Expectation-maximization algorithm for multi-pitch estimation and separation of overlapping harmonic spectra

Roland Badeau, Valentin Emiya, Bertrand David

EXPECTATION-MAXIMIZATION ALGORITHM FOR MULTI-PITCH ESTIMATION AND SEPARATION OF OVERLAPPING HARMONIC SPECTRA

Roland BADEAU, Valentin EMIYA, Bertrand DAVID

CNRS LTCI, TELECOM ParisTech (ENST)
roland.badeau@telecom-paristech.fr

ABSTRACT

This paper addresses the problem of multi-pitch estimation, which consists in estimating the fundamental frequencies of multiple harmonic sources, with possibly overlapping partials, from their mixture. The proposed approach is based on the expectation-maximization algorithm, which aims at maximizing the likelihood of the observed spectrum, by performing successive single-pitch and spectral envelope estimations. This algorithm is illustrated in the context of musical chord identification.

Index Terms— Spectral analysis, Maximum likelihood estimation, Algorithms, Harmonic analysis, Envelope detection.

1. INTRODUCTION

Estimating the fundamental frequency (or pitch) of a harmonic signal is a prominent task in audio signal processing. Its main difficulty lies in the intrinsic ambiguity of the harmonic model, which typically leads to octave errors. However this problem can be circumvented by taking the smoothness of the spectral envelope into account [1]. In presence of multiple sources, the multi-pitch estimation task is even more difficult, because of the spectral overlap between the harmonic components (see [2,3] for a review). There are two categories of methods for performing this task:

- iterative approaches, which recursively estimate the dominant pitch, and remove its harmonics from the mixture;
- joint approaches, which simultaneously estimate all the fundamental frequencies by optimizing a joint criterion.

Methods in the first category are fast, but they tend to accumulate errors at each iteration, because of the spectral overlap between partials belonging to different sources. Methods in the second category do not accumulate errors, but they are more computationally demanding, since they require to explore a space whose dimension is the number of harmonic components in the mixture (which can be circumvented by reducing this search space beforehand, like in [4]).

In this paper, we propose a new approach for addressing the multi-pitch estimation task, which combines the advantages of both categories: we optimize a joint criterion by means of a recursive algorithm which performs successive single-pitch estimations. This is achieved by applying the expectation-maximization (EM) approach [5] to an appropriate spectrum model, which provides a proper statistical framework to our method. The paper is organized as follows: the spectral mixture model is introduced in Section 2, and the EM algorithm is described in Section 3. Numerical simulations are presented in Section 4, and the conclusions are summarized in Section 5. The following notations will be used throughout the paper:

- we use normal symbols for scalars, underlined symbols for vectors, and doubly underlined symbols for matrices;
- for any integer variable N, $\overline{N}$ denotes the set $\{0 \dots N-1\}$.

The research leading to this paper was supported by the French GIP ANR under contract ANR-06-JCJC-0027-01, Décompositions en Éléments Sonores et Applications Musicales - DESAM, and by the European Commission under contract FP6-027026, Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content - KSPACE.

2. SPECTRAL MIXTURE MODEL

We denote $\{Y_i\}_{i\in\overline{I}}$ the samples of the outcome power spectrum (where $\overline{I} = \{0 \dots I-1\}$ is the set of all frequency bins), and $\underline{Y}$ the random vector $[Y_0 \dots Y_{I-1}]^T$. This power spectrum is modeled as the squared magnitude of a sum of J complex spectra $\underline{X}_j = [X_{0,j} \dots X_{I-1,j}]^T$, whose presence or absence at frequency i is indicated by a Boolean variable $B_{i,j}$, plus a complex spectrum $\underline{N} = [N_0 \dots N_{I-1}]^T$ corresponding to an additive noise present at all frequencies, so that $Y_i = |N_i + \sum_{j\in\overline{J}} B_{i,j} X_{i,j}|^2$. This model should be understood as follows: since we aim at modeling a mixture of harmonic spectra, $\underline{X}_j$ is the j^{th} complete spectrum, and the Booleans $B_{i,j} \in \mathcal{B} = \{0,1\}$ act as a selector for the harmonic frequencies (vector $\underline{B}_j = [B_{0,j} \dots B_{I-1,j}]^T$ is typically shaped as a comb).

2.1. Spectral envelope model

As usually assumed in the literature, for all $j \in \overline{J}$, we suppose that $\underline{X}_j$ is a complex Gaussian random vector, of zero mean, and covariance matrix $\mathrm{diag}(\underline{s}_j)$, where $\underline{s}_j = [s_{0,j} \dots s_{I-1,j}]^T$ (the outcome values at all frequency bins are assumed independent). In the same way, $\underline{N}$ is a complex Gaussian random vector, of zero mean, and covariance matrix $\mathrm{diag}(\underline{s}_J)$. All those spectral components are assumed mutually independent. Moreover, $\forall j \in \overline{J}\cup\{J\}$ ($j = J$ referring to the noise component), the smooth spectral envelope $\underline{s}_j \in \mathbb{R}_+^I$ is parameterized by a moving average (MA) model of order K:

$$s_{i,j} = \sigma_j^2 \left| \sum_{k=0}^{K} \alpha_{k,j} e^{-2i\pi k \frac{i}{I}} \right|^2 \tag{1}$$

where $\forall j \in \overline{J} \cup \{J\}$, $\alpha_{0,j} = 1$.

2.2. Harmonicity model

All Booleans $B_{i,j}$ are assumed independent, and we denote $\pi_{i,j} \in]0,1[$ the probability that $B_{i,j} = 1$. The harmonicity of the spectral components is expressed as follows: we consider L subsets $\mathcal{H}_l$ of $\overline{I}$, which locate the harmonic frequency bins for L pitch candidates. We further assume that for all the spectral components, *i.e.* $\forall j \in \overline{J}$, there is a unique pitch identified by an index $l_j \in \overline{L}$, and two

constants $\dot{\pi}_j$ and $\mathring{\pi}_j$ in $]0,1[$, such that $\forall i \in \mathcal{H}_{l_j}, \pi_{i,j} = \dot{\pi}_j$, and $\forall i \notin \mathcal{H}_l, \pi_{i,j} = \mathring{\pi}_j$ (all harmonic frequency bins in $\mathcal{H}_{l_j}$ have the same high probability of presence $\dot{\pi}_j$, and the other ones have the same low probability of presence $\mathring{\pi}_j$, as illustrated in Fig. 1). For all $i \in \overline{I}$, let $\underline{B}^i = [B_{i,0} \dots B_{i,J-1}]^T$. Since all Booleans $B_{i,j}$ are independent, for any vector $\underline{b} \in \mathcal{B}^J$, the probability that $\underline{B}^i = \underline{b}$ is

$$\pi_{i,\underline{b}} = \prod_{j \in \overline{J}} (\pi_{i,j})^{b_j} (1 - \pi_{i,j})^{1-b_j}. \tag{2}$$

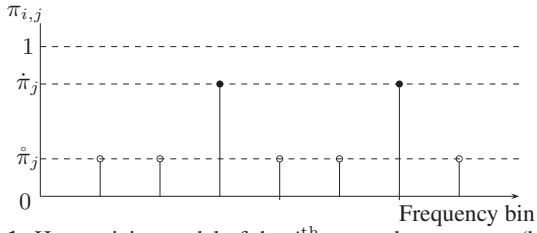

Fig. 1. Harmonicity model of the j^{th} spectral component (here the fundamental frequency equals 3 Fourier bins; • stands for frequencies in $\mathcal{H}_{l_j}$, ∘ stands for frequencies not in $\mathcal{H}_{l_j}$).

2.3. Resulting mixture model

For any vector $\underline{b} \in \mathcal{B}^J$, the mixture $N_i + \sum_{j \in \overline{J}} b_j X_{i,j}$ is a complex Gaussian random variable, of zero mean, and variance

$$s_{i,\underline{b}} = s_{i,J} + \sum_{j \in \overline{J}} b_j s_{i,j}, \tag{3}$$

where $s_{i,J}$ refers to the noise component (see section 2.1). The conditional law of Y_i given that $\underline{B}^i = \underline{b}$ is a chi-square distribution with two degrees of freedom, of probability density function

$$p(Y_i | \underline{B}^i = \underline{b}) = \frac{1}{s_{i,\underline{b}}} e^{-\frac{Y_i}{s_{i,\underline{b}}}}. \tag{4}$$

3. EXPECTATION-MAXIMIZATION ALGORITHM

In this section, we propose an EM algorithm to estimate the spectral mixture model. Here the observations are the power spectrum samples Y_i, and the hidden states are the Booleans $B_{i,j}$, arranged in the $I \times J$ matrix $\underline{\underline{B}}$. Their joint probability is written in the form

$$p(\underline{Y}, \underline{\underline{B}}) = \prod_{i \in \overline{I}} \sum_{\underline{b} \in \mathcal{B}^J} \mathbf{1}_{\{\underline{B}^i = \underline{b}\}} \pi_{i,\underline{b}}\, p(Y_i | \underline{B}_i = \underline{b}),$$

where $\pi_{i,\underline{b}}$ was defined in equation (2). Substituting equation (4) into the above equation leads to the following expression of the joint log-likelihood $L(\underline{Y}, \underline{\underline{B}}) = \ln(p(\underline{Y}, \underline{\underline{B}}))$:

$$L(\underline{Y}, \underline{\underline{B}}) = \sum_{i \in \overline{I}} \sum_{\underline{b} \in \mathcal{B}^J} \mathbf{1}_{\{\underline{B}^i = \underline{b}\}} \left(\ln \left(\frac{\pi_{i,\underline{b}}}{s_{i,\underline{b}}} \right) - \frac{Y_i}{s_{i,\underline{b}}} \right)$$

To estimate the set of prior probabilities $\pi_{i,j}$, arranged in the $I \times J$ matrix $\underline{\underline{\pi}}$, and the set of envelope coefficients $s_{i,j}$, arranged in the $I \times (J+1)$ matrix $\underline{\underline{s}}$, the EM algorithm generates a sequence $(\underline{\underline{\pi}}^n, \underline{\underline{s}}^n)_{n \geq 0}$ by recursively maximizing the conditional expectation $Q^n_{\underline{\underline{\pi}}, \underline{\underline{s}}} = \mathbb{E}\left[L(\underline{Y}, \underline{\underline{B}}) | \underline{Y}; \underline{\underline{\pi}}^n, \underline{\underline{s}}^n \right]$ with respect to (w.r.t.) the model parameters: $(\underline{\underline{\pi}}^{n+1}, \underline{\underline{s}}^{n+1}) = \underset{\underline{\underline{\pi}}, \underline{\underline{s}}}{\operatorname{argmax}}\, Q^n_{\underline{\underline{\pi}}, \underline{\underline{s}}}$. An important result about the EM algorithm is that the log-likelihood of the sequence $(\underline{\underline{\pi}}^n, \underline{\underline{s}}^n)_{n \geq 0}$ generated in this way is non-decreasing.

Here $Q^n_{\underline{\underline{\pi}}, \underline{\underline{s}}}$ is the sum of two terms $Q^n_{\underline{\underline{\pi}}}$ and $Q^n_{\underline{\underline{s}}}$, defined as

$$Q^n_{\underline{\underline{\pi}}} = \sum_{i \in \overline{I}} \sum_{\underline{b} \in \mathcal{B}^J} \gamma^n_{i,\underline{b}} \ln(\pi_{i,\underline{b}}) \tag{5}$$

$$Q^n_{\underline{\underline{s}}} = \sum_{i \in \overline{I}} \sum_{\underline{b} \in \mathcal{B}^J} \gamma^n_{i,\underline{b}} \left(\ln \left(\frac{1}{s_{i,\underline{b}}} \right) - \frac{Y_i}{s_{i,\underline{b}}} \right) \tag{6}$$

where $\gamma^n_{i,\underline{b}}$ is the posterior probability that $\underline{B}^i = \underline{b}$ given Y_i.

3.1. E-Step: updating the posterior probabilities

The Bayes theorem proves that

$$\gamma^n_{i,\underline{b}} = \frac{\pi^n_{i,\underline{b}}\, p(Y_i | \underline{B}^i = \underline{b}; \underline{s}^n)}{p(Y_i; \underline{\underline{\pi}}^n, \underline{\underline{s}}^n)} \propto \frac{\pi^n_{i,\underline{b}}}{s^n_{i,\underline{b}}} e^{-\frac{Y_i}{s^n_{i,\underline{b}}}}.$$

according to equation (4). These posterior probabilities are thus calculated by using the property $\sum_{\underline{b} \in \mathcal{B}^J} \gamma^n_{i,\underline{b}} = 1$. The posterior marginal probability $\gamma^n_{i,j}$ that $B_{i,j} = 1$ given Y_i is then obtained according to

$$\gamma^n_{i,j} = \sum_{\underline{b} \in \mathcal{B}^J \,\&\, b_j = 1} \gamma^n_{i,\underline{b}}. \tag{7}$$

3.2. M-step: updating the model parameters

3.2.1. Updating the prior probabilities

We are now looking for the prior probabilities $\underline{\underline{\pi}}^{n+1}$ which maximize $Q^n_{\underline{\underline{\pi}}}$ defined in equation (5). First, equations (2) and (7) prove that $Q^n_{\underline{\underline{\pi}}}$ can be decomposed in the form $Q^n_{\underline{\underline{\pi}}} = \sum_{j \in \overline{J}} Q^n_j$, where $Q^n_j = \sum_{i \in \overline{I}} \gamma^n_{i,j} \ln(\pi_{i,j}) + (1 - \gamma^n_{i,j}) \ln(1 - \pi_{i,j})$. The prior probabilities are thus estimated by independently maximizing each Q^n_j w.r.t. the I variables $\pi_{0,j} \dots \pi_{I-1,j}$, which are entirely defined by the three parameters $(l_j, \dot{\pi}_j, \mathring{\pi}_j)$, as described in section 2.2. In particular, this maximization involves estimating the pitch the j^{th} component, via parameter l_j. We thus write Q^n_j as a function of the triplet $(l, \dot{\pi}, \mathring{\pi}) \in \overline{L} \times]0,1[\times]0,1[$, which will have to be maximized to obtain the estimates $(l^{n+1}_j, \dot{\pi}^{n+1}_j, \mathring{\pi}^{n+1}_j)$:

$$Q^n_j(l, \dot{\pi}, \mathring{\pi}) = |\mathcal{H}_l|\, \dot{Q}^n_{l,j}(\dot{\pi}) + (I - |\mathcal{H}_l|)\, \mathring{Q}^n_{l,j}(\mathring{\pi}) \tag{8}$$

where $|\mathcal{H}_l|$ denotes the cardinality of set $\mathcal{H}_l$ and

$$\begin{cases} \dot{Q}^n_{l,j}(\dot{\pi}) = \frac{1}{|\mathcal{H}_l|} \sum_{i \in \mathcal{H}_l} \gamma^n_{i,j} \ln(\dot{\pi}) + (1 - \gamma^n_{i,j}) \ln(1 - \dot{\pi}) \\ \mathring{Q}^n_{l,j}(\mathring{\pi}) = \frac{1}{I - |\mathcal{H}_l|} \sum_{i \notin \mathcal{H}_l} \gamma^n_{i,j} \ln(\mathring{\pi}) + (1 - \gamma^n_{i,j}) \ln(1 - \mathring{\pi}) \end{cases} \tag{9}$$

For all $l \in \overline{L}$, we independently maximize functions $\dot{Q}^n_{l,j}$ and $\mathring{Q}^n_{l,j}$ w.r.t. $\dot{\pi}$ and $\mathring{\pi}$, resulting in the optimal values

$$\begin{cases} \dot{\pi}^n_{l,j} = \frac{1}{|\mathcal{H}_l|} \sum_{i \in \mathcal{H}_l} \gamma^n_{i,j} \\ \mathring{\pi}^n_{l,j} = \frac{1}{I - |\mathcal{H}_l|} \sum_{i \notin \mathcal{H}_l} \gamma^n_{i,j}. \end{cases} \tag{10}$$

Substituting equation (10) into equation (9), we obtain the maximal values of $\dot{Q}^n_{l,j}(\dot{\pi})$ and $\mathring{Q}^n_{l,j}(\mathring{\pi})$ w.r.t. $\dot{\pi}$ and $\mathring{\pi}$:

$$\begin{cases} \dot{Q}^n_{l,j} = \dot{\pi}^n_{l,j} \ln(\dot{\pi}^n_{l,j}) + (1 - \dot{\pi}^n_{l,j}) \ln(1 - \dot{\pi}^n_{l,j}) \\ \mathring{Q}^n_{l,j} = \mathring{\pi}^n_{l,j} \ln(\mathring{\pi}^n_{l,j}) + (1 - \mathring{\pi}^n_{l,j}) \ln(1 - \mathring{\pi}^n_{l,j}) \end{cases}$$

Equation (8) becomes $Q^n_{l,j} = |\mathcal{H}_l|\,\dot{Q}^n_{l,j} + (I - |\mathcal{H}_l|)\,\mathring{Q}^n_{l,j}$. Maximizing this expression w.r.t. l yields $l_j^{n+1} = \underset{l}{\operatorname{argmax}}\, Q^n_{l,j}$, which identifies the estimates $\dot{\pi}_j^{n+1} = \dot{\pi}^n_{(l_j^{n+1},j)}$ and $\mathring{\pi}_j^{n+1} = \mathring{\pi}^n_{(l_j^{n+1},j)}$. The joint probabilities $\pi^{n+1}_{i,\underline{b}}$ are then calculated via equation (2).

3.2.2. Updating the envelope coefficients

Now we are looking for the coefficients σ_j^{n+1} and $\alpha_{k,j}^{n+1}$ which maximize $Q^n_{\underline{s}}$ defined in equation (6). Note that in this equation, $s_{i,\underline{b}}$ depends on σ_j and $\alpha_{k,j}$ via equations (3) and (1) (condition $b_j = 1$ is relevant for $j \in \overline{J}$; for $j = J$, this condition is assumed always true in the following developments).

3.2.2.1. Estimation of the variances

Differentiating equation (6) w.r.t. σ_j^2 shows that $\forall j \in \overline{J} \cup \{J\}$, $\sigma_j^2 \frac{\partial Q^n_{\underline{s}}}{\partial \sigma_j^2} = \rho_j^+ - \rho_j^-$, where $\left\{ \begin{array}{rcl} \rho_j^- & = & \sum\limits_{i\in\overline{I}} \sum\limits_{\underline{b}\in\mathcal{B}^J \,\&\, b_j=1} \gamma^n_{i,\underline{b}} \frac{s_{i,j}}{s_{i,\underline{b}}} \\ \rho_j^+ & = & \sum\limits_{i\in\overline{I}} \sum\limits_{\underline{b}\in\mathcal{B}^J \,\&\, b_j=1} \gamma^n_{i,\underline{b}} \frac{s_{i,j}}{s_{i,\underline{b}}} \frac{Y_i}{s_{i,\underline{b}}} \end{array} \right.$

Zeroing this derivative does not admit a closed-form solution, but it can be proved that the multiplicative update rule $\sigma_j^2 \leftarrow \frac{\rho_j^+}{\rho_j^-}\,\sigma_j^2$ forms an ascent method which converges to the maximum of $Q^n_{\underline{s}}$. In practice, we perform one such update per iteration of the EM algorithm, which still guarantees that the log-likelihood $Q^n_{\underline{s}}$ is non-decreasing.

3.2.2.2. Estimation of the MA parameters

Differentiating (6) w.r.t. $\alpha_{k,j}$ yields $\forall k \in \{0 \dots K\}$, $\forall j \in \overline{J} \cup \{J\}$, $\frac{1}{\sigma_j^2} \frac{\partial Q^n_{\underline{s}}}{\partial \alpha_{k,j}} = \sum\limits_{\kappa=0}^{K} \alpha_{\kappa,j}\, r_j^+(k-\kappa) - \sum\limits_{\kappa=0}^{K} \alpha_{\kappa,j}\, r_j^-(k-\kappa)$, where

$$\left\{ \begin{array}{rcl} r_j^-(k) & = & \sum\limits_{i\in\overline{I}} 2\cos\left(2\pi k \frac{i}{I}\right) \sum\limits_{\underline{b}\in\mathcal{B}^J \,\&\, b_j=1} \frac{\gamma^n_{i,\underline{b}}}{s_{i,\underline{b}}} \\ r_j^+(k) & = & \sum\limits_{i\in\overline{I}} 2\cos\left(2\pi k \frac{i}{I}\right) \sum\limits_{\underline{b}\in\mathcal{B}^J \,\&\, b_j=1} \frac{\gamma^n_{i,\underline{b}}}{s_{i,\underline{b}}} \frac{Y_i}{s_{i,\underline{b}}} \end{array} \right.$$

Let $\underline{\alpha}_j = [\alpha_{0,j} \dots \alpha_{K,j}]^T$, and $\underline{\underline{R}}_j^-$ and $\underline{\underline{R}}_j^+$ be the $(K+1) \times (K+1)$ Toeplitz matrices whose coefficients are $r_j^-(k)$ and $r_j^+(k)$, respectively. Again, the maximum does not admit a closed-form expression, but the multiplicative update rule $\underline{\alpha}_j \leftarrow (\underline{\underline{R}}_j^-)^{-1}(\underline{\underline{R}}_j^+)\underline{\alpha}_j$ guarantees that the log-likelihood $Q^n_{\underline{s}}$ is non-decreasing. The convergence of such multiplicative rules for estimating MA models was studied in [6]. Note that at each iteration, vector $\underline{\alpha}_j^n$ should be remapped into the set of minimum phase filters, following the approach proposed in [6].

3.3. Discussion

The two main advantages of this EM algorithm are its low complexity (multi-pitch estimation is performed by means of J successive single-pitch estimations, instead of exploring a vector space of dimension J), and its ability to handle spectral overlap between the harmonic components (by taking the smoothness of the spectral envelopes into account). However, the log-likelihood function $Q^n_{\underline{\pi},\underline{s}}$ to be optimized is generally not smooth, and the algorithm tends to stay trapped in local maxima[1]. Therefore it is necessary to develop strategies to escape from these maxima, for instance by testing multiples or sub-multiples of the current estimated frequencies. Initialization is also an important point. An approach (illustrated in section 4) would be to use the EM algorithm as a refinement, after a first stage of basic multi-pitch estimation (any fast multi-pitch estimator can be used, such as [3]). Indeed, we observed that this algorithm is capable of correcting some wrongly estimated pitches when initialized properly. To summarize, the algorithm requires the following inputs:

- the initial values of the multiple pitches via the indexes l_j. We then propose to initialize the a priori probabilities $\pi_{i,j}$ as harmonic combs, as illustrated in Fig. 1, with $\dot{\pi}_j = 1 - \varepsilon$, $\mathring{\pi}_j = \varepsilon$, and $\varepsilon \ll 1$. Finally, the posterior probabilities $\gamma_{i,j}$ are taken equal to $\pi_{i,j}$.
- the initial spectral envelopes. We propose to take flat MA envelopes ($\alpha_{k,j} = 0\ \forall k \geq 1$), with variances of different orders, and to let the multiplicative updates presented in section 3.2.2 converge (contrary to the EM algorithm itself, where only one update is performed per iteration).

Note that in practice, like any EM-based approach, this algorithm has to be implemented very carefully to avoid numerical errors, due to possible divisions by numbers much smaller than one.

4. SIMULATION RESULTS: MUSICAL CHORD IDENTIFICATION

In this section, our EM algorithm is applied to an audio-like synthetic signal. Music signals indeed provide an interesting, challenging field of application to our algorithm, since they often contain mixtures of harmonic components with a substantial spectral overlap (which is typically the case of consonant chords). Here we use a sampling frequency of 22 kHz, and we consider a chord formed of $J = 4$ notes, of fundamental frequencies 330, 440, 550, and 660 Hz (corresponding to musical notes E4 - A4 - C♯5 - E5, forming a major triad involving an octave, which generates a large spectral overlap). For modeling the spectral envelopes of the three notes, we use MA models of order 5, whose coefficients are normalized so that the maximal values of the envelopes of the four notes are -30, -10, -20, and 0 dB, respectively. The amplitudes $X_{i,j}$ of the partials of the four notes are then obtained by sampling their MA envelope at the harmonic frequencies (their phases being generated as independent, uniformly distributed random variables). The additive noise N_i is then synthesized as a white Gaussian noise, whose variance is chosen so that its constant power density function (PSD) is -60 dB. We compute $I = 1000$ samples of this mixture signal, corresponding to a 45 ms-long frame, which is a typical analysis length for multi-pitch estimation. The digital Fourier transform of the signal is then computed with the same number of samples I, without applying a tapering window, in order to avoid spectral leakage, and correlation between adjacent Fourier bins. The periodogram of the resulting signal is represented in Fig. 3-(a).

The EM algorithm is applied with $L = 88$ pitch candidates, distributed according to the piano MIDI scale which ranges from note A0 to C8, including the $J = 4$ true pitches. To simulate a first stage of basic multi-pitch estimation, the algorithm is initialized with the following pitches : 330, 880, 748, and 660 Hz, corresponding to notes E4, A5, F♯5, and E5 (two erroneous pitches out of four). The MA envelopes are initialized as constant PSD of magnitude 0 dB, and -40 dB for the noise part. Fig. 2 represents the 4 detection functions $l \mapsto Q^n_{l,j}$ obtained after $n = 25$ iterations of the algorithm. It is worth noticing that the multi-pitch estimation problem is reduced to four independent single-pitch estimation problems, and that the detection functions admit a strong maximum at the true pitch values.

[1] Local maxima are inherent to the pitch estimation problem. Such maxima appear for instance in the pitch detection functions represented in Fig. 2.

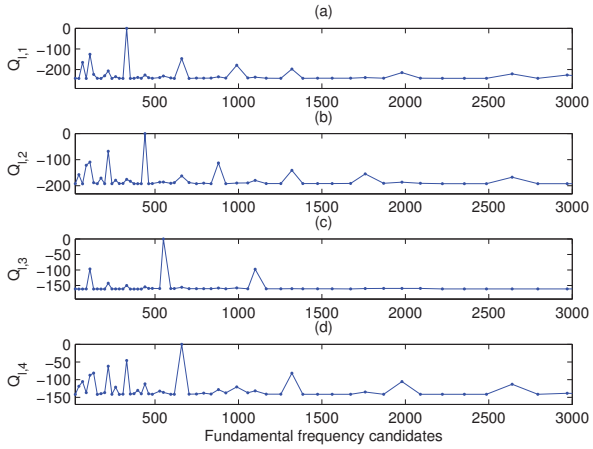

Fig. 2. Pitch detection functions (dB)
(a) E4, 330 Hz, (b) A4, 440 Hz, (c) C♯5, 550 Hz, (d) E5, 660 Hz

The output of the algorithm also permits to separate the magnitude spectra of the four notes. Indeed, for all $j \in \overline{J} \cup \{J\}$, the squared magnitude of the j^{th} spectrum component $|X_{i,j}|^2$ (if $j < J$) or $|N_i|^2$ (if $j = J$) can be estimated as a weighted sum of products between the outcome spectrum Y_i and the Wiener-like filters $\frac{s_{i,j}^2}{s_{i,\underline{b}}^2}$: $\widehat{|X|}_{i,j}^2 = \sum_{\underline{b} \in \mathcal{B}^J \,\&\, b_j = 1} \pi_{i,\underline{b}} \frac{s_{i,j}^2}{s_{i,\underline{b}}^2} Y_i$. The separated spectra are represented in Fig. 3-(b-f) (solid lines), superimposed with the original MA envelopes (dashed lines). Note that the spectral components are efficiently estimated. However in the case of overlapping partials, the amplitude of the weakest one tends to be under-estimated (the estimation of the even harmonics of note E4 in Fig. 3-(b) is perturbed by the partials of note E5, represented in Fig. 3-(e)).

5. CONCLUSIONS

In this paper, we introduced a novel approach for multi-pitch estimation, based on the statistical framework of the EM algorithm. The proposed method is particularly promising, due to its robustness to overlapping partials, and its capacity to simplify the multi-pitch estimation task into successive single-pitch estimations. It requires a proper initialization, involving a first stage of basic multi-pitch estimation for instance, and could advantageously make use of heuristics, in order to avoid to stay trapped in local maxima. The effectiveness of this approach is confirmed by our simulations, performed on audio-like synthetic signals. In order to apply this algorithm to real audio signals, and compare its performance to that of competing methods, some additional improvements will be helpful:

- The design of the sets $\mathcal{H}_l$ of harmonic frequencies may be refined by means of peak-picking in the Fourier spectrum. In the case of inharmonic instruments such as the piano, the inharmonicity coefficient may also be included in the model.
- Since the estimation of a PSD usually involves windows such as Hann or Hamming, the spectral leakage and correlation between adjacent Fourier bins should be integrated in the model.
- The number J of harmonic components, which is supposed known herein this paper, could be estimated by incorporating a statistical criterion, such as BIC [7].

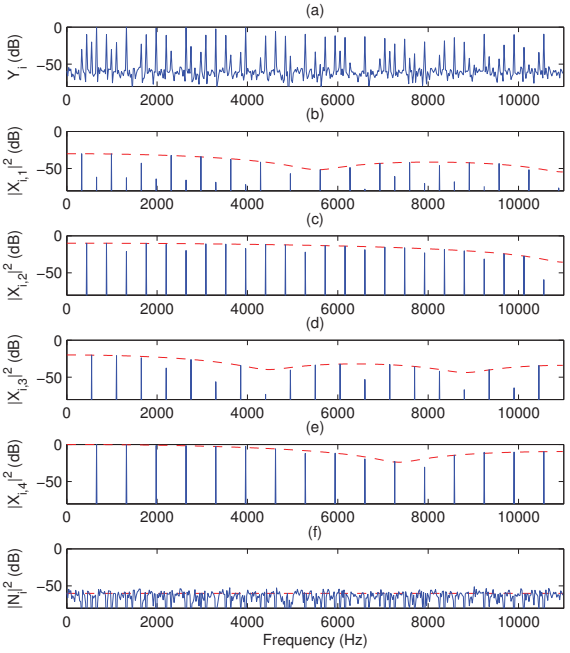

Fig. 3. Observed (a) and estimated (b-f) spectra

Besides, the proposed framework permits to chose various models for the spectral envelopes, which could also be parameterized by using cepstral representations for instance, or mixtures of template spectra, either defined according to a psychoacoustic frequency scale, such as Mel, Bark, and ERB, or learned from a database of harmonic signals.

6. REFERENCES

[1] A. Klapuri, "Multiple fundamental frequency estimation by harmonicity and spectral smoothness," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 6, pp. 804–816, Nov. 2003.

[2] A. Klapuri and M. Davy, Eds., *Signal Processing Methods for Music Transcription*, Springer, New York, 2006.

[3] A. de Cheveigné, D. Wang, and G.J. Brown, Eds., *Computational Auditory Scene Analysis: Principles, Algorithms and Applications*, chapter Multiple F0 estimation, Wiley-IEEE Press, Piscataway, NJ, 2006.

[4] V. Emiya, R. Badeau, and B. David, "Multipitch estimation of inharmonic sounds in colored noise," in *Proc. of 10th International Conference on Digital Audio Effects DAFx-07*, Bordeaux, France, Sept. 2007, pp. 93–98.

[5] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[6] R. Badeau and B. David, "Weighted maximum likelihood autoregressive and moving average spectrum modeling," in *Proc. of 2008 International Conference on Acoustics, Speech, and Signal Processing ICASSP'08*, Las Vegas, Nevada, USA, Apr. 2008, pp. 3761–3764, IEEE.

[7] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.