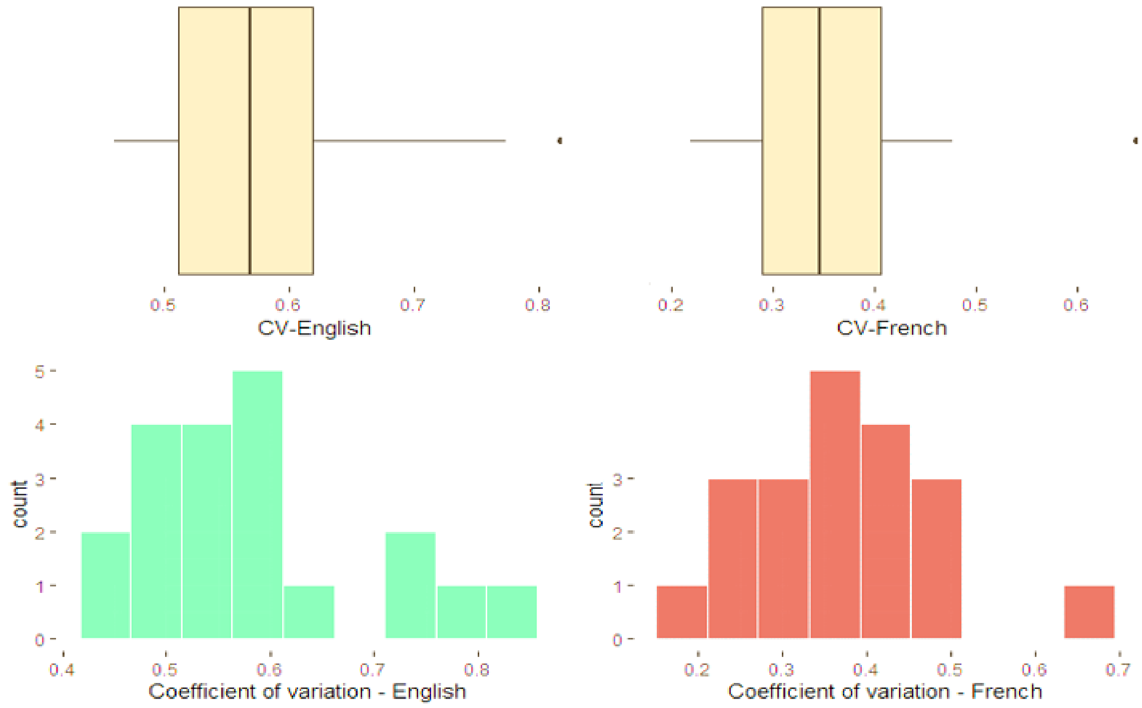

Figure 11: *Boxplots and histograms for English and French CV. The y axis in the histograms show how many times the CV score was observed on the x axis.*

The Monte Carlo results for speech are portrayed in Figure 12. The Monte Carlo analysis was performed in order to investigate whether CV differences between the two languages are responsible for linguistic nPVI differences, given that the linear regressions before showed a relationship. It is important to keep in mind that the consecutive order of elements matters for nPVI while in variability it does not. This plots the distribution of nPVI differences between English and French speech when the order of vowel durations is scrambled and the nPVI difference between the two languages is computed after 1000 iterations (the same approach was done by Patel et al. (2006)). The actual nPVI difference for speech is 24.1 points which is presented by a red line in Figure 12. The probability of an nPVI difference of 24.1 points or greater for speech is p=0.01. This shows that it is highly unlikely that variability differences account for nPVI difference, thus verifying the results of the previous study. This means that the order of the elements is important for nPVI measurements.

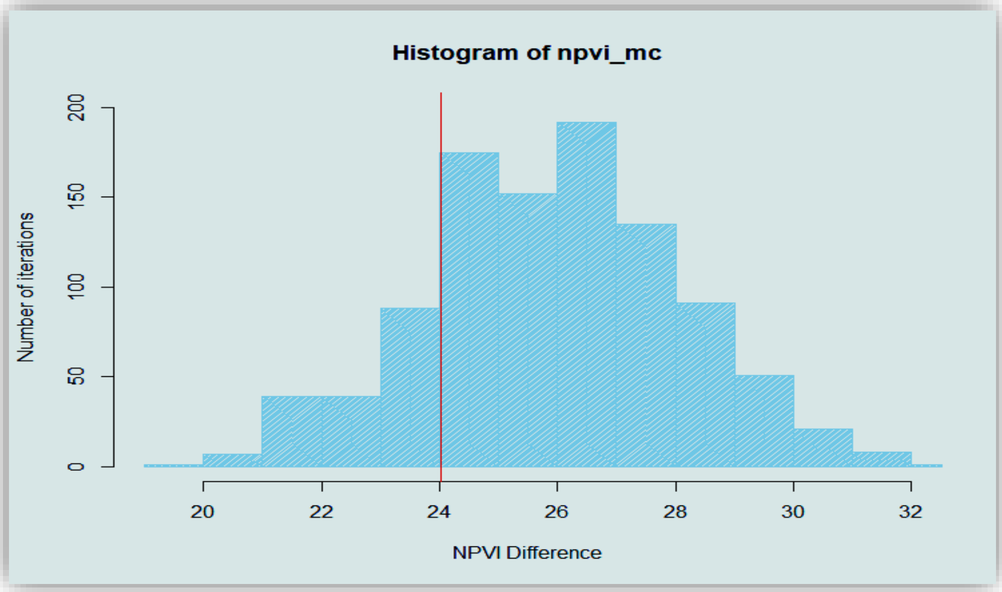

Figure 12: *MonteCarlo Speech*

6.1.2 Music Rhythm

The same results presented here indicate that the relationship appears to be stronger in speech than in music. Table 3 shows that the difference in nPVI measurement is 7.5 while the difference in CV seems to be insignificant with a reported difference of 0.03.

	nPVI (s.e)	CV (s.e)	nPVI (s.e) *Patel*	CV (s.e) *Patel*
English Music n=136	**47.6 (1.8)**	**0.60 (0.02)**	*47.1 (1.8)*	*0.61 (0.02)*
French Music n=180	**40.1 (1.8)**	**0.57 (0.02)**	*40.2 (1.9)*	*0.58 (0.02)*
p	**< 0.01**	**0.27**	*< 0.01*	*0.34*

Table 3*: nPVI and CV values for Music*

Figure 13 shows that in English music, the expected nPVI if CV is constant, is 27.67 while for a unit increase in CV, the nPVI is increased by 33.13. Moreover, small *p* values show that the effect in question is significant. It also indicates that the relationship between nPVI and CV due to chance is rejected. The significance of this variable (∗∗∗) indicates that there is a highly significant feature. R^2 adjusted, the proportion of variance show that CV in English music can explain 11% of the variability of nPVI. The *F* statistic is equal to 18.35 and the *p* value < 0.05 showing that there is a significant relationship between nPVI and CV, visualized by the linear regression in Figure 13.

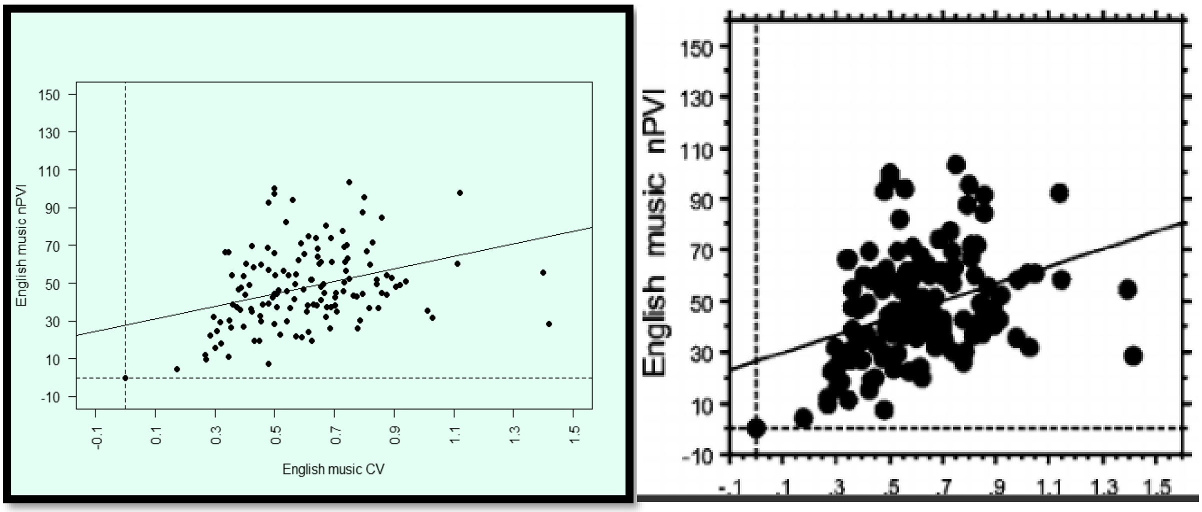

Figure 13: *nPVI and CV plot for English Music:*

$$nPVI = 27.67 + 33.13 \cdot CV, R^2 = 0.11, df = 134, p < 0.001$$

In French music, as illustrated in Figure 14, the expected nPVI is 9.17, when CV is constant. For a unit increase in CV, nPVI is increased by 54.17 units. Moreover, small *p* values show that the effect in question is highly significant. R^2 adjusted, the proportion of variance, shows that CV in French music can explain 31% of the variability of nPVI. *F*. statistic is equal to 79.79 and the *p*.value < 0.05. This shows that there is a relationship between nPVI and CV, visualized by the linear regression in Figure 14.

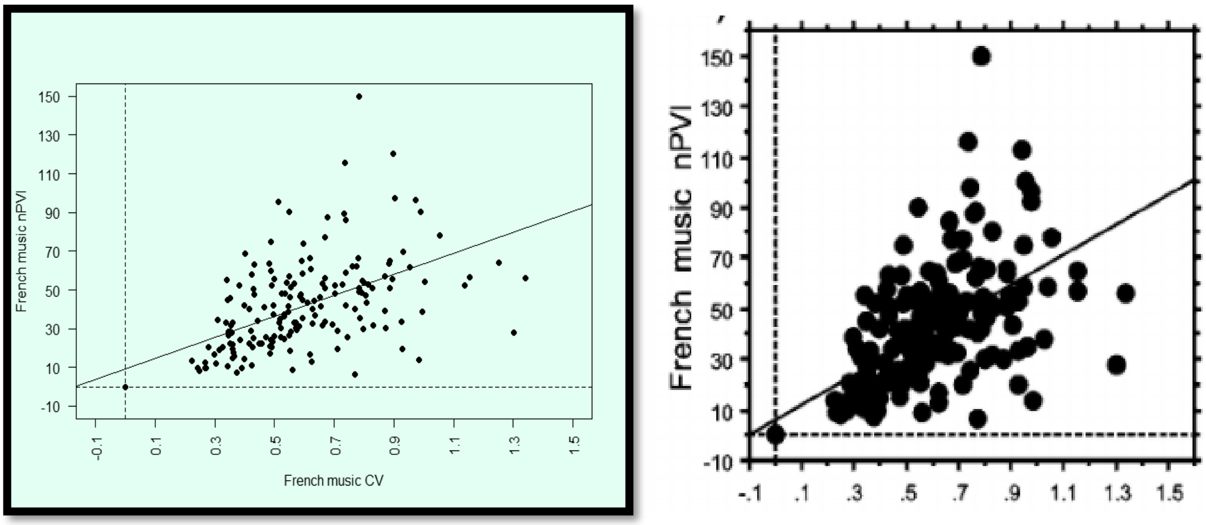

Figure 14: *nPVI and CV plot for French Music:*

$$nPVI = 9.17 + 54.17 \cdot CV, R^2 = 0.31, df = 178, p < 0.001$$

The Monte Carlo results are presented in Figure 15. The actual nPVI difference for music found to be at 7.5 points, which is represented again by a red line. The probability of the observed difference in nPVI, given the observed difference in variability in music, results a p − value of 0.001. As in speech, this indicates that it is highly unlikely that variability differences account for nPVI difference in either domain, verifying the results of the previous study.

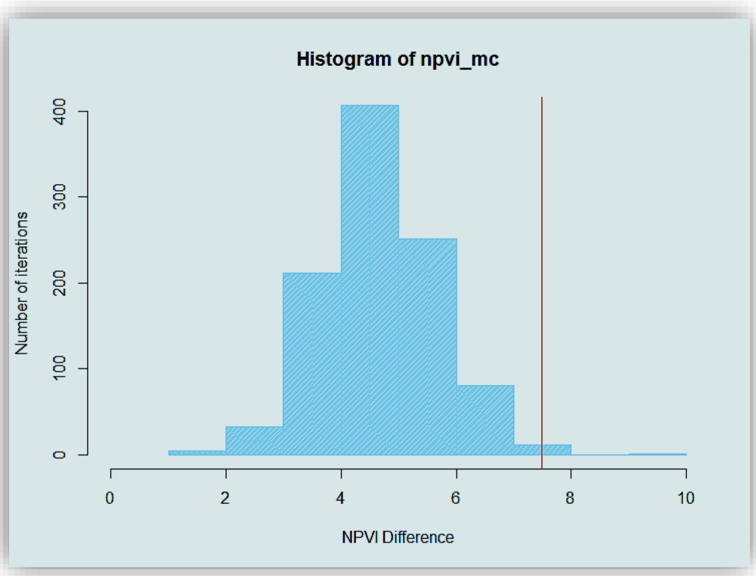

Figure 15: *MonteCarlo Music*

6.2 Melody

6.2.1 Speech

Table 4 shows the results of pitch height and pitch interval variability measurements for speech. In English speech, the pitch height was found at 0.73 which is 0.02 points higher than the initial study. When in English pitch interval, reported results were 0.03 points lower. In French, the pitch height appeared to be 0.01 higher where the pitch interval was identical. This difference might be due to the fact of different segmentation and the updated version of the Prosogram program.

	Pitch Height	Pitch Interval	Pitch Height *P*	Pitch Interval *P*
English Speech n=20	**0.73 (0.05)**	**0.85 (0.04)**	*0.71 (0.04)*	*0.88 (0.05)*
French Speech n=20	**0.77 (0.03)**	**0.68 (0.03)**	*0.75 (0.04)*	*0.68 (0.03)*
p	**0.12**	**< 0.01**	*0.32*	*< 0.01*

Table 4: *Pitch height and Pitch intervals CV values for Speech*

6.2.2 Music

The results showed in Table 5. depict the same notion as in speech. Differences in pitch height are not significant when in pitch interval, English music is higher by 0.05 points which has identical results as in the previous study. This is normal as the pitch points were extracted from music notation. This data showes that the linguistic difference in pitch interval variability between English and French speech is much more pronounced than the musical difference.

	Pitch Height	Pitch Interval	Pitch Height P	Pitch Interval P
English Music n=136	**0.69 (0.01)**	**0.76 (0.02)**	*0.69 (0.01)*	*0.76 (0.02)*
French Music n=138	**0.71 (0.01)**	**0.71 (0.02)**	*0.71 (0.01)*	*0.71 (0.02)*
p	**0.14**	**0.03**	*0.14*	*0.03*

Table 5: *Pitch height and Pitch intervals CV values for Music*

Additionally, Figure 16 shows the pitch interval variability in both speech and music which is described by the CV of the absolute interval size between pitches in a sequence. Reported with red colour is speech while in black colour is music. The difference of Pitch Interval Variability in music is about 23.5% of the counterpart of speech.

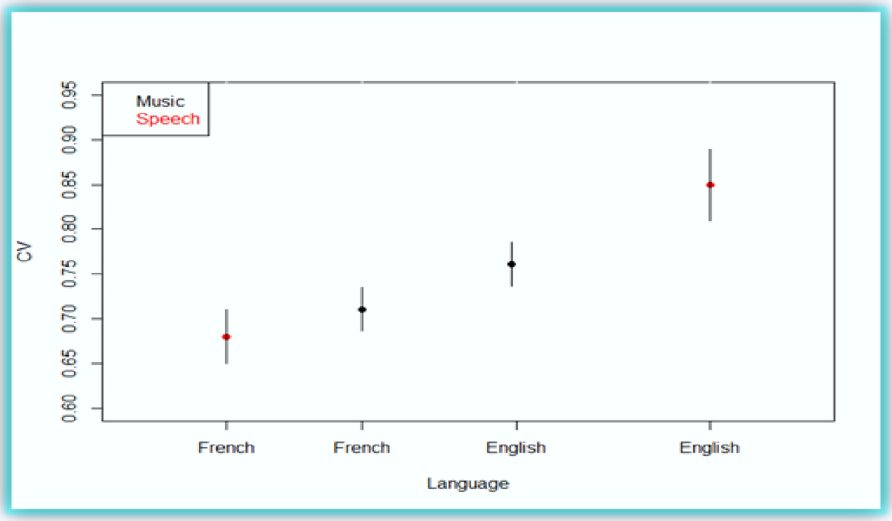

Figure 16: *Pitch Interval Variability*

In Figure 17, a representation of nPVI values is shown. The red dots describe speech while the black ones describe the music for both English and French. Like Figure 13, this representation has the same drift as in the previous study (Patel et al., 2006).

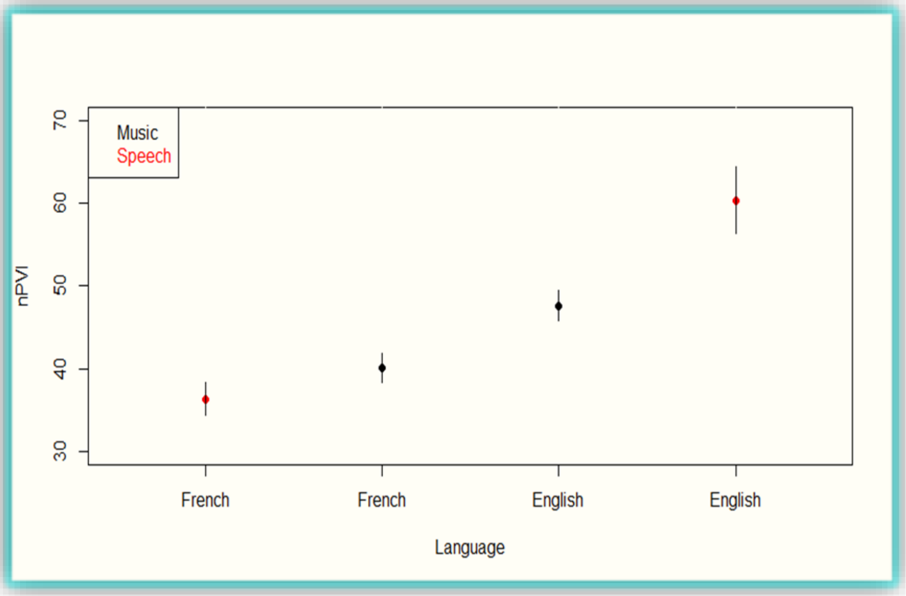

Figure 17: *nPVI values for English and French*

6.2.3 Slope Variability Results

As the results of nPVI and CV pitch interval yielded the same effect as in Patel, this allows us to assume these metrics are usable enough in categorizing language and music based on their national characteristics. In addition, the technical rational of slope variability was tested.

Table 6. shows the results for both speech and music slope variability differences in the same data set. For speech and music, each pitch interval was divided by the duration of the initial vowel/note within the tested pair. The CV of the slope of each pair was calculated as per standard deviation divided by the mean. Reported p-values in the following tables were computed using the Mann-Whitney U-test as the same test was used for nPVI and pitch height and pitch interval. However, variability of the slope appeared to be higher in French speech than in English speech. For the former, the reported slope variability found to be at 3.2, where the latter was lower, resulting in a slope variability of 1.9 with the *p* value of 0.001. Similarly,

the difference of slope variability in French music was slightly higher by 0.1 in contrast to the same slope variability in English music with the *p* value of 0.05.

In this case, the results of the slope variability portray the opposite effect from the one in the study by Patel et al. (2006) for pitch interval variability. This questions the quantitative method's ability to categorize language and music based on their national characteristics. The opposite effect is observed in speech with French resulting in a higher slope CV than English speech. The comparison of the difference with respect to slope CV variable in the two groups (English - French speech) shows that both groups have significant differences in the mean values of their slope variability with the $p < 0.001$. For music, the effect in question appears to be insignificant. The mean values of their slope variability has a *p* value of 0.05. Moreover, the results of the slope CV are not aligned with the results of nPVI and CV of pitch interval, as expressed by Patel et al. (2006) and verified by the replication of this study, calling into question the method's ability to make such predictions.

	English Speech	French Speech	P value	English Music	French Music	P value
CV Slope	**1.9 (0.6)**	**3.2 (0.9)**	***< 0.001***	***1.2 (0.4)***	***1.3 (0.4)***	**= 0.05**

Table 6: *CV(s.e) Values for English/French Speech and Music Slope*

Boxplots and Histograms for English and French speech slope CV, as illustrated in Figure 18, show that the median is almost in the middle of the interquartile range (IQR). However, from the histogram, it is well observed that the scatter is indeed too large for these values, leading to the assumption that the distribution of CV English is not normal. Contrarily, in French, the scatter is very small but also heavily positively skewed. Furthermore, an extreme value can be observed with a range value above 5.

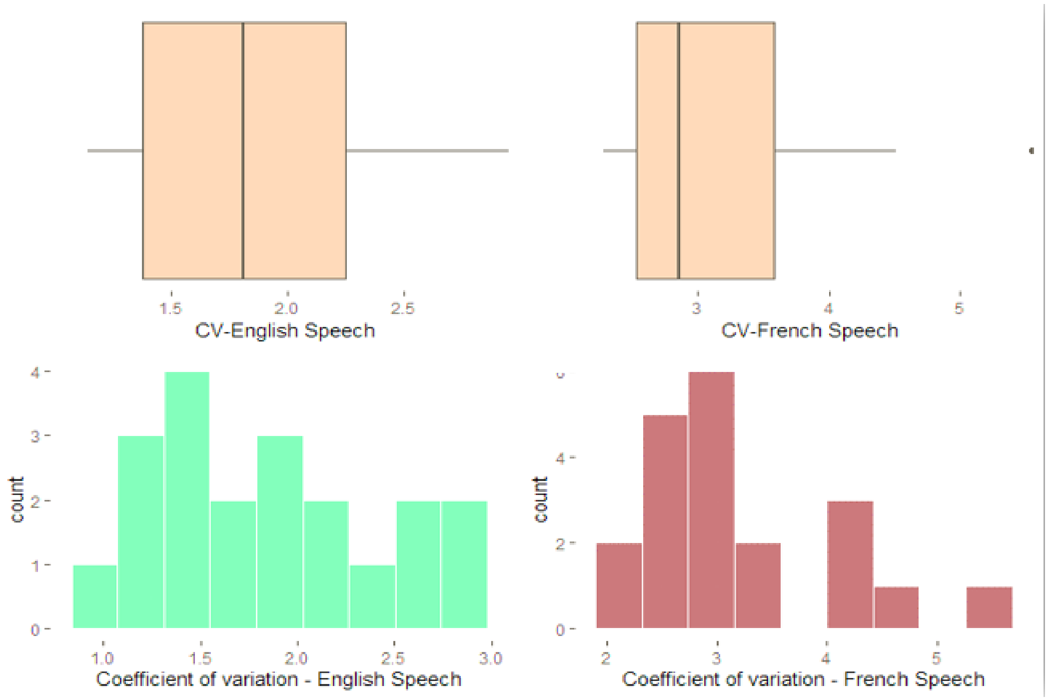

Figure 18: *Boxplots and Histograms of Slope in English and French Speech*

Following the above, Figure 19 represents Boxplots and Histograms for English and French Music Slope. The former is positively skewed while the latter seems normally distributed with some extreme values. Additionally, CV English portrays a greater scatter in the boxplot.

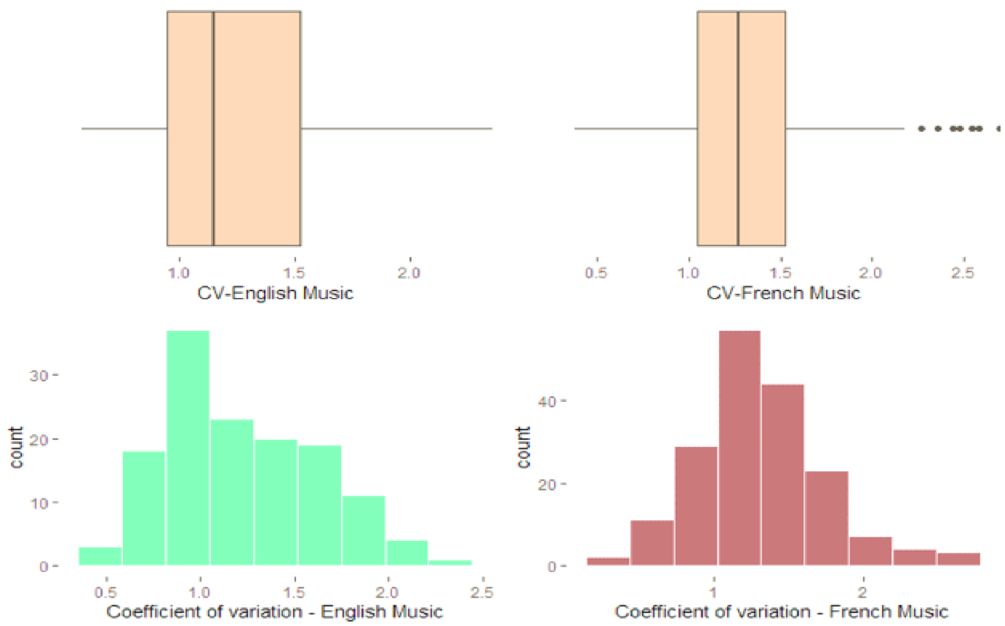

Figure 19: *Boxplots and Histograms of Slope in English and French Music*

Figure 20 illustrates CV slope for both music and speech. The difference here appears to be again greater in speech than in music, where the CV slope appears to be higher in French. In contrast with Figure 13 of pitch interval and Figure 14 of nPVI, French CV slope shows significantly higher variability than English, for both speech and music. In speech the difference of CV slope was 1.3 higher for French. In music, the difference was significantly lower but with French music scoring 0.1 higher values than English music.

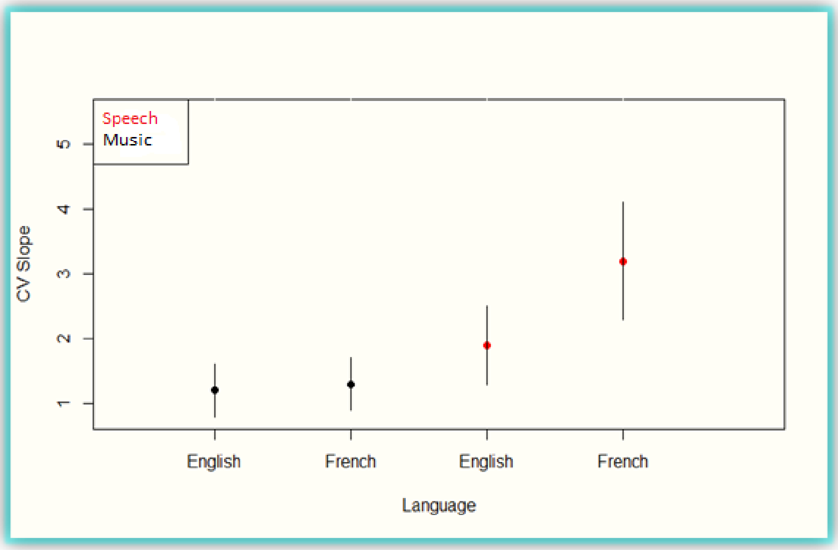

Figure 20: *Slope Interval Variability*

Figure 21 shows all 3 figures of CV slope, nPVI of duration and CV of pitch interval with the language labels in the same order for comparability, representing CV slope in the order of nPVI and CV pitch interval. As explained before, the opposite effect is observed for CV slope.

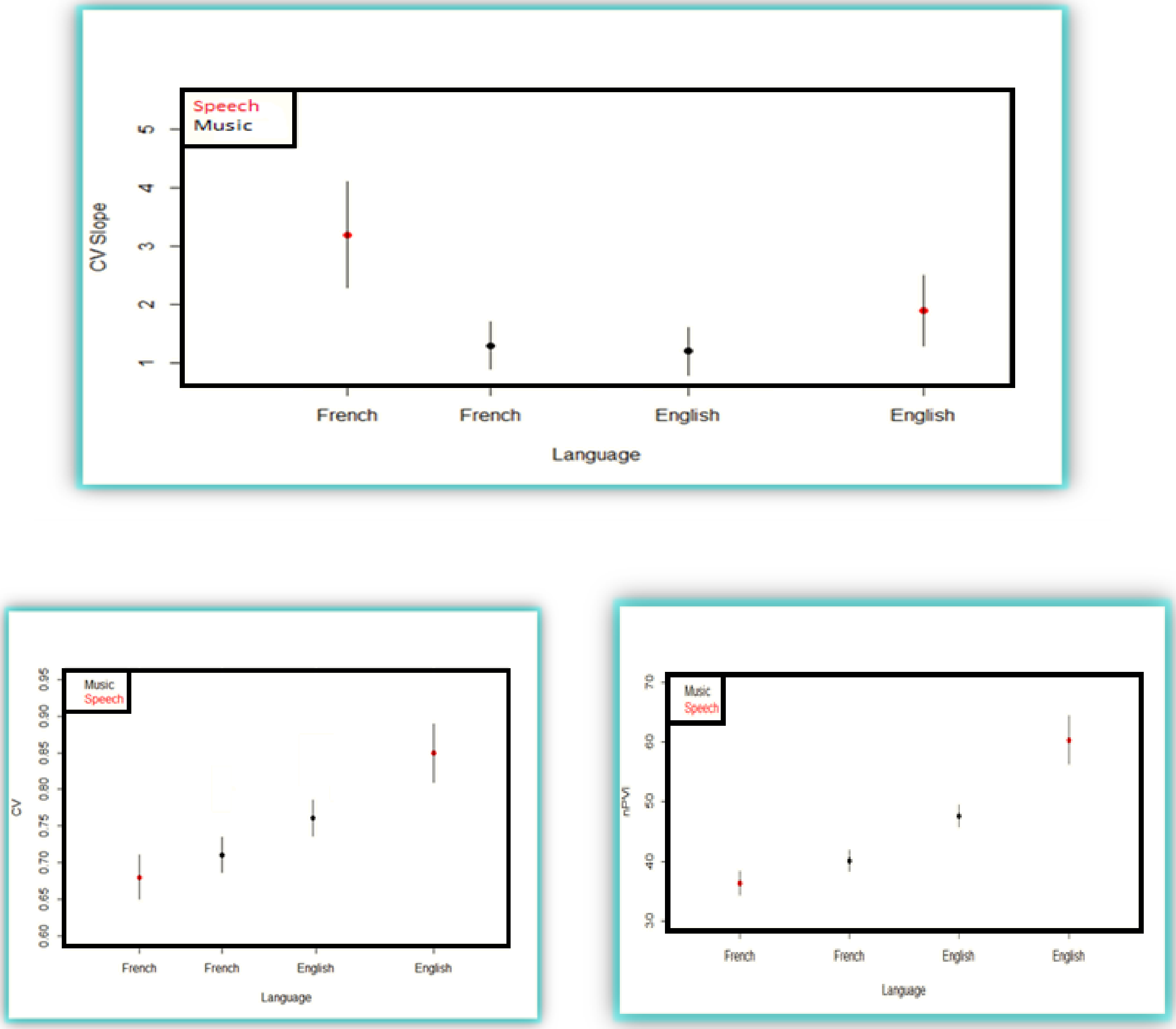

Figure 21: *A combination of CV slope, CV pitch interval, nPVI*

7 Discussion and Conclusion

7.1 Discussion on Rhythm

Results of the study in rhythm showed that nPVI is a good technical tool that can measure the variability of the ratio of durations in speech and music. Results of the nPVI metric for both English and French speech and music yielded the same effect as in Patel et al. (2006), showing that English has a higher nPVI than French for both speech and music. The results here confirmed the analysis by Patel et al. (2006) and showed that the same effect is present in the specific dataset. Moreover, the results were −almost− identical, and allowed us to move to the next parameter, testing the extra technical rationale of the slope variability. The CV slope did not have the same effect as nPVI. Previous research has shown that the rhythmic properties of music and speech, as quantified by the nPVI, vary as a function of whether the composer's native language was stress-timed (e.g., English) or syllable-timed, e.g., French (Patel & Daniele, 2003). Grabe and Low (2002) have used the nPVI metric to examine the pattern of vowel durations in sentences of several languages. The results showed that stress-timed languages such as English and German have a higher nPVI score as opposed to syllable-timed languages such as French and Italian, a finding tool that follows the results reported here for English and French. However, they found that Tamil, which is a syllable-timed language, has a high nPVI which questions if the effect of the nPVI's metric is accurate or universal in categorizing languages in stress-timed or syllable-timed ones. The CV slope also showed that French has higher variability than English which is opposite to the nPVI findings. The difference in English, in terms of variability, appears to be higher than the one in French. On the one hand, it is uncertain if these findings can serve as a principal guideline and categorize languages. On the other hand, nPVI measurement does seem to provide a measurement of rhythmic complexity which can be seen as changing over time and through different social circles (VanHandel, 2017). Of course, timing is just one of the contributors of rhythm and other factors, such as tone and intonation, must be considered. Additionally, it is at least uncertain that this metric can provide such information on the analysis of rhythm but specifically in such a way that can yield safe metric conclusions regarding nationality. Daniele and Patel (2003) hypothesized that the distancing from the Italianate style in the late eighteenth century could be observed by an increase of nPVI measurement. Vukovic and Shanahan (2020) reported that the Italian Slope is also ascending, indicating that according to their syllable-timed/stress-timed groupings, they should be observing the opposite trend. In that manner, they suggest that nPVI can be used as a marker of intentional aspects of musical style and not as a factor of linguistic origin or nationality. Arvaniti (2010) tested metrics proposed to quantify rhythm with nPVI to be included in one of those. She reported that the metrics show a substantial inter-speaker variation affected by the corpus used to calculate them. She explained that segmental timing is affected by other factors that are not clearly related to syllable structure, e.g., the extent of phrase final lengthening, the phonetic inventory of the language, and the extent of vowel elision. The uses of nPVI have gone beyond simply language and music comparisons. Daniele (2015) argues that nPVI provides a quantitative mechanism for comparison and support of the historical record and theory. Arvaniti (2010) argues that the metrics are unreliable predictors of rhythm, providing no more than a crude measure of timing. It is further argued that timing is distinct from the rhythm and that equating them has led to circularity and a psychologically questionable conceptualization of rhythm in speech. Future studies should use other quantitative measurements that focus on aspects other than crude segmentation which is what nPVI measurement uses.

7.2 Discussion on Melody

The melody study results portrayed an unclear effect of pitch interval to mirror the same difference, as the one in nPVI, between English and French for speech and music. The results on the replicated study verified the results as introduced by Patel et al. (2006). As Bolinger asserted *"since intonation is synonymous with speech melody, and melody is a term borrowed from music, it is natural to wonder what connection there may be between music and intonation"* (Boligner, 1985:28). Following this, several researchers tested the pitch contour. Schön et al. (2004) found that similar cognitive computations and neural systems are involved in integrating pitch processing in both language and music. Besson et al. (2007) reported that a set of standard processes might be responsible for pitch processing in music and speech.

Additionally, Magne et al. (2006) found behavioural evidence for a standard pitch processing mechanism in language and music perception. However, the statistical evidence to track the national origin, as expressed by pitch, varies. On the one hand, the CV of the pitch height on French appears to be higher than in English. On the other hand, the CV of pitch interval portrays a reverse effect with English scoring higher numbers than French. For the latter, pitch interval has a higher score for English speech and music, "mirroring" the same effect of nPVI differences as reported before. To conclude, the CV of pitch interval has -almost- the same drift for English and French as the nPVI metric. Although, the CV of pitch height reported an opposite effect on the two languages which raise some questions on why pitch interval is eligible for categorizing languages based on their ethnic footprint and pitch height is not.

7.3 Discussion on Slope

The results of the CV slope technical rationale are that it shows an opposite effect compared to the nPVI and the CV of the pitch interval. Results show that the CV slope, when taking into consideration the durational property of the pitch interval, the variability in French speech is higher than that of the English speech. Something that is in contrast with the results of nPVI and CV of pitch interval. Returning to our hypothesis, the results of the CV slope raise some questions on the methodology as done by Patel et al. (2006). The results of CV slope revealed that the durational property of the pitch interval, has an effect in terms of variability. As explained before, nPVI, being a tool used to quantify the rhythmic variability of this experiment, measures only variability of time. The CV of pitch interval, used to quantify the melodic variability, measures only pitch intervals. Moreover, the CV slope results for music show the difficulty to discriminate French and English on that matter, with French music scoring insignificant higher numbers (1.3) than English music (1.2) while in nPVI and pitch interval CV, English scored higher numbers than French. Figure 18., as shown before, provides a visualization of these findings. The opposite effect of the CV slope parameter, as expressed by the results of this thesis, is questioning the robustness of the nPVI and CV pitch interval to categorize English and French based on their national characteristics. Many researchers have highlighted the importance of rhythm and pitch and how these two categories are not as distinct as one may think. The CV slope "breaks" the barriers of this distinction. For example, Jones, Boltz, and Kidd (1982) demonstrated that listeners are better able to recognize the pitch of a tone when it is rhythmically accented than when it is not. Furthermore, it has been reported that higher pitch was systematically associated with longer duration judgement and that there seems to be a universal tendency for higher pitch sounds to be perceived longer when all the other sound parameters are kept constant (Martti Vainio et al., 2013). It is well known that the duration of segments is affected by all sorts of factors, and this thesis used the CV slope to investigate whether it can provide the same information as nPVI and CV of pitch interval. Results in the CV slope parameter show that we can not assume that measurements that rely only on ratio of durations or on pitch interval, as in Patel et al. (2006), are sufficient and other factors (e.g., the extent of phrase final lengthening, the phonetic inventory of the language, and the extent of vowel elision) must be considered.

7.4 Conclusion

The replication of the rhythm and melody experiments found that the same effect is present (with relatively insignificant differences) for English and French as in the study by Patel et al. (2006) in terms of nPVI and CV of pitch interval variability. Thereplication also tests how different phonetic segmentation (interactive playback) in speech will affect the outcome or the categorization. This was done to examine if the effects of nPVI and CV pitch interval are robust enough and if CV slope will resemble the same effect as earlier. In addition, findings of the CV slope experiment show that a relationship between English and French in terms of rhythmic and melodic variability and the information that can be derived for its national footprint are not as straightforward as one would like, or at least nPVI and CV of pitch interval can not provide such information. The potentiality of nPVI and pitch interval to categorize languages as stressed-timed or syllable-timed has been questioned. The example of Tamil that is a syllable-timed language and nPVI categorized it as a stressed-timed raise some critical questions on the nPVI's ability to conclude in such categorizations. The CV slope takes into account the durations of the pitch intervals provided higher scores for French than English, which raises some important questions regarding linguistic rhythmic classes. If we assume that nPVI and

CV of pitch interval can categorize languages based on national characteristics, this would mean that CV slope categorized French as stressed timed language which is something that goes against our theoretical background. This thesis underlines that attention is needed when using quantifying methods for rhythm categorization purposes and results from CV slope compromise the distinction between stress-timed and syllable-timed languages. Of the same importance is that nPVI scored higher numbers for Tamil. This is something that goes against the theory too or at least provides us with some information on what nPVI metric can not predict and this might be a rhythmic categorization of languages. CV of pitch interval had the same effect for English and French as the nPVI which can raise some questions for the pitch interval. This thesis will not argue that CV slope is a more robust method in categorizing languages. As the literature is skeptic on the nPVI and CV of pitch interval ability to make such categorizations, this thesis used both ratio of durations and pitch interval, combined them and calculated their CV slope to compare them with the nPVI and CV of pitch interval findings. The CV slope in theory seems promising because it accounts for both ratio of durations and pitch interval. A closer examination of the robustness of the CV slope can be derived if the CV slope of Tamil can be calculated so it can be compared with the theoretical findings. If CV slope result a higher score for Tamil than for English, it might possibly mean that higher score of CV slope accounts for syllable-timed languages and lower CV slope is a measurement for stress-timed languages. What we know so far is that CV slope is able to make a distinction between English and French but that it doesn't follow the effect that the replicated study showed. This thesis found that when duration is taken into consideration the effect of Patel's (2009) findings is reversed. Both language and music have rhythm and melody. This does not mean that the methods used so far to quantify rhythm and melody are applicable on language and music too. Findings of CV slope showed that the categorization of languages to stress-timed and syllable-timed is not straightforward as one would like. To conclude, this thesis introduces a new technical rationale to inspect the eligibility of the nPVI and CV of pitch interval ability to categorize languages on their ethnic footprint, which is worthy of further investigation.

8 Acknowledgments

Foremost, I would like to thank my supervisor Professor Yiya Chen for her help, support and insightful remarks. Her passion, guidance and comments helped me conduct the research and write this thesis. I couldn't have asked for a better supervisor that motivated me and broaden my understanding. I would like to thank Dr. Daan van de Velde for his help and detailed explanations at each part of this thesis. His questions, comments, suggestions and mentoring have been especially valuable. It was a great privilege working and studying under their guidance.

I would also like to thank and express my gratitude to Professor Patel for sharing his data and for answering my questions regarding the methodology and analysis.

Moreover, I would also like to thank my friends from Leiden, Amsterdam and Piraeus for their morale support, questions, brainstorming and long discussions during the Covid period. Without them this will not have been feasible.

At long last, a special thank you to my family, whose value to me only grows with time.

A Appendix: ENGLISH AND FRENCH SENTENCES

- A hurricane was announced this afternoon on the TV.
- My grandparent's neighbor's the most charming person I know.
- Much more money will be needed to make this project succeed.
- The local train left the station more than 5 minutes ago.
- The committee will meet this afternoon for a special debate.
- The parents quietly crossed the dark room and approached the boy's bed.
- This supermarket had to close due to economic problems.
- In this famous coffee shop you will eat the best donuts in town.
- This rugby season promises to be a very exciting one.
- Science has acquired an important place in western society.
- The last concert given at the opera was a tremendous success.
- In this case, the easiest solution seems to appeal to the court.
- Having a big car is not something I would recommend in this city.
- They didn't hear the good news until last week on their visit to their friends.
- Finding a job is difficult in the present economic climate.
- The library is open every day from 8 a.m. to 6 p.m.
- The government is planning a reform of the education program.
- This year's Chinese delegation was not nearly as impressive as last year's.
- The city council has decided to renovate the Medieval center.
- No welcome speech will be delivered without the press offices' agreement.
- Les parents se sont approchés de l'enfant sans faire de bruit.
- Cette boulangerie fabrique les meilleurs gâteaux de tout le quartier.
- La femme du pharmacien va bientôt sortir faire son marche.
- Les voisins de mes grandparents sont des personnes très agreeables.
- Il faudra beaucoup plus d'argent pour mener à bien ce projet.
- Le magasin est ouvert sans interruption toute la journée.
- Les mères sortent de plus en plus rapidement de la maternité.
- L'été sera idyllique sur la côte méditerranéenne.
- Ils ont appris l'évènement au journal télévisé de huit heures.
- La nouvelle saison thé trale promet d'être des plus intéressante.

- Un tableau de très grande valeur a été récemment dérobé.
- Le plus rapide est encore le recours auprès de la direction.
- Les récents événements ont bouleversé l'opinion internationale.
- Le train express est arrivé en gare ilyamaintenant plus de 5 minutes.
- La reconstruction de la ville a commencé après la mort du roi.
- L'alcool est toujours la cause d'un nombre important d'accidents de la route.
- Aucune déogation ne pourra être obtenue sans l'avis du conseil.
- Les banques ferment particulièrement tôt le vendredi soir.
- Trouver un emploi est difficile dans le contexte économique actuel.
- Le ministère de la culture a augmenté le nombre de ces subventions.

B Appendix: COMPOSERS AND MUSICAL THEMES

English: **Bax** b508, b509, b510, b511, b515, b517, b518, b519, b520. **Delius** d189, d191, d192, d193, d194, d195, d196, d197, d198, d199, d200, d201, d202, d205, d208, d214, d215, d216, d219. **Elgar** e3, e4, e7, e8, e13, e14, e15, e16, e17, e18, e19, e20, e21, e23, e27, e28, e30, e31, e33, e34, e35, e51, e52, e53, e56, e58, e60, e61, e62, e63, e64, e66, e67, e68, e70, e71, e72, e73a, e73b, e73c, e73d, e73f, e73h, e73i, e73j. **Holst** h798, h799, h801, h803, h804, h805, h806, h807, h810, h811, h813, h814, h817, h818, h819, h820. **Ireland** i95, i97, i98, i102, i104, i105, i109, i110, i111, i112, i113. **Vaughan Williams** v4, v5, v6, v7, v8, v12, v13, v14, v17, v18, v19, v20, v21, v22, v23, v24, v26, v27, v28, v29, v30, v31, v32, v33, v34, v35, v37, v38, v39, v40, v41, v42, v43, v44, v45, v49.

French: **Debussy** d13, d14, d20, d21, d42, d43, d55, d57, d58, d62, d70, d71, d74, d77, d78, d80, d83, d85, d86, d87, d88, d90, d97, d98, d100, d105, d107, d108, d109, d113, d116, d117, d118, d122, d123, d124, d125, d126, d127, d129, d132, d134, d135, d138, d139, d140. **Faur´e** f60, f61, f62, f63, f72, f75, f76, f76d, f77, f78, f79, f80, f84, f85, f87, f89, f91, f92, f93, f94, f95, f97, f98, f101, f102, f103, f104, f105. Honegger h830, h832, h833, h834, h836, h842, h843, h844. **Ibert** i1, i3, i4, i6, i8, i13, i14, i24, i26, i27. **D'Indy** i31, i33, i40, i41, i42, i44, i47, i48. **Milhaud** m382, m383, m384, m386, m387, m394, m395. **Poulenc** p170, p171, p176, p177, p178. **Ravel** r124, r128, r129, r130, r132, r133, r147, r148, r150, r151, r152, r153, r154, r155, r156, r183, r184, r186. **Roussel** r407, r409, r410, r411, r412, r416, r417, r419, r420, r422, r423. **Saint-Sa¨ens** s18, s20, s21, s22, s26, s31, s32, s33, s34, s35, s36, s40, s42, s49, s50, s66, s69, s77, s79, s89, s92, s98, s99, s100, s102, s103, s104, s105, s106, s107, s108, s109, s110, s112, s114, s127, s129, s133, s134.

C Appendix: SLOPE CALCULATION: AN EXAMPLE

Example of B509 from the corpus.

The interval in absolute numbers of the first and second note is 3. The duration of the first note is 1. In order to compute the slope of this pair we divide the pitch interval (3) with the duration of the first note of that pair (1). The result for this one is: 3 divided by 1, equals 3. The slope of this note pair is 3. This computation was performed for all the note or vowel pairs in the corpus. The coefficient of variation per sentence and musical theme was calculated and the average of that resulted the CV for speech slope and music slope for English and French.

Interval of 1st note and 2nd note = 3

Duration of 1st note = 1

3/1 = 3

The slope for this note pair is 3.

References

Aalto, Daniel, Juraj Simko, and Martti Vainio (2013). “Language background ˇ affects the strength of the pitch bias in a duration discrimination task”. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH.

Aramaki, Mitsuko et al. (2005). “Comparison of Rhythmic processing in Language and Music : An interdisciplinary approach”. In: Journal of Music and Meaning 3.

Arbib, Michael et al. (2013). Language, Music, and the Brain: Probing a Mysterious Relationship.

Arvaniti, Amalia (2010). “The usefulness of metrics in the quantification of speech rhythm”. In: Journal of Phonetics 40.3. issn: 00954470. doi: 10 . 1016/j.wocn.2012.02.003.

Besson, Mireille and Daniele Sch¨on (2001). “Comparison between language and music”. In: Annals of the New York Academy of Sciences. Vol. 930. doi: 10.1111/j.1749-6632.2001.tb05736.x.

Besson, Mireille, Daniele Sch¨on, et al. (2007). Influence of musical expertise and musical training on pitch processing in music and language.

Bolinger, D and D. L. M. Bolinger (1989). “Intonation and its uses: Melody in grammar and discourse”. In: Stanford university press.

Burns, Edward M. (1999). “Intervals, Scales, and Tuning”. In: The Psychology of Music. doi: 10.1016/b978-012213564-4/50008-1.

Cummins, Fred (2012). “Looking for Rhythm in Speech”. In: Empirical Musicology Review 7.1-2. issn: 1559-5749. doi: 10.18061/1811/52976.

Daniele, Joseph R. and Aniruddh D. Patel (2013). “An empirical study of historical patterns in musical rhythm: Analysis of German & Italian classical music using the nPVI equation”. In: Music Perception 31.1. issn: 07307829. doi: 10.1525/MP.2013.31.1.10.

— (2015). “Stability and change in rhythmic patterning across a composer’s lifetime: A study of four famous composers using the nPVI equation”. In: Music Perception 33.2. issn: 15338312. doi: 10.1525/MP.2015.33.2.255.

Darwin, Charles (1871). “The descent of man and selection in relation to sex, in Charles Darwin, The origin of species and The descent of man (combined volume)”. In: Journal of anatomy and physiology 5.Pt 2. issn: 0028-0836.

Fishman, Joseph P. (2018). “Music as a matter of law”. In: Harvard Law Review 131.7. issn: 0017811X.

Grabe, Esther and Ee Ling Low (2002). “Durational variability in speech and the Rhythm Class Hypothesis”. In: Laboratory Phonology 7. doi: 10.1515/ 9783110197105.

Gussenhoven, Carlos and Wencui Zhou (2013). “Revisiting pitch slope and height effects on perceived duration”. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH.

Hall, R. A. (June 1953). Elgar and the intonation of British English. The Gramophone.

Hanson, Josef (2017). Rhythmic variability in language and music of Latino and Latino-inspired composers. doi: 10.1525/MP.2017.34.4.482.

Hausen, Maija et al. (2013). “Music and speech prosody: A common rhythm”. In: Frontiers in Psychology 4.SEP. issn: 16641078. doi: 10.3389/fpsyg. 2013.00566.

Honing, Henkjan (2010). “Lure(d) into listening: The potential of cognitionbased music information retrieval”. In: Empirical Musicology Review 5.4. issn: 1559-5749. doi: 10.18061/1811/48549.

House, D.R. (1990). “Tonal perception in speech”. In: Jackendoff, Ray (2009). “Parallels and nonparallels between language and music”. In: Music Perception. Vol. 26. 3. doi: 10.1525/mp.2009.26.3.195.

Jones, Mari Riess, Marilyn Boltz, and Gary Kidd (1982). “Controlled attending as a function of melodic and temporal context”. In: Perception & Psychophysics 32.3. issn: 00315117. doi: 10.3758/BF03206225.

Koelsch, Stefan et al. (2003). “Children processing music: Electric brain responses reveal musical competence and gender differences”. In: Journal of Cognitive Neuroscience 15.5. issn: 0898929X. doi: 10.1162/jocn.2003.15.5.683.

Kotz, Sonja A. and Michael Schwartze (2010). “Cortical speech processing unplugged: A timely subcortico-cortical framework”. In: Trends in Cognitive Sciences 14.9. issn: 13646613. doi: 10.1016/j.tics.2010.06.005.

Ladd, D. R. (2008). “Intonational phonology”. In: Cambridge University Press.

Ladd, D. Robert and Dwight Bolinger (1987). “Intonation and Its Parts: Melody in Spoken English”. In: Language 63.3. issn: 00978507. doi: 10 . 2307 / 415013.

Magdics, Kl´ara and Klara Magdics (1963). “From the Melody of Speech to the Melody of Music”. In: Studia Musicologica Academiae Scientiarum Hungaricae 4.3/4. issn: 00393266. doi: 10.2307/901395.

Mertens, Piet (2004). “The Prosogram: Semi-Automatic Transcription of Prosody Based on a Tonal Perception Model”. In: Proceedings of the 2nd International Conference on Speech Prosody. — (2020). “Prosogram user’s guide”.

Patel, Aniruddh D. (2012). Music, Language, and the Brain. doi: 10.1093/acprof:oso/9780195123753.001.0001.

Patel, Aniruddh D. and Joseph R. Daniele (2003). “An empirical comparison of rhythm in language and music”. In: Cognition 87.1. issn: 00100277. doi: 10.1016/S0010-0277(02)00187-7.

Patel, Aniruddh D., Edward Gibson, et al. (1998). “Processing syntactic relations in language and music: An event-related potential study”. In: Journal of Cognitive Neuroscience 10.6. issn: 0898929X. doi: 10.1162/089892998563121.

Patel, Aniruddh D., John R. Iversen, and Jason C. Rosenberg (2006). “Comparing the rhythm and melody of speech and music: The case of British English and French”. In: The Journal of the Acoustical Society of America 119.5. issn: 0001-4966. doi: 10.1121/1.2179657.

Paul, Boersma and Weenink David (2020). “Praat: doing phonetics by computer [Computer program]. Version 6.1.37, retrieved 19 December 2020 from http://www.praat.org/”.

Peterson, Gordon E. and Ilse Lehiste (1960). “Duration of Syllable Nuclei in English”. In: The Journal of the Acoustical Society of America 32.6. issn: 0001-4966. doi: 10.1121/1.1908183.

Pierrehumbert, Janet (1980). The phonology and phonetics of English intonation. doi: 10.1177/003368828401500113. Pike, K. L. (1945). “The Intonation of American English”.

Ramus, Franck (2002). “Acoustic correlates of linguistic rhythm: Perspectives”. Proceedings of Speech Prosody 2002.

Ramus, Franck, Marina Nespor, and Jacques Mehler (1999). "Correlates of linguistic rhythm in the speech signal". In: Cognition 73.3. issn: 00100277. doi: 10.1016/S0010-0277(99)00058-X.

Rousseau, Jean-Jacques (1998). "Essay on the Origin of Languages [1781]". In: Essay on the origin of languages and writings related to music.

Sch¨on, Daniele, Cyrille Magne, and Mireille Besson (2004). "The music of speech: Music training facilitates pitch processing in both music and language". In: Psychophysiology 41.3. issn: 00485772. doi: 10.1111/1469-8986.00172.x.

Seeger, Charles et al. (1948). "A Dictionary of Musical Themes". In: Notes 5.3. issn: 00274380. doi: 10.2307/890449.

Thompson, William Forde (1994). "Sensitivity to combinations of musical parameters: Pitch with duration, and pitch pattern with durational pattern". In: Perception & Psychophysics 56.3. issn: 00315117. doi: 10.3758/BF03209770.

Trehub, Sandra E. (2003). The developmental origins of musicality. doi: 10. 1038/nn1084. UNESCO (2021). Browse the Lists of Intangible Cultural Heritage and the Register of good safeguarding practices. Retrieved 15 August 2021, from https://ich.unesco.org/en/lists#2017.

VanHandel, Leigh (2017). "The War of the Romantics: An Alternate Hypothesis Using nPVI for the Quantitative Anthropology of Music". In: Empirical Musicology Review 11.2. issn: 1559-5749. doi: 10.18061/emr.v11i2.5474.

VanHandel, Leigh and Tian Song (2010). "The role of meter in compositional style in 19th century French and German art song". In: Journal of New Music Research 39.1. issn: 09298215. doi: 10.1080/09298211003642498.

Vukovics, Katherine and Daniel Shanahan (2020). "Rhythmic variability, language, and style: A replication and extension of nPVI findings with the RISM dataset". In: Journal of New Music Research 49.3. issn: 17445027. doi: 10.1080/09298215.2020.1751209.

Wells, Rulon S. and Kenneth L. Pike (1947). "The Intonation of American English". In: Language 23.3. issn: 00978507. doi: 10.2307/409880.

Zheng, Yi and Arthur G. Samuel (2018). “The effects of ethnicity, musicianship, and tone language experience on pitch perception”. In: Quarterly Journal of Experimental Psychology 71.12. issn: 17470226. doi: 10.1177/1747021818757435.