Musical Speech: a New Methodology for Transcribing Speech Prosody

Alexsandro R. Meireles[1], Antônio R. M. Simões[2], Antonio Celso Ribeiro[3], and Beatriz Raposo de Medeiros[4]

[1,3] Federal University of Espirito Santo, Brazil, [2]University of Kansas, USA, [4]University of São Paulo, Brazil

meirelesalex@gmail.com, asimoes@ku.edu, antoniocelsoribeiro@gmail.com, biarm@usp.br

Abstract

Musical Speech is a new methodology for transcribing speech prosody using musical notation. The methodology presented in this paper is an updated version of our work [12]. Our work is situated in a historical context with a brief survey of the literature of speech melodies, in which we highlight the pioneering works of John Steele, Leoš Janáček, Engelbert Humperdinck, and Arnold Schöenberg, followed by a linguistic view of musical notation in the analysis of speech. Finally, we present the current state-of-the-art of our innovative methodology that uses a quarter-tone scale for transcribing speech, and shows some initial results of the application of this methodology to prosodic transcription.

Index Terms: speech prosody, speech melodies, musical notation, quarter tones

1. Introduction

It is known among linguists that speech is composed of musical elements such as speech rhythm, intonation, tonicity, and speech dynamics. Speech Prosody is the area of Linguistics that investigates this musicality. In recent years there has been an ongoing interest in this field, and since 2002 a special conference Speech Prosody (http://isle.illinois.edu/sprosig/sp2002/), has been organized to investigate all interdisciplinary aspects of speech prosody. Although musical aspects are present in speech, researchers have rarely tried to represent speech using traditional musical notation. On the other hand, composers such as Bach, Beethoven, Schoenberg, Reich, and Janaček have represented speech using traditional scores. Indeed, Bach was one of the pioneers in musically scoring speech with the *recitativo secco*, a technique in which a singer is allowed to adopt the rhythms of ordinary speech [1, 17].

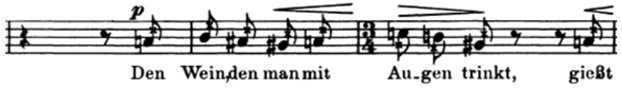

Figure 1: Excerpt of Schoenberg's speech-melody (represented by the crosses "x" on the notes) as seen in Pierrot Lunaire op. 21. The cross represents the speech notes. Freely available at http://imslp.org/wiki/Pierrot_Lunaire,_Op.21_(Schoenberg,_Arnold).

2. Systematic transcription of speech with musical notation

2.1. Janáček's Theory of Speech Melodies

According to Fiehler [6], Leoš Janáček (1854-1928) developed his Theory of Speech Melodies at the time he was writing his opera Jenůfa (1903). In this same text, the author states that the composer intended to faithfully represent speech. For this reason, as happens with speech, "his melodic fragment are unstructured, phrased unconventionally, unconstrained by key or meter" [6, p. 42].

Tyrrel [7, p.793] points out that Janáček related in 1916 that "tune is created by the word, the whole melody depends thus upon the sentence, it couldn't be otherwise". Nevertheless, even though the composer always denied he fits words to an existing tune, Tyrrel [7] shows in his paper that he frequently did it when he rewrote his musical pieces.

2.2. Joshua Steele's methods on melody and measure of speech (1775)

The idea of transcribing speech using music notation comes even before Janáček's time. In 1775 Joshua Steele wrote the book "An essay towards establishing the melody and measure of speech to be expressed and perpetuated by peculiar symbols" [9]. Even though he did not have the modern instrumental tools to help him with transcription, he developed a very interesting way of notating speech with music notation. Contrary to Janáček, though, Steele's was more interested in the linguistic side of speech and his main objective was to show that "the musical expression of speech may be described and communicated in writing" [9, p. 15].

3. A linguistic view on the musical aspects of speech

This paper intends to present a new methodology for transcribing speech prosody, which is inserted in a research program that proposes to develop a new model for studying speech prosody with music notation. As far as we know, no researcher has tried to fully develop a model for studying speech prosody using musical elements. Nevertheless, as seen before, some studies have transcribed speech using this method.

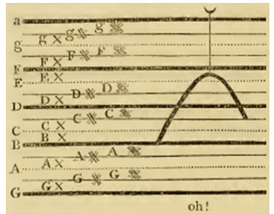

Figure 4: Steele's representation of speech with a quarter tone scale. Steele's transcription is very detailed and looks like a mix of f0 curves, music notation, and other representations of prosodic notation. Segmental symbols like notes and rests, as well as symbols representing musical dynamics were elements of Steele's detailed transcription, but to describe his notation system is beyond the scope of this paper.

4. Methods

In Simões and Meireles [12] we started to develop a method for transcribing speech prosody with music notation. The transcriptions of Brazilian Portuguese, Mexican Spanish and American English were automatically done using the software Ableton Live 9 Suite and then annotated in Finale 2011. After that, we compared the musical scores of some scores and realized that the musical transcriptions via wave to midi conversion were not very much precise. That is why we devised a more sophisticated method of transcription based on the acoustic signal.

Due to software limitations, we could only transcribe speech using a traditional 12-tone scale. Nevertheless, by measuring the frequencies in speech, we noticed that very often the syllable's frequencies did not correspond to the frequencies in this scale. That is why we thought in expanding this scale to include smaller intervals than the semitone. At first we thought using a 48-tone scale, but as divisions smaller than the quarter tone are very difficult for listeners (even trained musicians) to perceive, we chose to use a quarter-tone scale, which is used for example in Arabic music and is also known by electric guitarists. Therefore, our scale consisted of 24 tones equally spaced (see figure 6).

In order to divide the scale in 24 tones we used the formula, =f0*√(24&2^i), where f = frequency to be calculated, f0 is the first note of the scale, and i is one of the notes of the scale. As an example, let us observe below the frequency intervals from A to A# (=Bb).

Table 1: Frequency chart from A to Bb where f0 = 440 hertz. q is a quarter tone.

Note	Scale
A	440 (i=0)
Aq	452.89 i=1)
Aw	466.16 (i=2)

We created then a frequency chart using a quarter tone scale considering all the possible quarter tone frequencies in a piano keyboard from C0 to C8. After that, we divided the sentences in vowel-to-vowel (VV) units [cf. 14] in Praat [15] using the BeatExtractor plug-in [16], took the frequency mean for each VV unit using the script f0.praat3, and, by comparing this VV frequency with the frequencies in our 24 tone chart, we got the musical notes with its respective octave in a piano keyboard. Of course, it was hard to find a 100% correspondence of the speech notes with the chart. That is why we made a formula in Visual Basic within Microsoft Excel to automatically make this correspondence and find out what was the frequency in the chart that was closer to the mean VV frequency. For example, the VV unit "ood" in "flooded" (see texts in [12]), recorded by subject AEF (American English female), had a mean frequency of 210 Hz, which is in-between Ab3 (207.64 Hz) and Abq4 (213.73 Hz). By subtracting the VV frequency by the frequency to be compared (in Hz) we got that 210 Hz is closer in modulus to Ab3 (210 – 207.64 = 2.36) than Abq (210 – 213.73 = -3.73). Thus, for "ood", we found, as a result, the note Ab. We have followed this procedure for all the VV units in the corpus used by Simões and Meireles (2006) to transcribe speech with musical notation.

5. Discussion

In order to exemplify the transcriptions done in this part of our research, take a look at figure 6. As can be seen there, we have randomly used a 4/4 time measure, a piano score (2 clefs), and a C key for our transcriptions. Of course, this is a first step in using a quarter tone scale for transcribing speech and thus further improvement may be necessary.

Regarding the time signature, we have to study in the future the pattern of the phrases' stress groups, so that we may divide the scores in signatures that fit the phrases strictly within the bars (see for example figure 6).

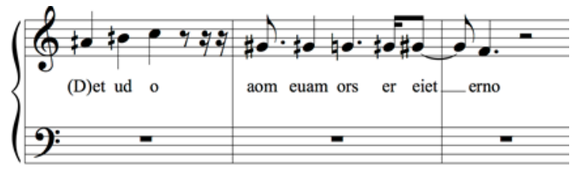

Figura 6. Example of a transcription in quarter tones, representing a passage of Soneto de Fidelidade by Vinícius de Moraes, read by a Brazilian female subject.

6. Conclusions

We presented here an innovative method of transcribing speech prosody as musical notation using a quarter-tone scale. As shown here, this method is much more reliable and precise than the technique we have used before. There are, though, some issues that need to be taken into consideration for further improvement of the method.

7. Acknowledgements

The authors would like to thank the São Paulo Research Foundation (FAPESP grant 2015/06283-0 to the fourth author) for supporting this research, and Daniel Hirst, who introduced us to the work of John Steele.

Bibliography

Bauer-Lechner, N. (1980). *Recollections of Gustav Mahler*. English translation, Dika Newlin, 1923, p.50. Faber Music. London: Faber & Faber, 1980.

Deutsch, D., Henthorn, T., and Lapidis, R. (2011). Illusory transformation from speech to song. *Journal of the Acoustical Society of America*, 2011, 129, 2245-2252.

Fiehler, J. (2013). The Search for Truth: Speech Melody in the Emergence of Janáček's Mature Style. In: Harmonia: Leoš Janáček: Life, Work, and Contribution. *The Journal of the Graduate Association of Musicologists and Theorists at the University of North Texas*, Special Issue, May 2013.

Harnoncourt, Nikolaus von (1973). Das quasi Wort-Ton-Verhältnis in der instrumentalen Barockmusik. In: *Colloquium Music and Word. Musik und Wort*, arranged and edited by Rudolf Pečman. Colloquia on the history and theory of music at the International Musical Festival in Brno 1969. Volume 4, 225–228.

Rapoport, E. (2004). Schoenberg – Hartleben's Pierrot Lunaire: Speech – Poem – Melody – Vocal Performance. *J. New Music Research* 33: 71-111.

Rapoport, E. (2016) "On the Origins of Schoenberg's Sprechgesang in Pierrot Lunaire." *Min-Ad: Israel Studies in Musicology* Online 2 (2016) at http://www.biu.ac.il/HU/mu/min-ad/06/Sprchgsng.pdf (accessed February 7, 2017).

---------- (2016).

Sherman, B. Inside Early Music: Conversations with Performers. New York : Oxford University Press, 1997.

Simões, Antônio R.M. and Alexsandro R. Meireles. 2016. *S*peech Prosody in Musical Notation: Spanish, Portuguese and English. In *Proceedings of the Speech Prosody Conference 8*, editors Jon Barnes, Alejna Brugos, Stefanie Shattuck-Hufnagel and Nanette Veilleux Boston, Mass, Boston University: 212-216.

Steele, J. (1775). *An essay towards establishing the melody and measure of speech, to be expressed and perpetuated by peculiar symbols*. London: Boyer and Nichols. 2nd. Edition 1779, Prosodia Rationalis. London: Nichols.

Tierney, A., Dick, F., Deutsch, D., and Sereno, M. (2012). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Córtex*.

Tyrrell, J. (1970). Janáček and the Speech-Melody Myth, in The Musical Times, Musical Times Publications Ltd., Vol. 111, No. 1530 (Aug., 1970), pp. 793-796. http://www.jstor.org/stable/955301.

Vainiomäki, T. (2012). *The Musical Realism of Leoš Janáček – From Speech Melodies to a Theory of Composition*, Ph.D. dissertation, Helsinki, Finland, University of Helsinki, Faculty of Arts, 356 pages. https://helda.helsinki.fi/bitstream/handle/10138/36087/themusic.pdf?sequence. Last Visited in October 2016.