MIREX 2013 “AUDIO ONSET DETECTION” SUBMISSION: MONOPHONIC-ATIC ALGORITHM

Emilio Molina, Lorenzo J. Tardón, Ana M. Barbancho, Isabel Barbancho
Dept. Ingeniería de Comunicaciones, E.T.S. Ingeniería de Telecomunicación,
Campus Universitario de Teatinos s/n, 29071, Málaga, Spain
emm@ic.uma.es, lorenzo@ic.uma.es, abp@ic.uma.es, ibp@ic.uma.es

ABSTRACT

In this extended abstract, we present an onset detection method for monophonic melodies based on hysteresis on the pitch-time curve. This method is especially designed to perform note segmentation in the case of a-capella singing, in which the pitch evolution during the same note can behave very unstable. The selected approach estimates the regions in which the chroma is stable. Then, a note segmentation stage based on pitch intervals of the sung signal is carried out. To this end, we perform an average of the pitch values of each new note as a representative value of its global pitch, and then we measure the instantaneous pitch deviations with respect to such average. When a sustained / large deviation is detected, a note change is considered, an onset mark is placed and the process restarts.

1. INTRODUCTION

In this extended abstract we describe a novel onset detection method especially designed for monophonic a-capella singing (like it happens in [1] and [2]), but also useful for other type of generic monophonic melodies. The proposed algorithm has been called *monophonic-ATIC algorithm*, and all the details about it will be soon described in [3] (paper under revision). This algorithm has been successfully applied as a singing transcription stage in previous applications [4–6]. Our approach consists of three stages: *low-level feature extraction*, *voicing* and *interval-based note segmentation*. These stages are described in the following sections.

2. LOW-LEVEL FEATURE EXTRACTION

We use the well-known Yin algorithm [7] for low-level feature extraction. The chosen implementation of Yin algorithm was made by its original author in Matlab [8], and it computes three different curves at frame level: fundamental frequency f_0, RMS power pwr and aperiodicity ap. We perform the later interval-based note segmentation making use of just these three curves. In Figure 1, we show the aspect of these curves for singing voice.

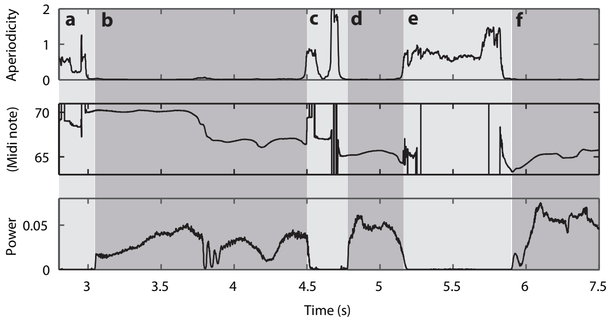

Figure 1. Output of the YIN algorithm for a child singing voice: fundamental frequency, power, and aperiodicity over time. In this figure, real singing notes has been marked with shadowed rectangles.

In order to avoid spurious errors, which could decrease the accuracy of later stages in the system, we average the f_0 curve with a median filter.

3. VOICING

In this section we propose a method to estimate whether a certain frame of the input signal is *voiced* or *unvoiced*. The proposed voicing method is based on the following hypotheses: (1) The pitch slope within a voiced sound does not overcome a certain threshold (apart from octave errors); (2) The energy during a voiced sound is high. It should correspond to stable high energy segments; (3) The aperiodicity during a voiced sound is low. It should correspond to stable low aperiodicity segments. Our method is related to the analysis of *pitch contours*. A pitch contour is a time continuous sequence of F0 candidates grouped using heuristics based on auditory streaming cues. In this extended abstract, we introduce the novel concept of *chroma contour*, which is an octave-independent version of the pitch contour. In our approach, only chroma contours are candidates to voiced regions within the input signal. Thus, our voicing method performs two steps: (1) Estimation of chroma contours (2) Voiced / Unvoiced classification of

This work has been funded by the Ministerio de Economía y Competitividad of the Spanish Governmet under Project No. TIN2010-21089-C03-02 and Project No. IPT-2011-0885-430000, by the Junta de Andalucía under Project No. P11-TIC-7154 and by the Ministerio de Educación, Cultura y Deporte through the 'Programa Nacional de Movilidad de Recursos Humanos del Plan Nacional de I-D+i 2008-2011, prorrogado por Acuerdo de Consejo de Ministros de 7 de octubre de 2011'. This work has been done in the setting Campus de Excelencia Internacional Andalucía Tech, Universidad de Málaga

chroma contours.

4. INTERVAL-BASED NOTE SEGMENTATION

In the following subsections, we expose the details behind our segmentation algorithm. First, in subsection 4.1 we introduce the concept of dynamic averaging, which is a curve that roughly estimates the pitch of each note even when their exact boundaries are unknown. Then, in subsection 4.2 we explain how the hysteresis-based relationship between the instantanous f_0 and the dynamic average is defined in order to detect meaningful pitch deviations.

4.1 Dynamic average

When a note change is detected, the pitch values of all the available frames of the new note are averaged. At the beginning of the note, the dynamic average is similar to the actual f_0 curve because few samples are available. As the note grows, the average becomes closer to a representative pitch value of the whole note. Finally, when a new note is detected the dynamic average is reset and this process starts again.

4.2 Hysteresis

We consider a note change when the area between the instantaneous pitch curve and the dynamic average gets over a certain threshold. This can be seen as a hysteresis process in pitch and time; after the pitch detected has defined the presence of a certain note, a strong and/or sustained deviation is needed to return to the previous note (or to evolve to a new note).

5. MIREX 2013 RESULTS

TO DO.

6. CONCLUSIONS

TO DO.

7. REFERENCES

[1] M. Ryynänen, (2006) “Singing transcription,” in Signal Processing Methods for Music Transcription (A. Klapuri and M. Davy, eds.), pp. 361390, Springer Science + Business Media LLC.

[2] Gómez, E., Bonada J. (2013). “Towards Computer-Assisted Flamenco Transcription: An Experimental Comparison of Automatic Transcription Algorithms As Applied to A Cappella Singing”. *Computer Music Journal*. 37(2), 73-90.

[3] E. Molina, L. Tardón, A.M. Barbancho, I. Barbancho, “Singing transcription method based on a pitch-time hysteresis loop,” *IEEE Transactions on Acoustics, Speech and Audio Processing*, (under revision)

[4] Molina, E., Barbancho I., Gómez E., Barbancho A. M., Tardón L. J. (2013). Fundamental frequency alignment vs note-based melodic similarity for singing voice assessment. *8th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).*

[5] Molina, E. (2012). Automatic scoring of singing voice based on melodic similarity measures. Master Thesis.

[6] Roig, C., Barbancho, I., Molina, E., Tardón, L. J., Barbancho, A. M. (2013). Rumbator: a Flamenco Rumba Cover Version Generator Based on Audio Processing at Note Level, *16th International Conference on Digital Audio Effects Conference (DAFx-13)*

[7] A. De Cheveigné and H. Kawahara, (2002) “YIN, a fundamental frequency estimator for speech and music,” *Journal of the Acoustical Society of America*, vol. 111, no. 4, p. 1917.

[8] A. De Cheveigné, (Last access. Sept. 2013) “Matlab Implementation of YIN algorithm. http://audition.ens.fr/adc/sw/yin.zip,”