A multipitch detection algorithm using a sparse decomposition with instrument-specific harmonic atoms

Pierre Leveau

August 31, 2007

Abstract

An algorithm is proposed for multi-pitch estimation. It is based on the sparse decomposition of a signal using instrument-specific harmonic atoms. Although not designed for a context where the instruments are not known, we propose to evaluate our algorithm in the open context proposed by MIREX.

1 Principle

The signal processing frontend has been previously described in [1, 2]. We described here the main features of the algorithm.

1. A dictionary of instrument-specific atoms is first built. The atoms are short waveforms parameterized by the amplitudes (resp. phases) of the partials A (resp. Φ), the time localization u, their scale s and fundamental frequency f_0:

$$h_{s,u,f_0,A,\Phi}(t) = \sum_{m=1}^{M} a_m \, e^{j\phi_m} g_{s,u,m.f_0}(t) \tag{1}$$

   The g atoms are Gabor atoms that represent the partials, and can be written as follows:

$$g_{s,u,f} = w\left(\frac{t-u}{s}\right) e^{2j\pi ft} \tag{2}$$

   where w is a time- and frequency- localized window.

   The amplitude vectors are learned on databases of isolated notes annotated in pitch and instrument, and quantized with a K-means algorithm to obtain 16 amplitude vectors per pitch. The database is made of IOWA, Studio Online et RWC databases.

2. The signal is then decomposed with this dictionary using a Matching Pursuit algorithm.

Once the decomposition algorithm has been performed, a specific post-processing step is performed to get the pitch values for each track (similar to the one proposed in [2]):

1. For each time bin, the extracted atoms whose time support overlaps with the time bin are selected,

2. The instantaneous energy of the atoms are computed at the time bin (same notations as [2]):

$$e_\lambda = (|\alpha_\lambda w(\frac{u - u_\lambda}{s_\lambda})|)^2 \tag{3}$$

3. The selected atoms are sorted in decreasing energy,

4. A parsimony criteria indicates which atoms must be kept and considered to belong to a music note:

$$P_n = \frac{\sqrt{\sum_{n'=1}^{n} e_{n'}}}{n^\beta} \tag{4}$$

   While P_n is increasing as a function of the atom index n in the atom list, the atoms are kept (similar to the criteria presented in [3]).

2 Algorithm parameters

The parameters for the decomposition are the following (refer to previous publications for explanations): $s = 46ms$, $\Delta u = 23ms$. The signal is previouslu subsampled at $Fs = 11050Hz$. f_0 is sampled logarithmically with a step of 1/10 ton. The decompositions are performed until the Signal-To-Residual ratio reaches 25 dB or 500 atoms per seconds.

References

[1] P. Leveau, E. Vincent, G. Richard, and L. Daudet. Mid-level sparse representations for timbre identification: design of an instrument-specific harmonic dictionary. In *1st Workshop on Learning the Semantics of Audio Signals*, dec 2006.

[2] P. Leveau, D. Sodoyer, and L. Daudet. Automatic Instrument Recognition in a Polyphonic Mixture using Sparse Representations. In *Proc. of Int. Conf. on Music Information Retrieval (ISMIR)*, Vienne, Autriche, sep 2007.

[3] A. P. Klapuri. Multiple fundamental frequency estimation by summing harmonic amplitudes. In *Proc. of Int. Conf. on Music Information Retrieval (ISMIR)*, 2006.