Automatic Melodic Transcription of Flamenco Singing

Emilia Gómez
Music Technology Group, Universitat Pompeu Fabra, Spain
emilia.gomez@iua.upf.edu - http://mtg.upf.edu

Jordi Bonada
Music Technology Group, Universitat Pompeu Fabra, Spain
jordi.bonada@iua.upf.edu - http://mtg.upf.edu

In: C. Tsougras, R. Parncutt (Eds.)
Proceedings of the fourth Conference on Interdisciplinary Musicology (CIM08)
Thessaloniki, Greece, 2-6 July 2008, http://web.auth.gr/cim08/

Background in ethnomusicology and music analysis. Melody in flamenco music is characterized by the presence of microtonality, pitch glides (portamenti or smooth transitions), a short pitch range or tessitura (usually limited to a range of a sixth and characterized by the insistence on a note and its contiguous ones), the use of enharmonic scale (having microtonal intervallic differences between enharmonic notes), a Baroque ornamentation (with an expressive function) and the use of a Greek Dorian mode in the most traditional songs (Fernández 2004).

Background in computing, mathematics and statistics. State of the art techniques in melodic analysis of audio allow us to obtain different representation levels of a music recording (Gómez et alt. 2003A, Gómez et alt. 2003B). There are different representation levels for melody. Energy (computed for the whole spectrum or in different frequency bands) and fundamental frequency (related to its perceptual correlate pitch) curves are the main low-level melodic features. They represent the instantaneous evolution of intonation and intensity. In a higher structural level, note duration and pitch provide a symbolic representation, which can be the input to higher-level analyses. Finally, deviations of the analyzed recording with respect to the obtained score are related to expressivity.

Aims. The goal of this work is to provide a tool for the automatic melodic transcription of flamenco pieces in different representation levels, focusing on monophonic singing excerpts. The tool should help finding a suitable representation level for flamenco performances and be useful for the comparative analysis of singers and styles.

Main contribution. The steps for melodic description can be summarized as follows. We first compute a set of low-level features from the audio signal, including energy and fundamental frequency. Next, since the tuning reference is unknown, we estimate it with a method based on the equal-tempered chromatic scale. The audio signal is then segmented into notes using a dynamic programming algorithm whose cost functions take into account pitch error, energy variations and note durations. The obtained notes are labeled in terms of pitch and duration. Finally, vibrato depth and rate descriptors are computed for each note. We have worked with a music collection of 135 monophonic pieces corresponding to deblas, martinetes and tonás. These categories of flamenco styles (called *palos*) are traditionally sung a cappella or in some cases with some percussion.

Results. The tool provides a way to visualize different aspects of the analyzed performances, as the instantaneous evolution of energy and pitch, the sequence of notes and some expressive aspects (modulations and transitions).

Conclusions. The work presented here provides a set of techniques for the automatic melodic description of flamenco singing. Quantitative evaluation is a difficult task, as it is necessary to have a set of manually annotated performances by expert flamencologists. This is an ongoing work, although some informal tests have proven the usefulness of the approach. Actually, the proposed tool is being used in the context of a research project on comparative analysis of a cappella flamenco styles (Escobar et alt. 2007).

Implications. These techniques can support research on comparative analysis of singers and styles, and they provide an automatic and objective way to transcribe flamenco performances.

Introduction

Motivation

Flamenco is a music tradition mostly originally from Andalusia in southern Spain. The origin and evolution of the different flamenco genres (*palos*) and variations have been studied by different disciplines, including ethnomusicology, literature and anthropology (Katz 2006).

The fact that flamenco has traditionally been an oral tradition has made the number of available transcriptions scant and incomplete. Transcriptions have been usually manually provided by experts on the field, so that particular characteristics of flamenco music are implicitly considered in the transcription

process. Manual analyses provide very accurate and expert information, although manual annotation is very time consuming and might be subjective or prone to errors. These problems have been pointed out by Toiviainen & Eerola (2006) and Lesaffre et alt (2004) in different musical contexts. Furthermore, in flamenco music there is an additional problem related to transcription, namely, that of traditional western notation might not be suitable to represent all expressive resources used in flamenco. This fact is even more significant when considering singing voice in comparison with guitar playing.

Characteristics of flamenco singing

Flamenco singing is characterized by a set of features that differentiates it from other musical styles. Several of these traits are related to the melodic structure of the pieces (which might depend of the flamenco style or "*palo*"), some others relate to the expressive resources of a particular singer and some others to her/his timbre characteristics. The unavailability of scores and the oral character of flamenco have made discrimination between these diverse types of expressive assets very difficult and demanding, no matter if we deal with a certain style or an individual singer. This fact is linked to a long-standing discussion in the flamenco community around the definition of *styles* and *variants*.

There are different representation levels for melody. Energy and fundamental frequency curves are the main low-level melodic features. Energy is associated with loudness and fundamental frequency is related to its perceptual correlate, pitch. From now on, we will use the term pitch referring to fundamental frequency.

Some of the most relevant features of flamenco sung melodies (Fernández 2004, Mora 2008) are:

- Instability of pitch and energy envelopes.
- Microtonality, i.e., the use of intervals smaller than the equal-tempered semitones of the Western tradition.
- Pitch glides: Portamenti or smooth transitions between notes. Notes are not very clearly attacked.
- Pitch range: A short pitch range or tessitura, usually limited to a sixth and characterized by the insistence on a note and its contiguous ones. Flamenco female singers are usually contralto, in order to gain intelligibility, and male singers are tenors or baritones.
- Enharmonic scale: Presence of microtonal intervallic differences between enharmonic notes.
- Presence of melismas: High degree of ornamentation.
- Different timbres: Depending on the period we find some trends in timbre characteristics of flamenco singers. In general, there is a distinction into "Gypsy" and "non-Gypsy" voices.
- Breathiness: Presence of air in the voice.
- Formants: Absence of high frequency (singer) formant, which is characteristic of classical singing styles.

These characteristics quite contrast with classical singing styles, where it is very important to get a good tuning and timing, and where the timbre is characterized by its stability, absence of air, and the presence of high-frequency formants (i.e. the singer formant) (Sundberg 1987). Some attempts have been already devoted to manually and graphically represent these expressive resources of flamenco singing (Donnier 1997) by using the Gregorian chant neumes.

Automatic transcription of sung melodies

State of the art techniques in melodic analysis of audio allow us to obtain different representation levels of a music recording (Gómez et alt. 2003A, Gómez et alt. 2003B). There are different representation levels for melody. Energy (computed for the whole spectrum or in different frequency bands) and fundamental frequency (pitch) curves are the main low-level melodic features. They represent the instantaneous evolution of intonation and intensity. In a higher structural

level, note duration and pitch provide a symbolic representation, which can be the input to higher-level analyses. Finally, deviations of the analyzed recording with respect to the obtained score are related to expressivity. Some research has also focused on adapting general transcription systems to singing voice, mainly focusing on classical or mainstream popular music (de Mulder et alt 2003, Ryynänen 2006).

Aims

The goal of this work is to provide a tool for the automatic melodic transcription of flamenco singing in different representation levels.

This tool should help to analyze the listed particularities of flamenco singing regarding intonation and dynamics, and it should provide ways to move trough different representation levels of flamenco melodies and compare different singers and styles.

Method

The steps for melodic description can be summarized as follows:

1. Low-level feature estimation. We compute a set of low-level features from the audio signal: energy in different frequency bands and pitch. The pitch estimation algorithm is based on amplitude correlation in the frequency domain.

2. Tuning frequency estimation. As we analyze singing voice performances, the used frequency reference (with respect to 440 Hz) is unknown. In order to locate the main pitches, we perform an initial estimation of the tuning frequency (i.e. reference frequency used by the singer to tune the piece), assuming an equal-tempered scale system. This tuning frequency is computed by minimizing the estimated instantaneous pitch error weighted average. The weights are computed by combining energy and first and second pitch derivatives.

3. Short note transcription: the audio signal is then segmented into short notes by using a dynamic programming algorithm based on finding the segmentation that maximizes a set of probability functions. Those functions consider pitch error, energy variations and note durations.

4. Iterative note consolidation and tuning frequency refinement: the estimated tuning frequency is now refined according to the obtained notes. In order to do so, we minimize the note pitch error weighted average, where weights depend on note durations. Then, existing consecutive notes with the same pitch and a soft transition between them are consolidated. This process is repeated until there is no consolidation.

5. Vibrato estimation: finally, vibrato depth and rate descriptors are computed from the zero-crossings obtained by subtracting the transcribed note nominal pitch to the pitch function.

Case Study

We have worked with a music collection of 135 monophonic pieces corresponding to deblas, martinetes and tonás. These categories of flamenco styles are traditionally sung a cappella or in some cases with some percussion.

Figure 1 shows an example of the obtained representation for a flamenco audio excerpt sung by a female singer. The top view represents the waveform. The middle view shows the obtained note transcription, including quantized notes (up to a semitone) and the instantaneous pitch envelope, where the presence of vibratos, pitch glides and melismas can be observed. The bottom figure represents the energy envelop, which is related to the sound intensity.

Conclusions and future work

The presented work provides a set of techniques for the automatic melodic description of flamenco singing.

Quantitative evaluation is a difficult task, as it is necessary to have a set of manually annotated performances by expert flamencologists. This is an ongoing work, although some informal tests have proven the usefulness of the approach. It has been found

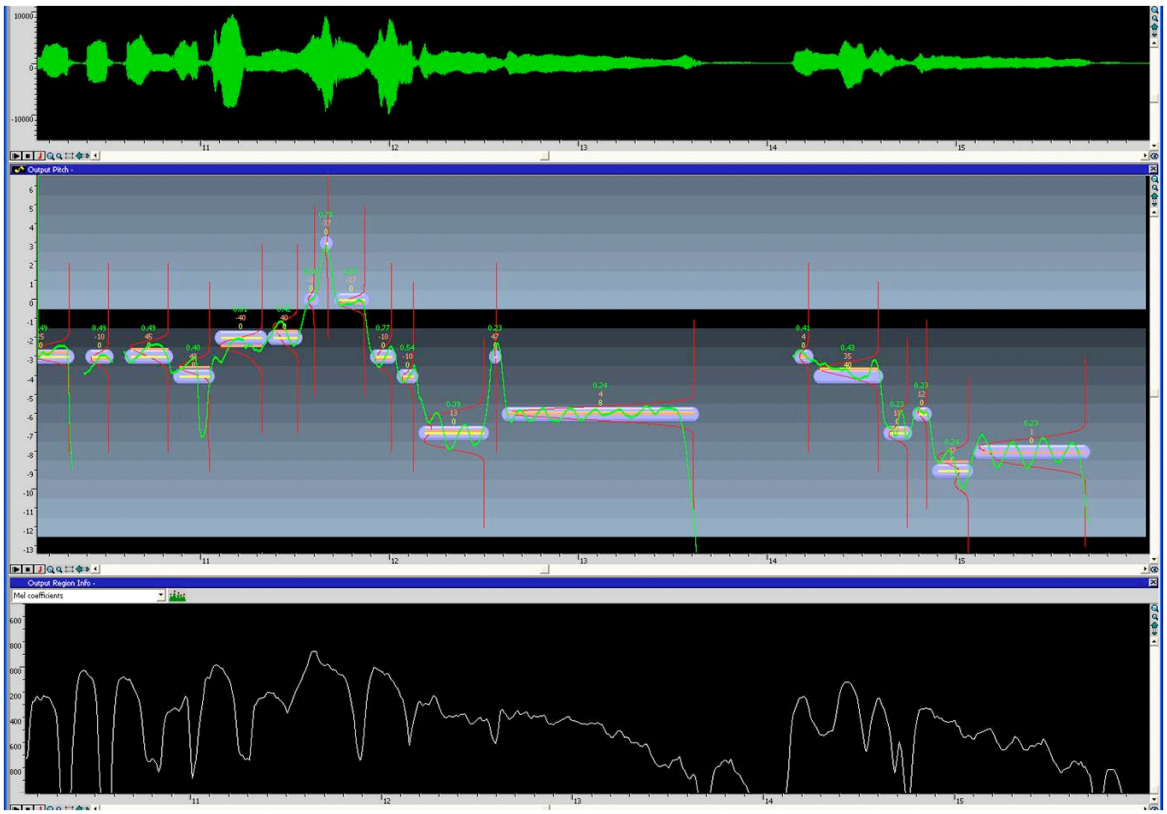

Figure 1. Example of an obtained melodic description

that experts usually use implicit information about flamenco (e.g. which the relevant note within a certain style are) to perform manual transcription.

Actually, the proposed tool is being used in the context of a research project on comparative analysis of a cappella flamenco styles (Escobar et alt. 2008).

Although current results are encouraging, there are still several areas to improve. Just to mention a few, vibratos on top of glissandos are not well detected, and rapid short notes are sometimes considered to be part of longer notes. Besides, in a capella singing, it is not that rare that the reference frequency varies along the song, but the current implementation of the transcription algorithm doesn't allow that frequency variation.

Acknowledgments. The authors would like to thank Joaquín Mora and Jose Miguel Díaz-Báñez, from University of Seville, for providing the music collection and expertise in flamenco, Francisca Merchán for her work on defining expressive features of flamenco singing and Paco Gómez from his review and comments. Acknowledgements. This research has been partially supported by the eContent VARIAZIONI[1] project (ECP 038264).

References

De Mulder, T., Martens, J. P. Lesaffre, M., Leman, M., De Baets, B., De Meyer, H. (2003), *An Auditory Model Based Transcriber of Vocal Queries*, Proceedings of the International Conference on Music Information Retrieval.

Donnier, P (1997*). Flamenco: elementos para la transcripción del cante y la guitarra*, Proceedings of the IIIrd Congress of the Spanish Ethnomusicology Society.

Escobar, F.J., Díaz-Báñez, J.M., Gómez, E., Mora, J., and Cabrera, J. (2008). *Comparative Melodic Analysis of a Capella*

[1] www.variazioniproject.org

Flamenco Cantes, Proceedings of the Conference on Interdisciplinary Musicology, CIM08.

Fernández, L (2004). *Flamenco music theory*. Acordes concert.

Gómez, E. Grachten, M. Amatriain, X. Arcos, J. (2003). *Melodic characterization of monophonic recordings for expressive tempo transformations*, Proceedings of the Stockholm Music Acoustics Conference.

Gómez, E. Klapuri, A. Meudic, B. (2003). *Melody Description and Extraction in the Context of Music Content Processing*, Journal of New Music Research Vol.32 .1

Israel J. Katz. *Flamenco*. Grove Music Online ed. L. Macy (Accessed 16 May, 2006), www.grovemusic.com

Lesaffre, M., Leman, M., De Baets, B. and Martens, J.-P. (2004). *Methodological considerations concerning manual annotation of musical audio in function of algorithm development*, Proceedings of the International Conference on Music Informaiton Retrieval.

Mora, J. (2008). *Características expresivas del cante flamenco*, COFLA meeting, Personal Communication, http://mtg.upf.edu/~egomez/cofla

Ryynänen, M. P. (2006). *Singing transcription*, in *Signal processing methods for music transcription* (A. Klapuri and M. Davy, eds.), Springer.

Sundberg, J. (1987). *The Science of the Singing Voice*. DeKalb, IL: Northern Illinois Univ. Press.

Toiviainen, P. & Eerola, T. (2006). Visualization in comparative music research. In A. Rizzi & M Vichi (Eds.), COMPSTAT 2006 - Computational Statistics. Heidelberg: Physica-Verlag, 209-221.