MULTIPITCH ESTIMATION AND TRACKING OF INHARMONIC SOUNDS IN COLORED NOISE

Valentin Emiya, Roland Badeau and Bertrand David
École Nationale Suprieure des Télécommunications (ENST),
46, rue Barrault, 75634 Paris cedex 13, France
valentin.emiya@enst.fr

ABSTRACT

We propose an algorithm for the multiple F_0 estimation and tracking task of the Third Music Information Retrieval Evaluation eXchange (MIREX 2007). It has been developped for automatic transcription of piano music, based on a maximum likelihood approach for joint estimation of multiple F_0's. The whole algorithm performs a segmentation of the signal by detecting onsets. In each segment, a set of candidates is then setected, followed by the multiple F_0 estimation. The robustness of the multipitch detection is further enhanced by embedding the maximum likelihood estimation into a Hidden Markov Model framework.

1 SYSTEM DESCRIPTION

The algorithm is composed of five successive stages, which are illustrated in figure 1. Firstly, the signal is segmented with respect to the onset estimates obtained by the algorithm by Alonso [1]. Each of these variable-length segments is then splitted into overlapping frames of constant length.

Secondly, for each segment, we select a presumably oversized set of F_0 candidates by analyzing the first frames and selecting the largest peaks of the function defined by:

$$S\left(f_0\right) = \frac{1}{H(f_0)} \ln \prod_{h=1}^{H(f_0)} \left|X\left(f_h\right)\right|^2 \qquad (1)$$

where $H(f_0)$ is an average number of overtones for fundamental frequency f_0, $X\left(f\right)$ is a whitened version of the Discrete Fourier Transform (DFT) of the current frame and f_h is the frequency of the overtone with order h defined by $f_h = hf_0\sqrt{1+\beta\left(h^2-1\right)}$, β being the so-called inharmonicity coefficient of the piano tone [2].

Thirdly, for each frame t and for each combination $\mathcal{C}$ of notes among the selected candidates, the likelihood $L_{X_t}\left(\mathcal{C}\right)$ of the observed spectrum X_t is derived according to the method described in [3]. It consists in using an autoregressive envelope model for spectral coefficients associated to notes and a moving-average model for the remaining coefficients, related to noise. Once the model

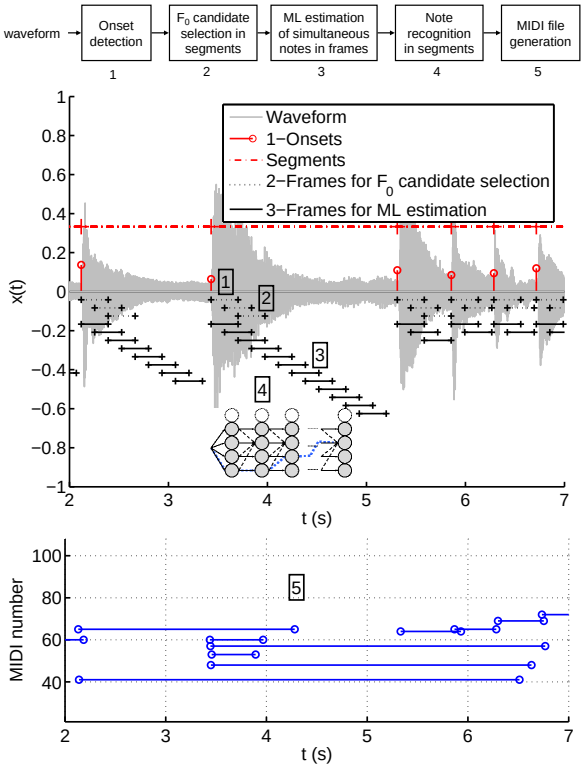

Figure 1. System overview

parameters have been estimated, either for the note or for the noise models, they are used to inverse-filter the observations. The likelihood is then derived as the spectral flatness of the resulting data.

Fourthly, the likelihood estimation scheme is mapped into a hidden-variable Markovian modelization: we create one Hidden Markov Model [4] per segment, the hidden variables being the chords (*i.e.* the combinations of simultaneous notes), the observations being the DFT of the frames, the likelihood of the observations X_t in a state $\mathcal{C}$ being $L_{X_t}\left(\mathcal{C}\right)$. We use a uniform law for initial probabilities of chords composed of F_0 candidates. Only transitions to a subset of the current chord are allowed (*i.e.* notes are allowed to continue or die but not to begin in non-initial states). The viterbi algorithm is applied to extract the most likely state sequence, *i.e.* to track to the most likely chord.

Fifthly, detected pitches (rounded to the closest semitone) in consecutive frames and segments are merged to create notes. Two output files are written, the first one reporting the active pitches every 10 ms, the second one containing onset, offset and pitches of notes.

Parameters of the system are set as follows:

- Sample frequency: 22050 Hz
- Frame length: 93 ms
- Overlap between frames: 50%
- Number of F_0 candidates: 9
- Maximum polyphony: 5

2 RESULTS

Results are reported in two ways: in a frame-by-frame basis (task one) and in a note basis (task two). As our system has been designed for piano music, we will focus on results of tests on piano pieces, which have been used in task two only.

According to the F-measure criterium, which gives a synthetic overview of true positives, false positives and false negatives, our system reaches the eighth rank (out of eleven) for Piano Results based on Onset Only and the seventh rank (out of eleven) for Piano Results based on Onset-Offset. The system suffers from a low recall rate, which has been identified by the authors as coming from low-pitched notes. The fifth rank is reached with respect to the averaged overlap ratio for Piano Results based on Onset Only.

3 CONCLUSIONS

A new approach for automatic transcription of piano music has been proposed for MIREX 2007 Multiple Fundamental Frequency Estimation and Tracking task. The system is based on a recent spectral approach for joint-F_0 estimation. Evaluation results on piano music pieces show the low recall rate trend of the system, which is ranked at the 7th or 8th position out of the 11 competing systems on piano music tests.

4 ACKNOWLEDGEMENT

The authors would like to thank the MIREX 2007 organization committee, and more particularly Mert Bay for the Multiple Fundamental Frequency Estimation and Tracking task.

The research leading to this paper was supported by the Groupe des Écoles des Télécommunications (GET) under project *Automatic Transcription of Music: Advanced Processing and Implementation - TAMTAM* and by the French GIP ANR under contract ANR-06-JCJC-0027-01, Décomposition en Éléments Sonores et Applications Musicales - DESAM.

5 REFERENCES

[1] M. Alonso, G. Richard, and B. David, "Extracting note onsets from musical recordings," in *Proc. of the ICME*, Amsterdam, The Netherlands, July 6-8 2005, pp. 1–4.

[2] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*. Springer, 1998.

[3] V. Emiya, R. Badeau, and B. David, "Multipitch estimation of inharmonic sounds in colored noise," in *Proc. of the 10th International Conference on Digital Audio Effects (DAFx-07)*, Bordeaux, France, September 10-15 2007.

[4] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.