FROM TRANSCRIPTION TO SIGNAL REPRESENTATION: PITCH, RHYTHM AND PERFORMANCE

Marie Tahon
IRISA
6 rue Kérampont
Lannion, France
marietahon@noath.net

Pierre-Eugène Sitchet
Paris-Sorbonne University, IReMus
2 rue de Louvois
Paris, France
zin@ginositson.com

ABSTRACT

Musical transcription is a real challenge, and more so in the case of folk music. Signal visualization tools may be of interest for this kind of music. The present paper is a comparison between a musical transcription and two signal representations (pitch and rhythm) applied to a song taken from the Gwoka repertoire. The study aims at finding similarities and differences in pitch, rhythm and performance features between the transcription and the signal visualization. Signal visualization is based on vowel segmentation, and on extraction of pitch and duration information. Transcription provides general characteristics about the music (harmony, tonality and rhythmic structure), while signal visualization provides performance-related characteristics. The main conclusion is that both approaches are of great interest for understanding folk music.

Keywords : Musical transcription, Signal visualization, Pitch, Rhythm, Performance, Gwoka.

1. INTRODUCTION

Transcription is described in ethnomusicology as a complex concept which overlaps musical analysis and musical culture [1]. Although transcription is a very difficult task, it may be a good support for analyzing musical features. Therefore, there is a growing interest in ethnomusicology for computational methods and their application to audio data collections [2]. Folk music recordings are very challenging because (i) instrument models do not necessarily exist ; (ii) audio quality is not always satisfactory ; (iii) sung languages have not necessarily been studied from a phonetic and cultural point of view. Automatic music transcription has been developed extensively during the past five years. This development is related to the huge amount of audio data recorded by everyone around the world and shared on databases. Thanks to multi-pitch detection and development of instrument physical models, automatic transcription is becoming more and more efficient. Added specific preliminary knowledge is of great importance for improving the efficiency of the system. As for example, knowledge related to morphological features of the sound of an instrument, sympathetic resonances, inharmonicity, harmonic transitions, tempo or rhythm. Automatic music transcription provides onsets, durations and pitch information. However such information is not sufficient for researchers who are looking for sound quality and music performance.

At present, many audio interfaces have been developed for analysis purposes. Among others, we can cite EAnalysis [3] and SonicVisualizer [4]. These tools are very powerful for giving a real-time representation of the music. They have been developed mainly for instrumental music. Some researchers in ethnomusicology try to use spectrograms for annotating timbral features [5], while others have used automatic transcription [6] and pitch analysis [7]. All these tools are a great help for analyzing music from scores or from audio recordings, but they are performance-specific. For analyzing music from recordings it is necessary to be able to separate performance features (musician, room, context, audio quality) from musical features (structure, harmony, tonality). In improvised folk music such a separation is very uneasy as these features are usually a combination of linguistic, musical and cognitive ones. In order to generalize the resulting analysis, many performances of the same song must be automatically analyzed.

The present work aims at validating an automatic analysis process by comparing manual transcription to signal representation obtained for a single performance. The present study has been carried out within the framework of a research project on Gwoka that aims at finding musical characteristic elements as well as, performance criteria of one specific singer.

The comparison between the manual transcription made by one of the authors and the signal representation obtained by the other underlines similarities and differences between elements such as pitch, rhythm or performance features. Energy level has not been investigated in the present study.

Section 2 summarizes the musical context of Gwoka. Transcription and signal representations are described in detail in section 3. Section 4 is devoted to the comparison between the two approaches. Conclusions and perspectives are drawn in the last section.

2. MUSICAL CONTEXT

Gwoka is a musical genre that emerged in the 17th century during the transatlantic slave trade. Today it has be-

Copyright: ©2016 Marie Tahon et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

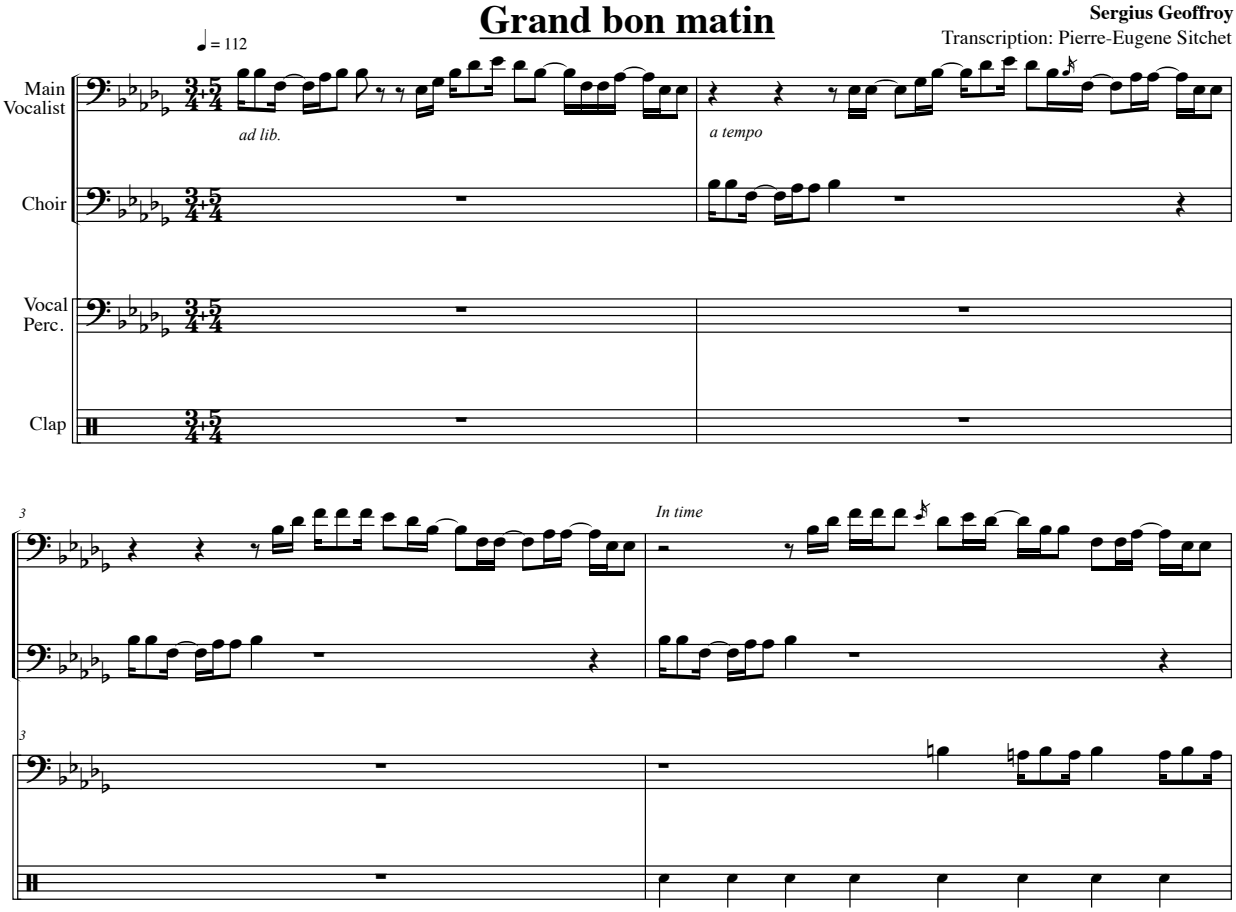

Figure 1. Musical transcription from Pierre-Eugène Sitchet of the song *Dimanche Gran Bon Maten* interpreted by Sergius Geoffroy. The lyrics are : *Dimanche gran bon maten / Gadé koudpyé nonm la té lansé an ka Ibana / Mwen di frè koudpyé nonm la té voyé an kaz a Ibana / Ka mandé lé répondè frapé lanmen an kaz a Ibana*

come an important element of the cultural heritage of Guadeloupe. Combining drum, song and dance, Gwoka is for those who practice it an artistic mode of expression, as well as an assertion of identity, a state of mind and a way of life. Through its history, Gwoka music has played —and still plays— an outlet role, offering an opportunity for freedom of expression. It has also become a sort of catharsis for Guadeloupeans [8].

The "Code Noir" forbade the use of any kind of drums in the practice of music. The slaves used a vocal technique, called Bouladjel, which imitates drums. While Gwoka was frowned for for years, it is now one of the most famous music and dance genres in Guadeloupe. Gwoka has been added to the UNESCO "Representative List of the Intangible Cultural Heritage of Humanity" on November 26, 2014. This musical genre is characterized by improvisation, syncopated rhythm and question-answer structure between a choir and the soloist.

Gwoka is also a kind of verbal joust between singers. Since it is a challenging performance [9] the singer has to hold the audience's attention for a long time. The singer's performance is generally assessed according to two criteria : "Santiman" (evaluation of the emotional expressivity of the singer) and "Lokans" (evaluation of the singer's power and his/her improvisation skills).

Gwoka is sung only in Creole [10]. Antillean Creole language came emerged during the slavery era. African slaves were forced to develop a new form of verbal communication by relying on what they heard from their French masters and from other African slaves. As a result, Creole is a combination of European words and of African expressions and sentence structure.

3. MATERIALS AND METHOD

In the present work the authors have made use of a song of Gwoka performed by Sergius Geoffroy and compared its transcription to the signal annotations. The entire song lasts 150 seconds. It alternates solo-improvised variations and a choir (répondé). The study focuses on the first four variations of the first 25 seconds.

3.1 Musical transcription

The musical transcription was based on a recording of the song. The first instants are reported in Figure 1. Four elements are considered : the soloist, the choir, vocal drums and claps. The choir is always singing the same melodic pattern repeating the title of the song : *Dimanche Gran Bon Maten*. The introduction is performed by the soloist ad libitum and the definitive tempo is reached at the third variation when claps come in. Bars have been chosen from a paradigmatic point of view. Each bar is a melodic, rhythmic and linguistic variation of the first one. The introduction and the first four variations contain 60 notes. The lyrics have also been transcribed, those transcriptions corresponding to the extract that was studied are shown in Figure 1.

Some "statistical" features —in opposition to "syntactical" features such as pitch and durations— are included in the score, for example : *ad libitum, a tempo*. Such annotations will never be able to provide information about voice quality or vocal techniques such as *vibrato, grupetto* because there are an infinite number of ways of doing them [11].

3.2 Signal visualization

3.2.1 Vowel segmentation

Because the study focuses on vocal music, speech and singing voice analysis features have been adapted to its specificities. The present contribution focuses on vowels only because these are homogeneous in time and spectral domains. Other speech sounds, such as voiced consonants, have been discarded because their spectral characteristics differ from those of vowels. Such tools are based on the extraction on vowels of different features. These features are related to pitch, formants, durations and voice quality. The extraction of such features from other speech signals (consonants) or instrumental signals (claps) would not have any physical meaning.

Creole speech is a highly voiced language in comparison with French, and an automatic segmentation based on voiced parts was not possible. Moreover, since this is a language that has not been widely studied, automatic syllable segmentation would probably result in many errors. For these reasons, the authors decided to segment and annotate manually vowels in sung and spoken tracks. Since the segmentation of vowels is mainly used to study their duration, they are annotated according to basic French vowels /a,e,i,o,u/. Nasals and diphthongs are annotated according to the closer basic vowel. Segmentation and annotation have been done with Transcriber [12]. A complete phonetic transcription would be interesting for studying pronunciation variants, but it has a higher annotation cost.

Pitch is extracted with Praat [13] tools every 10 ms on voiced signals only. Praat tools can make octave errors. To avoid such errors, only vowel signals that last more than 40 ms are considered. Duration corresponds to the length of the vowel signal that has been segmented. In the first four variations, 65 vowels were segmented. The number of notes and the number of vowels are not identical : some of the notes contain two vowels, some of the vowels are too short and therefore are not taken into account during the transcription process.

3.2.2 Pitch representation

Signal features are plotted on a graph in Figure 2. Continuous black lines correspond to the F key stave. Dotted black lines correspond to each semitone, the reference frequency being A4 at 440 Hz. The authors have added red notes that correspond to the pitches and durations obtained during the transcription process. Durations are computed on the basis of a 112 bpm tempo. The beginning of each note is synchronized with the beginning of each vowel.

3.2.3 Rhythm representation

Rhythm transcription has been done in order to highlight the global structure of the song. A bar is shown at the end of each melodic variation. In this section, the authors compare onsets and durations of each note or vowel. The transcribed score at relative level gives a note's duration ; the manually recognized tempo (112 bit per minute) gives their absolute values in seconds. These durations are discrete values. The manual segmentation process gives a vowel's duration. These durations are continuous values. It is important to consider that a note usually corresponds to a whole syllable (consonant and vowel) ; therefore the duration of a vowel will probably be shorter than that of a note.

Rhythm is visualized on a 2D graph (see Figure 3). Abscissa is time in seconds while ordinate is the duration value in seconds. Blue lines represent the duration of vowels and red lines the duration of notes. Onsets are used to plot each line at the right time. Black vertical lines join the beginning of a vowel to its assumed corresponding note. If a vowel has no corresponding notes it is linked to zero value.

At first the plot (not included here) showed that onsets and durations of extracted vowels and transcribed notes were not comparable. Onsets and durations of notes were thus adjusted to vowel onsets and durations (Table 1). The following two adjustments were made :

— The first note onset of each variation is synchronized with the first vowel onset.

— Tempo is adapted for each melodic variation using equation 1 where the transcribed variation duration is the number of beats of the variation and the vowel variation duration is the duration in seconds of the variation.

$$Tempo = 60 \times \frac{TranscriptedVarDuration}{VowelVarDuration} \qquad (1)$$

4. COMPARATIVE ANALYSIS

In the present section, the authors discuss the comparison between transcription and signal representation.

4.1 Pitch analysis

4.1.1 Note decision

During the transcription process, the musical ear needs to decide the pitch of the note. This decision is not easy

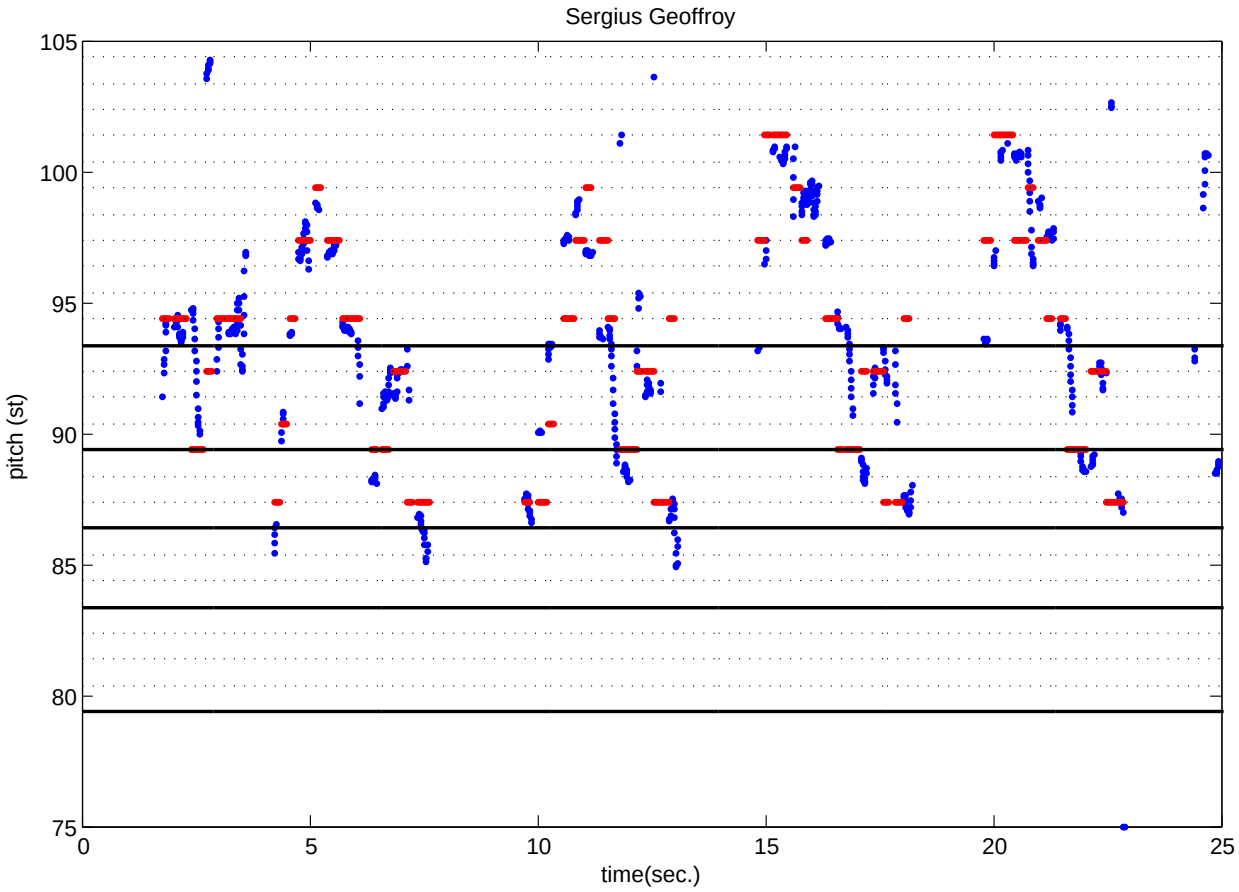

Figure 2. Pitch representation : extracted pitch values in blue and transcripted notes in red, according time.

Variation	Onset (sec.)	Tempo (bpm)
Intro	1.79	80.21
Var1	4.22	77.81
Var2	9.71	95.65
Var3	14.85	95.93
Var4	19.81	103.13

Table 1. Adjustments used for rhythmic comparison.

and is usually the result of a cognitive process that requires an interpretation of the song. There are some meaningful differences between transcribed notes and extracted pitch values. The transcriber usually chooses either :

— The last pitch he heard. It generally occurs after a glissando.
— The first pitch he heard. It generally occurs at the end of the variation.
— A kind of average of the pitches he heard. It generally occurs during a vibrato or even a glissando.

4.1.2 Tonal and harmonic a priori knowledge

Gwoka is a modal music with strong tonal characteristics [14]. The transcriber has in mind an a priori interpretation. For example, at the very begining of the second variation, the vocalist performs several glissandi from Db2 to D3, while the transcription shows successive notes from Eb2 to Eb3 from the dominant chord.

One can see that intonation is increasing with time. This evolution is discarded in the transcription, but is remarkable on the pitch signal representation. This information may be useful for understanding the performance, but is harmful for generalization purposes.

4.2 Rhythm analysis

4.2.1 Tempo stability

The tempo is not stable. In order to synchronize transcription with signal rhythm representation, one needs to adjust the tempo at each variation. During the performance different bars do not last the same time. Here again, this information is useful for understanding the performance and harmful for generalization purposes.

4.2.2 Vowel accentuation

Most vowels are shorter (in the signal representation) than the corresponding transcribed notes. However, some vowels are longer (for example the 6th vowel in Figure 3). This information helps the musicologist to locate accentuations in a given musical phrase. These usually occur at the same time as a vibrato, thus showing the importance of this particular vowel. Of course, these accentuations must be compared to the linguistic content.

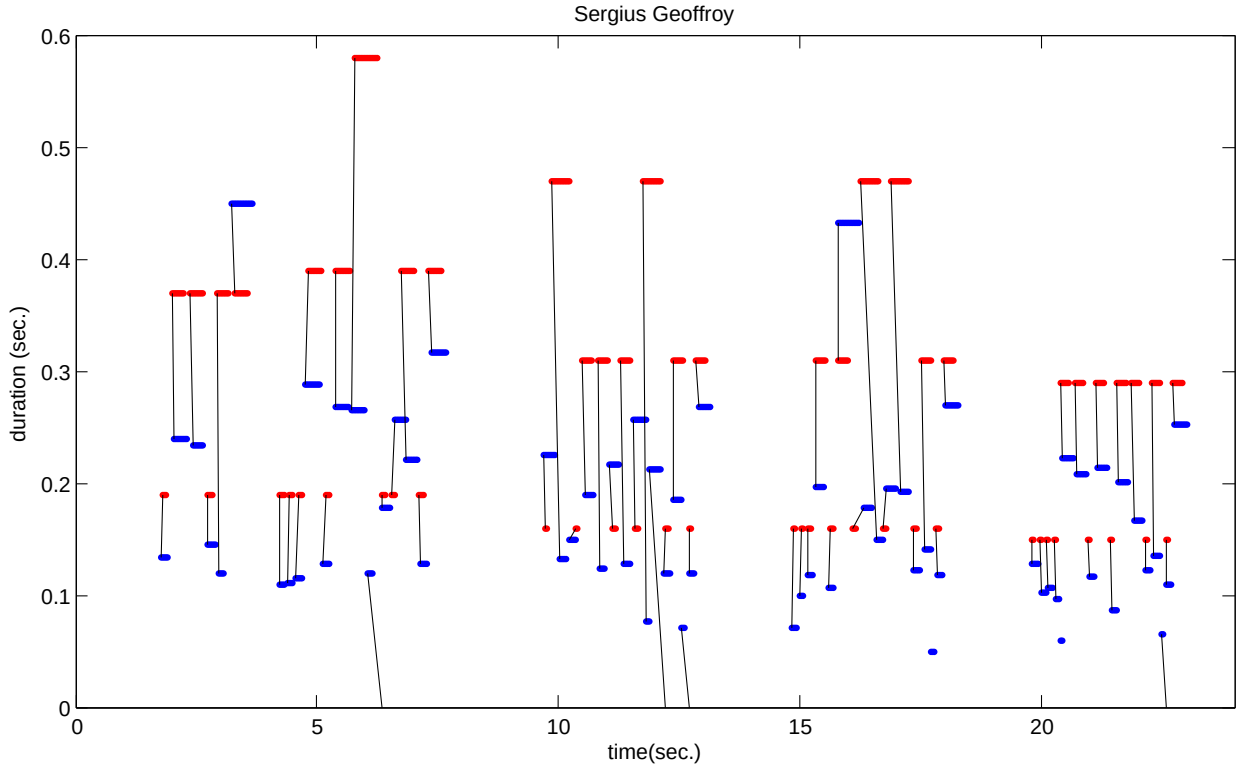

Figure 3. Rhythm visualization. Segmented vowel durations in blue and transcripted note durations in red according time.

4.3 Voice quality analysis

4.3.1 Vocal techniques

Annotation of vocal techniques is possible on the transcription score. An example is given in Figure 1, at the second variation. Some written annotations are usually added to the score in order to help musicians to interpret the score. However, the score has no vocation to show vocal quality or vocal techniques used by the singer during his performance [15].

Transcription in ethnomusicology aims at understanding and analyzing performances. Therefore indications on the techniques used are of great importance. Such indications are highly related to the musical context. Information about vibrato, glissando or grupetto can be seen on pitch graphs. Of course, a spectrogram of the signal greatly helps to detect these vocal techniques.

4.3.2 Linguistic content

In Gwoka, the linguistic content and the musical performance are strongly linked. Creole lyrics have also been manually transcribed (see Figure 1). However, the transcription has been made according to the guadeloupean creole alphabet and does not reflect the real pronunciation. An automatic speech recognition system trained for French Creole could enrich the transcription with phonological aspects.

For example, a previous work on Gwoka [16] shows that nasality is a specific-singer trait. Each singer has also his/her own way of saying the same word. Knowledge of the lyrics is very important to understand the song in the general context of Gwoka, while knowledge of phonological aspects helps the musicologist to understand the performance.

5. CONCLUSION AND DISCUSSION

This paper focuses on a single song of the Gwoka repertoire from Guadeloupe. It presents a comparison between a manual transcription and a signal representation through three main aspects : pitch, rhythm and performance.

Preliminary results validate an automatic transcription process thanks to a comparison with a manual transcription. While manual transcription generalizes the analysis of a specific song by introducing an a priori knowledge, automatic transcription is specific to one particular performance. Manual transcription highlights specific tonalities and rhythmic structures through a priori knowledge of harmony, rhythmic structure, etc. On the contrary, signal representation is entirely related to the song that is being analyzed (recording conditions, singer, context of the performance, etc.). The automatic process of statistical analysis can be done for a large number of performances of the same songs, making it possible to generalize the analysis. Thus it turns out to be a very nice tool for oral folk song analysis.

The present comparison between musical transcription

and signal representations allowed the authors to describe some characteristic features of Gwoka and to understand how a singer performed his own interpretation. The singer's performance features result from several vocal techniques (grupetto, glissando, and vibrato), accentuation of some vowels, an increasing tempo and an increasing intonation.

The main conclusion is that both manual transcription and signal processing are complementary for oral music analysis. The fact that the two analyses were conducted independently avoids bias. Such an approach should provide researchers with tools for a better analysis of music.

Acknowledgments

The authors would like to thank all singers from Guadeloupe who have participated to the recordings.

6. REFERENCES

[1] B. Nettl, "I Can't Say a Thing until I've Seen the Score : Transcription" in *The study of ethnomusicology : Thirty-one issues and concepts.* Urbana, IL : University of Illinois Press. pp. 74-91, 2005.

[2] S. Canazza, A. Camurri and I. Fujinaga, "Special section : ethnic music audio documents : From the preservation to the fruition", *Signal Processing*, 90(4), pp. 977-1334, 2010.

[3] `http://logiciels.pierrecouprie.fr/?page_id=402`

[4] `http://www.sonicvisualiser.org/`

[5] A. Blackburn and J. Penny. "Timbral Notation from Spectrograms : Notating the Un-Notatable ?" in *Proc. of the Int. Conf. on New Tools for Music Notation and Representation TENOR 2015*, Paris, France, pp.218225, 2015.

[6] D. Cazau, O. Adam and M. Chemillier. "Automatic music transcription of the marovany zither, based on knowledge from musical acoustics." in *Proc of the 5th International Workshop on Folk Music Analysis - FMA2015*, Paris, France, pp. 10-12, 2015.

[7] S. Abdallah, A. Alencar-Brayner, E. Benetos, S. Cottrell, J. Dykes, N. Gold, A. Kachkaev, M. Mahey, D. Tidhar, A. Tovell, T. Weyde and D. Wolff. "Automatic transcription and pitch analysis of the British Library World & Traditional music collections". in *Proc of the 5th International Workshop on Folk Music Analysis - FMA2015*, Paris, France, pp. 10-12, 2015.

[8] M.H. Laumuno. *Gwoka et Politique en Guadeloupe, 1960-2003 : 40 ans de construction du pays . [Gwoka and politics in Guadeloupe, 1960-2003 : 40 years of local establishment.]* l'Harmattan, Paris, 2011.

[9] M.H. Pichette, M. Desroches. "Comprendre le Gwoka par la performance. [Understanding Gwoka with performance.]" in *Mise en scne des territoires musicaux : tourisme, patrimoine et performance [Musical territories staging : tourism, heritage and performance.]*, Presses de l'Universit de Montral, 2011.

[10] M.C. Lafontaine. "Terminologie musicale en Guadeloupe : ce que le crole nous dit de la musique. [Musical terminology in Guadeloupe : what creole tells about music.]" *Language et Socit*, vol. 32, pp. 7-24, 1985.

[11] L. B. Meyer,"A Universe of Universals", *The Journal of Musicology*, Vol. 16, No. 1, pp. 3-25, 1998.

[12] C. Barras, E. Geoffrois, Z. Wu, M. Liberman. "Transcriber : development and use of a tool assisting speech corpora production." *Speech Communication*, vol. 33 (1), pp. 5-22, 2000.

[13] `http://www.fon.hum.uva.nl/praat/`

[14] J. Camal. "From Gwoka modènn to jazz ka : music, nationalism and creolization in Guadeloupe", PhD thesis, Graduate School of Arts and Sciences of Washington University, 2011.

[15] P.-E. Sitchet and M. Tahon. "Le voisement dans le Gwoka : entre le parlé et le chanté" [Voicing in Gwoka : between speaking and singing], Journées d'Informatique Musicale, Albi, France, pp.153-159, 2016.

[16] M. Tahon and P.E. Sitchet. "Interest of a mixed approach combining paradigmatic method and signal processing. Case study : Analysis of two musical pieces of Gwoka from Guadeloupe" In : *International Colloquium : Innovant tools and methods for music and signal processing teaching*. St Etienne, France, 2015.