DOI: 10.17469/O2106AISV000021

MARTINA TURCONI, GIOVANNA MAROTTA

Speech and music rhythm: A pilot study on English and Italian songs[1]

Instrumental and vocal compositions agree with the mother tongue rhythm of the composer, proving that the same rhythmic categorization operates on both linguistic and music levels. In this study we aimed to verify how strong the mother tongue rhythm of the singers is in relation to the mother tongue rhythm of the composers, and if this is also able to influence the music rhythm of their songs. The results of the analysis of original English and Italian songs along with their foreign versions suggest that on the music level the *stress-timed* rhythm exerted a greater influence than the *syllable-timed* one, being able to influence the music rhythm of the Italian version of the English songs. On the contrary, on sung level both timing categories seem to be equally strong.

Key words: rhythm, sung speech, music rhythm, timing.

1. *Preliminaries*

The notion of "rhythm" is intrinsic to several areas of natural phenomena and human activity (e.g. wave sounds, the flow of the tide, and all those human activities that make use of rhythmic pattern, such as speaking and singing, but also playing music, dancing, ice-skating and so on). The English term is derived from the Greek word ῥυθμός (*rhythmós*), cognate with ῥέω (*réo*), which means "to flow". It was originally used by Ionic authors to mean "schema", whereas in the Attic age, Plato gave it a new meaning: "rhythm is order in movement" (Law, II book: 664e-665a). This definition immediately evokes a sense of periodicity, hierarchic regularity, and dynamic recurrences of durations. However, "rhythm is more than mere clock time, the invariant sequence of evenly spaced intervals, and yet of such regularity is rhythm born" (Cummins, 2015: 158). It is in music that the concept of rhythm in this sense, is most at home. In fact, a stable underlying beat is almost always present in music, and this allows for the organization of periodicity on multiple time scales in order to create musical meter (see Patel, Daniele, 2003a) and, undoubtedly, speech also involves dynamic recurrences of durations. Are the principles that drive rhythm in speech and music the same, or do they change?

This question has given rise to a number of studies that have aimed to uncover the underlying principles of rhythmical manifestations. One such principle is that both in music and linguistics, rhythm refers to patterns of duration (see London,

[1] Even though the article has been jointly discussed, the authorship of the paragraphs is intended to be divided as follows: M. Turconi § 1, 3, 4.1, 4.2, 5; G. Marotta § 2, 4.3, 6.

2001). Furthermore, each domain shares the rhythmic configurations of the *grouping structure*: both tones and words are grouped into phrases (see Selkirk, 1984; Nespor, Vogel, 1986, 2007), in "a 'configurational' patterning of stressed vs. unstressed syllables" (Patel, Daniele, 2003a: 140).

In the last decades of the Eighteen century, Joshua Steele (1775) applied the musical theory to the analysis of the English language, in an attempt to explore the common belief that English rhythm tends to form sequences of regular beats: "there is a general tendency to make the stress – points of stressed syllables follow each other at equal intervals of time, but [...] this general tendency is constantly interfered with by the variations in the number and nature of the sounds between successive stress – points" (Steele, 1775). The first person to attempt to empirically prove this rhythmical intuition was Classé (1939). Using a kymograph, he was able to represent the intensity wave form of English speech. The results demonstrated a tendency to isochronism and Classé hypothesized that the perceptive system might also play a key role in the perception of unequal intervals as being even. A few years later, Pike (1945) published an investigation on the intonation of American English, where he identified two rhythmic dichotomous units: *stress-timed* and *syllable-timed* rhythmic units. This pattern of temporal organization was then extended to all the natural languages by Abercrombie (1967). A third rhythmical category was later postulated: the *mora-timed* class. This class takes as its rhythmic unit the mora, which recurs in isochronous intervals in languages such as Japanese (Bloch, 1950; Han, 1962; Ladefoged, 1975; Nespor, Shukla & Mehler, 2011).

In the last part of the XX century, several research studies were conducted to confirm this timing classification. Lehiste (1977), Dauer (1983), and Bertinetto (1989) to name but a few, proved through various methods of analysis that the postulated isochronism did not correspond to reality, but that it is the perceptive mechanism that influences us and leads us to state that the speech rhythm units occur in isochronous intervals. Lehiste (1977) demonstrated that there is a perceptive threshold, that ranges from 30-100 milliseconds, within which a difference in duration can be perceived. If the interval is shorter than 30 milliseconds, it cannot be discriminated, and it can be considered equivalent to either a larger or smaller interval. Isochrony is, after all, a purely perceptual phenomenon. Later, Bertinetto (1989), who analyzed two Italian speakers, and Dauer (1983), who measured interstress intervals from readings of texts in English, Thai, Spanish, Italian, and Greek, both argued that it might not be possible to categorize the languages into two opposite and discrete rhythmical categories, but that they might rather be distributed along a continuum, whose poles are the pure stress-timed language and the pure syllable-timed language, which, actually, do not exist. "In ordinary speech and everyday prose [...in fact] the rhythmic effect is a purely automatic consequence of linguistic circumstances" (Dauer, 1983: 60).

The linguistic scenario drastically changed at the turn of the last century. Nazzi, Bertoncini, and Mehler (1998) had proven that newborns could discriminate two languages if they belonged to two different rhythmic classes (for example they could discriminate English from Japanese, but not English from Dutch). Given those results, Ramus, Nespor & Mehler (1999) reviewed the timing model of classification, stating that feet, syllables, and morae were not the most suitable rhythmic units, but that languages could be rhythmically classified according to the comparison of the proportion of vocalic intervals within a sentence (*%V*), and the standard deviation of the vocalic and intervocalic intervals within a sentence (*ΔV* and *ΔC)*. Indeed, through these metrics the authors could demonstrate that stress-timed languages were characterized by more complex syllable structures and greater vowel reduction than syllable-timed ones.

Later on, other researchers proposed new coefficients in order to normalize the speech rates not only between different languages, but also different speakers of the same language, as with the *Pairwise Variability Index* of Grabe and Low (2002), or the *varco* coefficient of variation. This latter variation was applied by Dellwo, Wagner (2003) and Dellwo (2006) to improve Ramus' coefficients. In their paper, Grabe & Low (2002) took into consideration many more languages than ever before and the results proved the continuum hypothesis of Dauer (1983) and Bertinetto (1989). In fact, many of these languages, such as Greek, Welsh, and Polish, could not be classified either as stress-timed (e.g. English, German or Dutch) or syllable-timed (e.g. Romance languages) because of their 'mix' character.

A few years later, Bertinetto, Bertini (2008) proposed a new model of rhythmic classification capable to capture the rhythmic tendencies of natural languages: the *Control/Compensation Hypothesis*. Based on the consideration that all languages have syllables and prominence contrasts, the authors agreed that what makes the difference, in rhythmic terms, is the prosodic structure built on this common ground. Indeed, the *Control/Compensation Index* (*CCI*) "aimed at describing the intra-syllabic behavior, which in turn affects (or is possibly affected by) the overarching accentual alternation" (Bertinetto, Bertini, 2008: 427).

Even though new recent perceptual experiments have confirmed the rhythmic classification of languages based on quantitative measurement methods (e.g. Hannon, 2009), part of the scientific community pointed out not only the inability of all metrics to successfully and unequivocally classify the rhythmically non-prototypical languages, such as Greek and Polish (see Grabe, Low, 2003), but also that metrics "are very sensitive to the effects that methodological choices have on the durations of consonantal and vocalic intervals" (Arvaniti, 2012: 354) and that the inter-speaker variability in timing is more extensive than previously suspected, mainly in the spontaneous speech. Thus, rethinking the validity both of metrics and rhythm theories, Tilsen, Arvaniti (2013) initiated a different approach to analyse speech rhythm that was not dependent on syllable structure. Using the technique known as *empirical mode decomposition* (*EMD*, see Huang, Shen, Long, Wu, Shih, Zheng, Yen, Tung & Liu, 1998)[2], a spectral analysis of the envelope of speech waveform, they attempted to identify rhythmic stretches of speech and compare rhythmic patterns across languages. The envelope metrics[3] the au-

[2] "This procedure extracts non-orthogonal basis functions from the envelope, each of which captures oscillations on a different time-scale" (Tilsen, Arvaniti, 2013: 629).

[3] Tilsen and Arvaniti (2013) elaborated three types of envelope metrics: the *power distribution metrics* are able to capture the relative power in syllabic vs. supra-syllabic oscillations in the envelope; the *rate*

thors used seem to be able to dissociate rate from power distributions, capturing different information from amplitude envelopes, and to exhibit weak dependence on syllable structure, being less dependent on syllable phonotactics than interval metrics. For these reasons Tilsen, Arvaniti (2013) judge their metrics to be an alternative approach to describe speech rhythm. Moreover, the investigation of other acoustic dimensions, such as pitch other than duration, could help in understanding the cognitive mechanisms of rhythm in speech production and perception.

Also, Brow and colleagues (2017) argue that the durational coefficients describe just the surface level of rhythm organization, not providing enough information about the relative duration of the syllables: "they are not able to specify the actual rhythm of any given utterance within a language" (Brown, Pfordresher & Chow, 2017: 97). In order to fill these gaps, Brown and colleagues (2017) proposed an effective but upgradeable musical model capable to predict the rhythmic representations of a sentence. In contrast with the previous studies, the authors recovered one of the characteristics of musical meter to analyse speech rhythm: although meter often remain stable across periods of time, it can occasionally change creating *heterometric sentences* (sentences where a change in focus-word changes the sentence rhythm)[4], which could well represent the complexity of speech timing along with *isodurational* (stable meter and invariant syllable durations, such as the nursery rhyme "Twinkle Twinkle little star. How I wonder what you are") and *isometric sentences* (stable meter but variable timing of syllables within measures, such as the sentence "Pamela purchased beautiful flowers Saturday morning all through the year", in which there is an alternation between 3-syllable and 2-syllable groupings). Comparing the timing of *prominence groups*[5] between an a priori rhythmic transcriptions of a set of sentences and their spoken versions by calculating the *coefficient of variation* (*CV*)[6], the authors found out that "most of the transcriptions were borne out by the productions, suggesting that metricality in speech can be measured reliably" (Brown et al., 2017: 108). In fact, musical transcriptions, thanks to its metrical structure, can be adapted to all the different realizations of a given sentence.

2. *Rhythm in music and speech*

At the very same time, some scholars returned to the roots of the timing debate, rediscovering the powerful link that connects speech and music. In fact, the quantitative measurement methods used in the linguistic research on rhythm were successfully applied by Patel, Daniele (2003a) to the analysis of musical compositions,

---

metrics, such as w_1 and w_2, register the frequencies of those oscillations; and the *rhythmic stability metrics* capture the stability of oscillations.

[4] For example, in the sentence "Miguel bought two yellow shirts at the men's store by the bay" either "yellow" or "two" could be accented, changing the meter of the sentence.

[5] The authors' notion of prominence group is similar to the concept of an 'inter-stress interval' that can be found in previous research on speech rhythm, such as Dauer (1983), Cummins, Port (1998).

[6] The *coefficient of variation* is the Standard Deviation of normalized prominence groups (Brown et al., 2017).

in order to verify if the mother tongue rhythm of a composer could influence the music rhythm of their compositions. They considered the *Pairwise Variability Index* (by Grabe, Low, 2002) the most suitable variable for musical corpora, and, using this, were able to quantify the degree to which successive units differ in duration. Therefore, to analyze music rhythm variability, the unit to be taken into consideration should be the note. After having computed all the variables on their instrumental music corpus[7] and having compared the results with the data of English and French speech by Ramus (2002), they were able to affirm that the mother tongue rhythm of the composers overlapped with the music rhythm of their compositions.

This pioneering study generated an interesting debate as well as a new branch of studies which aimed to prove these findings in extended corpora (see Huron, Ollen 2003; Patel, Daniele, 2003b; and London, Jones, 2011). In recent years, research has addressed new linguistic areas (e.g. Polish, see Jekiel, 2014) and new musical genres such as blues (Larroque, 2012), popular music and Japanese vocal music (Sadakata, Desian, Honing, Patel & Iversen, 2004), which all appear to confirm Patel, Daniele (2003) hypothesis.

Our study proceeds along the path traced by the studies of Patel, Daniele (2003) and fellow musicologists (Huron, Ollen, 2003; Patel, Daniele, 2003b; London, Jones, 2011) as well as linguistic studies on speech rhythm (Jekiel, 2014; Larroque, 2012; Sadakata et al., 2004). Firstly, we aim to prove that, even in modern vocal compositions, the mother tongue rhythm of the composer influences the music rhythm of their songs. Moreover, we want to verify how strong the mother tongue rhythm of the singers is in relation to the mother tongue rhythm of the composers, and if this is also able to influence the music rhythm of their songs. To do so we decided to resort to a particular type of vocal composition: original songs that had been rearranged in a foreign language version. Therefore, we considered also to attempt to analyse the rhythm of the same song sung in two different language types. Indeed, using the linguistic quantitative measurement methods proposed by Grabe, Low (2002), and Dellwo, Wagner (2003), we were able to analyse the *sung speech rhythm*, which can be considered as the timespace where music and speech rhythm are fully involved in each other[8]. Assuming that general rhythmic principles drive the rhythmic motor behavior of human beings on multiple levels, and since it has been demonstrated that there is a "significant phase-locked neural response to rhythm in both music and speech domains" (Harding, Sammler, Henry, Large & Kotz, 2019: 98), we argue that the rhythm of sung speech could also be classified along with the traditional rhythmic continuum, as it was postulated by Dauer (1983) (see §1).

[7] Patel and Daniele (2003a) selected English and French compositions written by author lived between the Nineteenth- and Twentieth-century. English composers: A. Bax, F. Delius, J. Ireland, E. Elgar, R.V. Williams and G. Holst; French composers: C. Debussy, V. D'Indy, G. Fauré, A. Honneger, J. Ibert, D. Milhaud, F. Poulenc, M. Ravel, A. Roussel and C. Saint-Saëns.

[8] During the second part of the XX century some of the characteristics of the singing voice were investigated, e.g. melody structure by Sundberg, Lindblom (1976), vowel quality and consonants palatalization by Oussilova (1995), and spectral modifications of French vowels by Florig (1995).

3. *Corpus*

Focusing on a type of musical composition that exhibits a robust relationship between speech and music rhythm (i.e. songs), we decided to analyze vocal music as opposed to instrumental music for its tight relationship between speech and music rhythm. The choice of the language was guided by the aim of having two languages with different timing classifications. Therefore, we opted for American English, because it has been traditionally classified as a stress-timed language, and Italian, a syllable-timed language. Regarding the music period, we chose songs from the 60s and the 70s, which allowed us to select original songs along with their foreign version. In those years the practice of creating a new original version of a well-known foreign song, while maintaining an almost unaltered melodic line, was very popular. Hence, we chose three Italian and three English songs and their corresponding foreign version. The list is presented in Table 1:

Table 1 - *Italian and English songs and their corresponding version*

Original Italian songs	**English versions**
Azzurro – Adriano Celentano (1968)	*Blue skies (Azzurro)* – Ken Dodd (1970)
Mi ritorni in mente – Lucio Battisti (1969)	*Wake me, I am dreaming* – The Love Affair (1971)
Balla Linda – Lucio Battisti (1968)	*Bella Linda* – The Grass Roots (1969)
Original English songs	**Italian versions**
Stand by me – Ben E. King (1961)	*Pregherò* – Adriano Celentano (1962)
I'm a believer – Niel Diamond (1966)	*Sono bugiarda* – Caterina Caselli (1967)
Elenore – The Turtles (1968)	*Scende la pioggia* – Gianni Morandi (1968)

All the songs were performed in an "Andante" tempo, neither too fast nor too slow, and the performance didn't overly alter the temporal organization of the songs.

It was easy to find a number of recordings of all the songs of our *corpus*, so we selected the best audio track from the ones available[9]. On the contrary, it was difficult to source a printed score for some of the songs, so we decided to transcribe the missing scores from the recording ourselves. We also noticed that some official tran-

[9] Italian songs: L. Battisti, *Mi ritorni in mente*, "*RememberingItaly*", track 15, © 2016 Codex, 2016; L. Battisti, *Balla Linda*, "*I Successi di Lucio Battisti*", track 8, © 1998 DV More, 2009; C. Caselli, *Sono bugiarda*, "*Qualcuno mi può giudicare*", track 8, © 1997 Sugar Srl, 2006; A. Celentano, *Azzurro*, "*…Adriano*", disk 3, track 11, © 2013 Clan Celentano Srl, 2013; A. Celentano, *Pregherò*, "*Milva, Mina & More from Italy*", track 5, © 2013 OVC Media Limited, 2013; G. Morandi, *Scende la pioggia*, "*C'era un ragazzo*", track 11, 2002. English songs: N. Diamond, *I'm a believer*, "*The Bang Years 1966-1968*", track 7, Universal Music Italia Srl, 2011; K. Dodd, *Blue skies*, "*All the songs I love*", disk 2, track 7, Parlophone Records Ltd, 2005; B.E. King, *Stand by me*, "*Stand by me*", track 1, The Web Engine, 2016; The Grass Roots, *Bella Linda*, "*Greatest Hits: The Grass Roots*", track 9, © 2012 One Media iP Ltd, 2012; The Love Affair, *Wake me, I am dreaming*, https://www.youtube.com/watch?v=2T9pmRzRPl4 [21/03/2019]; The Turtles, *Elenore*, "*The turtles present the Battle of the Bands*", track 3, © 1968 Flo & Eddie, Inc, 2005.

scriptions, such as "Balla Linda" by Lucio Battisti, adopted the common normalized version for the "swing" style: instead of writing the song in triplets, as it should be performed, the transcribers wrote it in a binary beat division. Therefore, we had to create a new written version using a ternary beat division so that the rhythm of the score adhered, as far as possible, to the rhythm of our audio track[10].

4. *Analysis of the music rhythm*

4.1 Method

Following the method proposed by Grabe, Low (2002) and adopted by Patel, Daniele (2003a), we considered the *Pairwise Variability Index* (PVI)[11] to be the most suitable variable to analyze a musical corpus. According to Patel, Daniele (2003a), in fact, the value of a note is easily comparable to a vocal interval. Given that, we decided to calculate the *normalized PVI* (*nPVI*) (see Figure 1) of the melodic lines from the scores to stay as much as possible in line with the method previously used by the authors who conducted similar studies (Patel, Daniele, 2003a and London, Jones, 2011, among others), and in order to compare their results with ours, especially for the English values. Since the scores played such an important role in our analysis, we decided to rewrite the ones that couldn't reflect the exact rhythm of the acoustic performance (see § 2).

Figure 1 - *Normalized Pairwise Variability Index (Grabe, Low, 2002)*

$$n\,PVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right]$$

The variability index was computed for the first verse and the first refrain/chorus, where it was possible. Some foreign versions didn't use the entire original song: "Wake me I am dreaming" was composed only on the musical line of the verse of "Mi ritorni in mente"[12]; others were performed so freely compared with the scores, that we were forced to discard those parts of the score that was rhythmically too far from the audio track.

We calculated the *nPVI* assigning an arbitrary duration of 1 to the first note of each selected theme; consequently, the durations of the remaining notes were expressed as multiples or submultiples of the value of the first note. When a rest was part of the theme (in this musical genre it is a common characteristic), we didn't

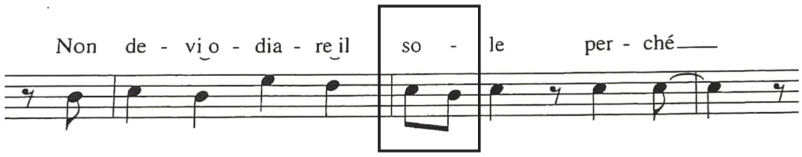
[10] See *Appendix*.

[11] The *Pairwise Variability Index* expresses the level of durational variability in successive measurements. The *normalized* version of *PVI* "is compiled by calculating the difference in duration between each pair of successive measurements, taking the absolute value of the difference and dividing it by the mean duration of the pair" (Grabe, Low, 2002: 523).

[12] See *Appendix*.

take into consideration the value of the rest in the algorithm (as in London, Jones 2011). On the contrary we did take into consideration the durations of all the notes making up the music line, even the ones belonging to the so called “melismatic configurations” (see Johnson, Huron & Collister, 2014)[13]. An example of such melismatic configuration is given in Figure 2: the syllable “so-” of the word “sole” in the song “Pregherò” by Adriano Celentano.

Figure 2 - *Example of a melismatic configuration*

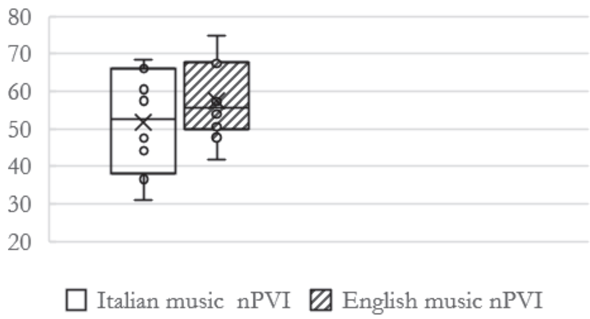


Lastly, to statistically prove our results, we carried out a *t-test analysis*.

4.2 Results

Figure 3 presents the box and whisker plot of the data collected: the average, the median and the range of the measurement of the Italian and English songs. Despite the wide-ranging results, it seems that there is a difference between Italian and English values.

Figure 3 - *Musical* nPVI *values for Italian (in white) and English songs (in black stripes)*



But how reliable is this account? The descriptive statistics of Table 2 as long as the conducted *t-test* (t(16.577) = -1.0713, p = 0.299) do not prove such a difference between Italian and English music *nPVI*.

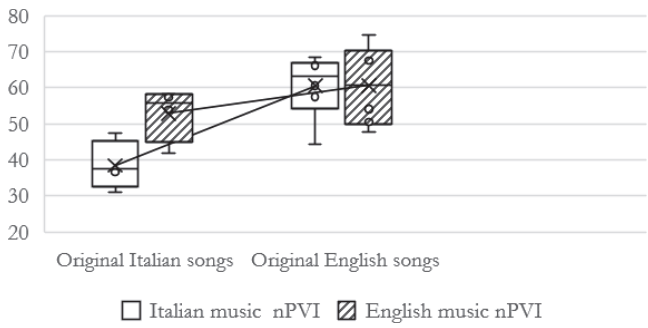
[13] In vocal compositions there are two types of note configuration: melismatic configurations and syllabic ones. In this second type, one single note is assigned one syllable, whereas in the melismatic configuration, each syllable can be sung as two or more notes.

Table 2 - *Mean, median and standard deviation of the Italian and English music nPVI, and their respective minimum and maximum values*

	Italian music nPVI	English music nPVI
Mean	51.68	57.48
Median	52.47	55.71
Std. Deviation	13.79	10.20
Minimum	31.04	41.97
Maximum	68.49	74.77

Looking in more depth at the data, the *nPVI* of the original Italian songs shows a noticeable difference from the original English songs *nPVI* (see Figure 4). But, if we look at the two groups, "Original Italian songs" and "Original English songs", we can see a different trend. The Italian versions of the original English songs do not show, as expected to be, a syllable-timed rhythm: the *nPVI* values (distributed in the white box at the right of Figure 4) suggest that they are carachterised by a stress-timed rhythm, as their English counterparts. Indeed, the difference between Italian orginal English songs and English original English songs is not statistically proven ($t(9.5609) = -0.00911$, $p = 0.992$), whereas it is proven the difference between the Italian original Italian songs and the Italian original English songs ($t(7.7364) = -4.399$, $p = 0.002$). On the other side, the English versions of the original Italian songs do not follow the syllable timing of Italian, mantaining their native stress-timed speech rhythm ($t(5.9447) = -2.8403$, $p = 0.029$).

Figure 4 - *The box and whisker plot shows the* nPVI *results for the first subgroup "Original Italian songs" on the left and the "Original English songs" on the right*



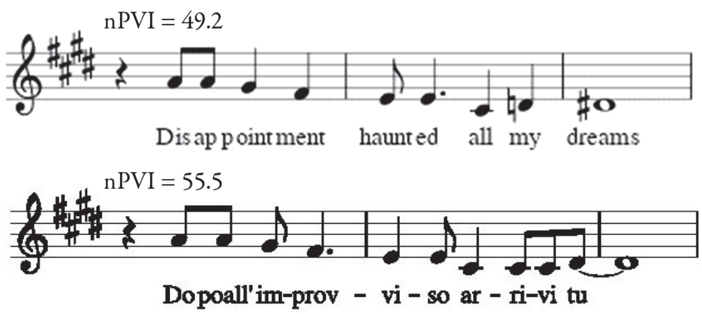
4.3 Discussion

The data collected allow us to state that, in our corpus, the rhythm of the English songs tends towards the stress-timed typology, because of the higher *nPVI* values. The situation regarding the Italian song rhythms, however, is much more complicated. In fact, with regard to the original Italian compositions (mean 38.45; Standard Deviation 6.8), the music rhythm tends towards the syllable-timed typology, but it does not do so for the Italian versions of the original English songs (mean 60.5;

Standard Deviation 8.9). It seems that only the English mother tongue rhythm is able to influence the rhythmic structure of the melodic line, thus making it more variable. Conversely, the Italian speech rhythm does not manage to reduce the variability of the rhythmic pattern, actually making it, at times, even more complex.

An example of this increased complexity can be found in the Italian version "Sono bugiarda" of the original English song "I'm a believer" by Niel Diamond, in which not only the rhythmic melodic structure is changed, but also the musical and linguistic metric one. Musically, the rhythmic structure of the English verse has a lower *nPVI*, and therefore the rhythm is less variable. In Figure 5 we present the verse "Disappointment haunted all my dreams" from the original English song "I'm a believer" by Niel Diamond at the top, followed by the verse "Dopo all'improvviso arrivi tu" from its Italian version "Sono bugiarda" by Caterina Caselli.

Figure 5 - *Comparison between two verses*



In the metric structure, the second accent falls on the beat in the English verse, whereas in the Italian one, it is anticipated on the upbeat through a syncope, which greatly alters the metric structure:

- Italian verse: /do.poal.lim.prov.ˈ**vi**.zo.ar.ri.vi.ˈ**<u>tu</u>**/
- English verse: /di.sə.pɔɪnt.mənt.ˈ**hɔ:n**.tɪd.ɔ:l.maɪ.ˈ**<u>dri:mz</u>**/[14]

5. *Analysis of sung speech rhythm*

The speech data regarding English speech and Romance languages (Ramus et al., 1999; Grabe, Low, 2003; Dellwo, Wagner, 2003) and the music data (Patel, Daniele 2003a) suggest that the same rhythmic categories operate not only on the linguistic level, but also on the musical one. Our data also support the hypothesis that the mother tongue rhythm of the composer almost always overlaps with the music rhythm of the composition.

What happens when the speech rhythm interacts with the music rhythm? How does speech rhythm behave when it is sung? Does the mother tongue rhythm of

[14] The accented syllables are written in bold letters, and the syllables where the second musical accent falls is underlined.

the singer overlap with the music rhythm of the songs? We have attempted to answer these questions by analysing the sung speech rhythm of our corpus. In order to analyze the rhythm of the sung speech we adopted two variables: the *normalize Pairwise Variability Index* by Grabe, Low (2003) and the *varcoΔV* and *varcoΔC* by Dellwo, Wagner (2003)[15]. These coefficients allowed us to normalize the tempo variability of the songs. In actual fact, even though we chose only songs, whose mood was "Andante", the term refers not to an exact speed rate, but to a range of rates.

After isolating the sentences of each song first verse and refrain with the software *Audacity*, we segmented the sonogram in vocalic and intervocalic intervals[16] using *Praat*. Then, we computed the variables with the software *Correlatore* and processed a statistical analysis to identify any possible outlier. Afterwards, we remove all the outliers to obtain a less scattered data. Lastly, to statistically prove our results, we carried out a *t-test analysis*.

5.1 Results and discussion

The results obtained for both the variables show a similar trend; but, only the results concerning the *nPVI* coefficients show a statistically significant difference between Italian and English rhythm (*vocalic nPVI*: $t(58.632) = -2.9077$, $p = 0.005$; *intervocalic nPVI*: $t(59.839) = 2.2641$, $p = 0.027$)[17]. The vocalic intervals are more variable for English sung speech, whereas the intervocalic intervals seem not to be subject to such a variation. Indeed, we have observed in various cases that the syllable structure of English sung speech changed, enduring reduction and deletion phenomena of consonants, mostly at the end of the phrases and the sentences. This could be an explanation of the low intervocalic values of English songs (see Table 3 and Table 4). With such preliminary overview, our hypothesis seems to be corroborated by our *nPVI* data. However, given that we took into consideration a corpus not adhering to stringent criteria such as the same number of syllables for each sentence, further analysis on syllable structure as well as syntactic and morphological analyses are needed.

Lastly, through in-depth analysis of the data, there is no trace of the contrast between the original and the foreign versions that was evident in the musical analysis[18].

[15] The *varco* coefficient is the percentage of ΔC of the mean value for consonantal intervals and the percentage of ΔV of the mean value for vocalic intervals (Dellwo, Wagner, 2003; Dellwo, 2006).

[16] To segment the vocalic and intervocalic intervals we followed the definition of Grabe, Low (2002: 520): "Vocalic intervals were defined as the stretch of signal between vowel onset and vowel offset, characterised by vowel formants, regardless of the number of vowels included in the section (a vocalic section could contain a monophthong, a diphthong, or, in some cases, two or more vowels spanning the offset of one word and the onset of the next). Intervocalic intervals were defined as the stretch of signal between vowel offset and vowel onset, regardless of the number of consonants included."

[17] The differences between English and Italian *varcoΔV* ($t(59.018) = -1.1$, $p = 0.275$) and *varcoΔC* ($t(53.535) = 1.2854$, $p = 0.207$) are not statistically proven.

[18] Data of the vocalic *nPVI* computed for the four groups: mean of the vocalic *nPVI* of original Italian songs: 57.90; mean of the vocalic *nPVI* of the English versions of the original Italian songs: 74.89; mean of the vocalic *nPVI* of the original English songs: 69.09; mean of the vocalic *nPVI* of the Italian versions of the original English songs: 61.00.

In fact, the hypothesized configuration occurs both for the originals and the transpositions; for example, as it can be seen in Figure 6, the English vocalic *nPVI* is higher than the Italian one in the group of "Original Italian songs" (t(15.784) = -2.8194, p = 0.012), whereas there is no significative difference between English and Italian vocalic *nPVI* of the group of "Original English songs" (t(31.617) = -1.4258, p = 0.163), as it was for the musical data (see § 4.2). Moreover, *t-test* analyses confirm that there is not a significative difference neither between Italian vocalic *nPVI*s (t(24.441) = 0.63665, p = 0.5303) nor between English vocalic *nPVI*s (t(21.597) = 0.86575, p = 0.3961). Consequently, English and Italian sung speech rhythm remain stable regardless the origin of the musical theme.

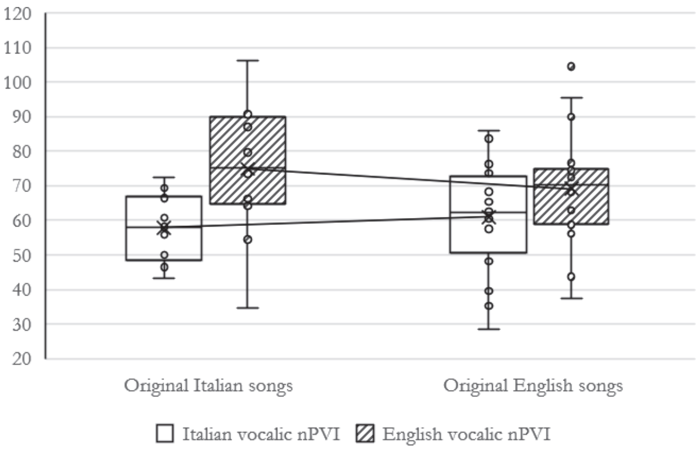
Table 3 - *Mean, median and standard deviation of the Italian and English vocalic and intervocalic nPVI, and their respective minimum and maximum values*

	Italian voc nPVI	**English voc nPVI**	**Italian interv nPVI**	**English interv nPVI**
Valid	29	30	30	28
Missing	2	1	1	3
Mean	59.61	71.41	67.90	55.88
Mediana	60.49	73.19	67.14	55.95
Std. dev.	13.57	17.47	19.98	15.62
Minimum	28.45	34.68	36.64	29.90
Maximum	86.03	106.3	107.6	85.39

Table 4 - *Mean, median and standard deviation of the Italian and English varcoΔV and varcoΔC coefficients, and their respective minimum and maximum values*

	Italian varcoΔV	**English varcoΔV**	**Italian varcoΔC**	**English varcoΔC**
Valid	29	30	30	29
Missing	2	1	1	2
Mean	61.67	69.63	56.26	52.68
Mediana	59.77	66.74	57.45	54.72
Std. Dev.	16.97	22.04	16.43	15.01
Minimum	30.74	28.08	29.04	22.79
Maximum	94.58	117.7	93.39	82.60

Figure 6 - *The box and whisker plot shows the vocalic nPVI results for the first subgroup "Original Italian songs" on the left and the "Original English songs" on the right*



6. *Conclusion*

Although our research might be considered a pilot study on sung speech rhythm, the data collected suggest that the mother tongue of the composer, as well as that of the singer, is mostly coherent with the rhythm of their music and songs. Italian versions of original English songs are in some way different from the original Italian songs inasmuch they retrieve the English rhythm. We know American music has exerted a greater influence on Italian composers of the last century, not allowing them to completely maintain their mother tongue rhythm in the compositions. On the contrary, this influence does not extend to sung speech rhythm, which mostly adhere with speech rhythm. However, larger corpora are needed in order to confirm this hypothesis, and an in-depth analysis of sung speech is required to understand which principles drive rhythm in songs and how sung speech behaves.

A question remains unanswered: is rhythm a prosodic property of speech which can be measured in the signal in simple numerical terms? "The quantitative algorithms do not allow us to relate acoustic data to phonological constraints, although they are widely used because of their explicitness and predictive value" (Marotta 2012). Moreover, Ramus (1999), Grabe, Low (2002), and Dellwo, Wagner (2003) rhythm metrics reflect only the duration and the clustering of segments, but we know that the complexity of the syllable structure and the reduction of unstressed vowels are fundamental to the perception of speech rhythm (see Dauer, 1983; Gil, 1986; Nespor et al., 2011). Besides, recent works pointed out the multidimensional complexity of the perception of rhythm, requiring a multimodal analysis of speech, but also of music, signal (Cummins, 2009).

In conclusion, speech rhythm as well as music rhythm could be epiphenomena driven by more general motor behaviors of human beings, that are biologically constrained.

Acknowledgment

We would like to thank Dr. Alessandro Cecchi, researcher in musicology, for his tireless support and assistance throughout this research.

Bibliography

Abercrombie, D. (1965). *Studies in Phonetics & Linguistics*. London: Oxford University Press, 16-25.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. In *Journal of Phonetics*, 40, 351-373.

Bertinetto, P.M. (1989). Reflections on the Dicotomy "Stress- vs. Syllable-timing". In *Revue de Phónetique Appliquée*, 91/93, 99-130.

Bertinetto, P.M., Bertini, C. (2008). On modelling the rhythm of natural languages. In *Proceedings of the 4th Speech Prosody Conference*, University of Campinas, 427-430.

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. In *Language*, 26, 86-125.

Brown, S., Pfordresher, P.Q., Chow, I. (2017). A musical model of speech rhythm. In *Psychomusicology: Music, Mind, and Brain*, 27, 2, 95-112.

Cummins, F. (2009). Rhythm as an Affordance for the Entrainment of Movement. In *Phonetica*, 66, 15-28.

Cummins, F. (2015). Rhythm and speech. In Redford, M.A. (Ed.), *The Handbook of Speech Production*. Wiley, 158-177.

Cummins, F., Port, R. (1998). Rhythmic constraints of stress timing in English. In *Journal of Phonetics*, 26, 145-171.

Dauer, R.M. (1983). Stress-timing and syllable-timing. In *Journal of Phonetics*, 11, 51-62.

Dellwo, V. (2006). Rhythm and speech rate: a variation coefficient for deltaC. In Karnowski, P., Szigeti, I. (Eds.), *Language and Language-processing*. Frankfurt am Meim: Peter Lang, 231-241.

Dellwo, V., Wagner, P. (2003). Relations between language rhythm and speech rate. In Solé, M.J., Recasens, D. & Romero, J. (Eds.), *Proceedings of 15th International Congress of Phonetic Sciences*, 471-474.

Florig, E. (1995). An acoustic study of French vowels in speech and singing voice. In Elenius, K., Branderud, P. (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences. ICPhS 1995. Stockholm, Sweden. 13-19 August 1995, vol. 1*. Stockholm: The Congress Organisers at KTH and Stockholm University, 210-213.

Gill, D. (1986). A prosodic typology of language. In *Folia Linguistica*, 20, 165-231.

Grabe, E., Low, E.L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In *Papers in Laboratory Phonology*, 7. Cambridge: Cambridge University Press, 515-546.

Han, M.S. (1962). The feature of duration in Japanese. In *Onsei no kenkyuu*, 10, 65-80.

Hannon, E.E. (2009). Perceiving speech rhythm in music: Listener classify instrumental songs according to language of origin. In *Cognition*, 111, 403-409.

Harding, E.E., Sammler, D., Henry, M.J., Large, E.W. & Kotz, S.A. (2019). Cortical tracking of rhythm in music and speech. In *NeuroImage*, 185, 96-101.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C. & Liu, H.H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proc. R. Soc. London*, Ser. A 454, 903-995.

Huron, D., Ollen, J. (2003). Agogic contrast in French and English themes: further support for Patel and Daniele (2003). In *Music Perception: An Interdisciplinary Journal*, 21, 2, 267-271.

Jekiel, M. (2014). Comparing rhythm in speech and music: the case of English and Polish. In *Yearbook of the Poznań Linguistic Meeting* 1. De Gruyter Open, 55-71.

Johnson, R., Huron, D. & Collister, L. (2014). Music and lyrics interaction and their influence on recognition of sung words: an investigation of word frequency, rhyme, metric stress, vocal timbre, melisma and repetition priming. In *Empirical Musicology Review*, 9, 1, 2-20.

Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Larroque, P. (2012). The Rhythm of English and Blues Music. In *American International Journal of Contemporary Research*, 2, 5, 123-136.

Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lehiste, I. (1977). Isochrony reconsidered. In *Journal of Phonetics*, 5, 253-263.

London, J. (2017). Rhythm. In *New Grove dictionary of music and musicians*. Oxford: Oxford University Press, 21, 277-309.

London, J., Jones, K. (2011). Rhythmic refinements to the nPVI measure: a reanalysis of Patel & Daniele. In *Music Perception: an Interdisciplinary Journal*, 29, 1, 115-120.

Marotta, G. (2012). Piedi metrici e sillabe orfane nella prosodia dell'italiano. In Schafroth, E., Selig, M. (Eds.), *Testo e ritmi. Zum Rhythmus in der italienischen Sprache. Studia Romanica et linguistica, 35*. Frankfurt a.M.: Lang, 89-103.

Nespor, M., Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications.

Nespor, M., Vogel, I. (2007). *Prosodic Phonology*: *with a new foreword*. Berlin-New York: Mouton De Gruyter.

Nespor, M., Shukla, M. & Mehler, J. (2011). Stress-timed *vs.* Syllable-timed languages. In Van Oostendorp, M., Ewen, C.J., Hume, E. & Rice, K. (Eds.), *The Blackwell companion to phonology II*. Oxford: Wiley-Blackwell, 1147-1159.

Oussilova, E. (1995). Effect of vowel modification on the phonemic accuracy of vowels and palatalization of the consonants in Russian vocalized speech. In Elenius, K., Branderud, P. (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences. ICPhS 1995. Stockholm, Sweden. 13-19 August 1995, vol. 1*. Stockholm: The Congress Organisers at KTH and Stockholm University, 210-213.

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

Patel, A.D., Daniele, J.R. (2003b). Language, music, syntax and the brain. In *Nature Neuroscience*, 6, 674-681.

Pike, K.J. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Publications.

Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In *1st International Conference on Speech Prosody*, France, 115-120.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. In *Cognition*, 73, 3, 287-288.

Sadakata, M., Desian, P., Honing, H., Patel, A.D. & Iversen, J.R. (2004). A cross-cultural study of the rhythm in English and Japanese popular music. In *Proceedings of the International Symposium on Musical Acoustics (ISMA)*, Nara, 41-44.

Selkirk, E. (1984). *Phonology and syntax: the relation between sound and structure.* Cambridge, MA: The MIT Press.

Steele, J. (1775). *An essay toward Establishing the Melody and Measure of Speech*. London: Printed by Bowyer, W., Nichols, J., for Almon, J.

Sundberg, J., Lindblom, B., (1976). Generative Theories in Language and Music Description. In *Cognition*, 4, 99-122.

Tilsen, S., Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. In *The Journal of the Acoustical Society of America*, 134, 1, 628-639.

Appendix

Balla Linda – Lucio Battisti

Bella Linda – The Grass Roots

Mi ritorni in mente – Lucio Battisti

Wake me I am dreaming – The Love Affair