HARMONIC TRACKING USING SEQUENTIAL MONTE CARLO

Corentin Dubois

IRCCyN UMR 6597
1 rue de la Noë, BP 92101
44321 Nantes Cedex 03, France
corentin.dubois@irccyn.ec-nantes.fr

Manuel Davy

LAGIS UMR 8146
Cité scientifique, BP 48
59651 Villeneuve d'Ascq Cedex, France
manuel.davy@ec-lille.fr

ABSTRACT

This paper extends our previous work about individual frequencies detection and tracking to the extraction of harmonic components from multicomponent signals. Once harmonic components are detected, we study their time evolution applying statistical filtering to track them in the time-frequency plane. A nonlinear observation model based on spectrogram is defined and a state-of-the-art particle filtering algorithm is developed. Applications are mainly audio signals.

1. INTRODUCTION

Several important applications require sequential frequency tracking. Typical examples are Computer Acoustic Scene Analysis (CASA), Speech/Music Processing and vibration monitoring of mechanical systems. This paper builts on our previous work [1] which addressed individual frequencies tracking. However, the signals involved in the above mentioned applications generaly have significant energy at overtone partial frequencies, in addition to energy at fundamental frequencies. This paper proposes a *harmonic model* aimed to track jointly fundamentals and overtone partials of possibly several harmonic components[1].

Time-Frequency Representations (TFRs) [2, 3] are a natural framework for joint detection and tracking of time-varying harmonic components. TFRs such as spectrograms usualy separate the individual frequency components and allow modelling of their evolution. In [1], we proposed a Markov model for several individual frequency components detection and tracking. This paper extends the previous model to Markov modelling of harmonic components, still in spectrograms. In the following, we use the standard spectrogram definition: let $\mathbf{x} = [x_1, x_2, \ldots]$, be a discrete time signal, its spectrogram (with window $\mathbf{w}$ of length $L_{\mathbf{w}}$) is

$$\mathrm{SPEC}^{\mathbf{w}}_{t,f}(\mathbf{x}) = \left|\mathrm{STFT}^{\mathbf{w}}_{t,f}(\mathbf{x})\right|^2, \qquad (1)$$

where t denotes discrete time and f denotes (normalized) discrete frequency. The Short Time Fourier Transform (STFT) is

$$\mathrm{STFT}^{\mathbf{w}}_{t,f}(\mathbf{x}) = \mathrm{DFT}(\mathbf{x} \cdot \mathbf{w}), \qquad (2)$$

where DFT denotes the Discrete Fourier Transform. The spectrogram maps the data into the Time-Frequency (TF) plane from which the estimation of the number k_t of harmonic components, the time-varying partial frequencies $\mathbf{f}_t$ and amplitudes $\mathbf{a}_t$ at time t is physically sound. The fallout of this choice is the loss of phase information. However, it is not penalizing because the final aim of our method is harmonic components tracking and not the reconstruction of the signal.

Similar to target tracking approaches, detection and tracking of harmonic components are performed within the sequential Bayesian framework. The statistical filtering problem consists of sequentially estimating the discrete parameter k_t and the continuous vector $\boldsymbol{\theta}_t = \left[\mathbf{f}_t\ ,\ \mathbf{a}_t\right]$ whose size depends on k_t. This requires the definition of a likelihood (observation equation) and a sequential prior (transition equation). The observation vector at time t is the current spectrogram "column":

$$\mathbf{y}_t = \mathrm{SPEC}^{\mathbf{w}}_{t,f}(\mathbf{x}) \text{ for } f \in \{0 \ldots 0.5\}. \qquad (3)$$

The likelihood is denoted $p\big(\mathbf{y}_t | k_t, \boldsymbol{\theta}_t\big)$ and the prior is denoted $p\big(k_t, \boldsymbol{\theta}_t | k_{t-1}, \boldsymbol{\theta}_{t-1}\big)$. Together with $p\big(k_0, \boldsymbol{\theta}_0\big)$, they form a *Jump Markov System* [4].

In order to perform detection and tracking, we use Sequential Monte Carlo (SMC) methods. SMC methods allow on-line parameters estimation by combining the powerful Monte Carlo sampling methods with Bayesian inference [5]. In this case, the algorithm is often called a *Particle Filter* (PF). We stress that, the observation model being nonlinear a Kalman filter cannot be implemented, whereas a PF can take into account the state changes of dimension and the observation model nonlinearities [6].

This paper is organized as follows: in Section 2, we introduce the Bayesian sequential harmonic model. In Section 3, we describe an efficient SMC algorithm (PF), spe-

[1]Throughout the paper, we term *harmonic component* a funtamental frequency together with its overtone partials.

cially designed for this problem. In Section 4, we present some results and conclusions are proposed in Section 5.

2. SEQUENTIAL BAYESIAN HARMONIC MODEL

At each time t, we assume that the observation $\mathbf{y}_t$ is the squared magnitude of the STFT of a sum of k_t harmonic components [7]. Each harmonic component j ($j = 1 \dots k_t$) is composed of H sinusoids with frequencies denoted $f_{t,j,h}$ for $h = 1, \dots, H$. The fundamental corresponds to $h = 1$ and H is the fixed number of overtone partials. Because of the harmonic structure, we have $f_{t,j,h} \approx hf_{t,j,1}$ for $h = 2, \dots, H$. In order to model possible inharmonicity[2], the overtone partials are also estimated. At time t, the frequencies vector

$$\mathbf{f}_t = \big[f_{t,1,1}\,,\, f_{t,1,2}\,, \dots,\, f_{t,1,H}\,, \quad \dots \\ \dots \quad ,\, f_{t,k_t,1}\,,\, f_{t,k_t,2}\,, \dots,\, f_{t,k_t,H}\big]^{\mathrm{T}} \tag{4}$$

and the amplitudes vector

$$\mathbf{a}_t = \big[a_{t,1,1}\,,\, a_{t,1,2}\,, \dots,\, a_{t,1,H}\,, \quad \dots \\ \dots \quad ,\, a_{t,k_t,1}\,,\, a_{t,k_t,2}\,, \dots,\, a_{t,k_t,H}\big]^{\mathrm{T}} \tag{5}$$

to be estimated have dimension k_tH.

2.1. Observation equation

The frequencies and the amplitudes are assumed stationnary with respect to the length $L_\mathbf{w}$ of $\mathbf{w}$ (this is the standard underlying assumption of the spectrogram). Let us define the function g_t by:

$$\begin{array}{rccc} g_t: & \mathbb{N} \times \mathbb{R}^{2Hk_t} & \longrightarrow & \mathbb{R}^{L_\mathbf{w}} \\ & (k_t, \boldsymbol{\theta}_t) & \longmapsto & \big|\mathrm{DFT}(\mathbf{s}_t \cdot \mathbf{w})\big|^2 \end{array} \tag{6}$$

with, for $\tau = 1, \dots, L_\mathbf{w}$

$$\mathbf{s}_{t,\tau} = \sum_{j=1}^{k_t} \sum_{h=1}^{H} a_{t,j,h} \cos(2\pi f_{t,j,h} \tau) \tag{7}$$

In practice, the actual number of overtone partials can be less than H if the frequency is higher than the Shannon frequency. The amplitude of such overtone partial is set to 0. The function g_t (Eq. (6)) maps the observation generative model (Eq. (7)) from the time domain to the frequency domain. Moreover, the phase information is lost by applying the function g_t, thus it needs not be estimated (signal reconstruction is not necessary in the foreseen applications). The observation equation is, for $\xi = 1, \dots, L_\mathbf{w}$:

$$\mathbf{y}_t[\xi] = g_t(k_t, \boldsymbol{\theta}_t) + v_t^\mathbf{y}[\xi] \tag{8}$$

[2]Harmonicity is a standard phenomenon in music which arises when overtone partial frequencies deviate from integer multiples of the fundamental frequency. Instruments such as Pianos produce inharmonic sounds.

where $v_t^\mathbf{y}$ is a zero-mean white Gaussian noise of variance $r_t^\mathbf{y}$. We allow the variance $r_t^\mathbf{y}$ to change with time according to $\log\left(r_t^\mathbf{y}\right) = \log\left(r_{t-1}^\mathbf{y}\right) + \varepsilon_{t-1}$ with $\varepsilon_{t-1} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Implementing time-varying variances allows the model to adapt to different situations such as stationnarity and quick frequency/amplitude changes[3]. To insure the positivity of the variances, the evolution model is on the variance logarithm. The value of σ_ε is fixed roughly and do not require fine tuning.

The likelihood is obtained from Eq. (8) and is given by:

$$p\big(\mathbf{y}_t|(k_t, \boldsymbol{\theta}_t)\big) = \\ \left(2\pi r_t^\mathbf{y}\right)^{-\frac{L_\mathbf{w}}{2}} \exp\left(-\frac{\left\|\mathbf{y}_t - g_t(k_t, \boldsymbol{\theta}_t)\right\|^2}{2r_t^\mathbf{y}}\right) \tag{9}$$

2.2. Transition equation

For generality reasons, we do not impose specific dynamic equations to manage the evolution of the unknown parameters from time $t-1$ to time t. For instance, such equations are hard to define for environmental sounds because they are composed of several kinds of sound (speech, transport sounds, urban noises, etc ...). Moreover, considering the standard stationnarity assumption of the spectrogram, there is no reason for abrupt change of frequency or amplitude. Thus, at time t, the transition equations are given by random walks centered on parameter estimate at time $t-1$. For $j = 1, \dots, k_t$ and $h = 1, \dots, H$:

$$f_{t,j,h} = f_{t-1,j,h} + v_{t-1,j,h}^\mathbf{f} \tag{10}$$

$$a_{t,j,h} = a_{t-1,j,h} + v_{t-1,j,h}^\mathbf{a} \tag{11}$$

where $v_{t-1,j,h}^{(.)}$ is a zero-mean white noise with Gaussian density of variance $r_{t-1,j,h}^{(.)}$. Similar to the observations, we allow the variance $r_{t-1,j,h}^{(.)}$ to evolve according to, for $j = 1, \dots, k_t$ and $h = 1, \dots, H$:

$$\log\left(r_{t,j,h}^\mathbf{f}\right) = \log\left(r_{t-1,j,h}^\mathbf{f}\right) + \varphi_{t-1,j,h} \tag{12}$$

$$\log\left(r_{t,j,h}^\mathbf{a}\right) = \log\left(r_{t-1,j,h}^\mathbf{a}\right) + \alpha_{t-1,j,h} \tag{13}$$

with

$$\varphi_{t-1,j,h} \sim \mathcal{N}(0, \sigma_\varphi^2) \tag{14}$$

$$\alpha_{t-1,j,h} \sim \mathcal{N}(0, \sigma_\alpha^2) \tag{15}$$

Our purpose is to allow parameters variance to change with time and a simple and efficient choice is again a random walk as dynamics to manage this evolution [4].

[3]We understand here quick changes as rapid evolutions of the frequency/amplitude at the time scale of the spectrogram, that is one order of magnitude higher than the window length.

	$k_{t-1} = k_{min}$	$k_{min} < k_{t-1} < k_{max}$	$k_{t-1} = k_{max}$
b_t	$1/10$	$1/10$	0
e_t	$9/10$	$8/10$	$9/10$
d_t	0	$1/10$	$1/10$

Table 1. Transition probabilities for the number k_t of harmonic components. The probabilities b_t, e_t and d_t are defined in Eq. (16).

The model also involves the discrete parameter k_t. Its transition equation is:

$$k_t = k_{t-1} + \begin{cases} 1 & \text{with probability } b_t \\ 0 & \text{with probability } e_t \\ \text{-1} & \text{with probability } d_t \end{cases} \quad (16)$$

where the probabilities b_t, e_t and d_t are given in Table 1. The precise values of the probabilities do not have so much importance. The aim is to allow k_t to increase or decrease but the probability to keep it constant must be prominent. This corresponds to the heuristic that the number of harmonic components changes at a time scale higher than time interval between $t-1$ and t. The number of components ranges from k_{min} to k_{max}.

3. PARTICLE FILTERING ALGORITHM

The main idea of SMC methods is the use of a random and adaptative grid to approximate the posterior density of the state [8, 9]. More precisely, the posterior $p(k_{0:t}, \boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ is defined, up to a constant, as the product between the likelihood and the prior. The particle filter provides, at each time t, a set of N weighted particles

$$\left\{\left(k_{0:t}^{(i)}, \boldsymbol{\theta}_{0:t}^{(i)}\right), \omega_t^{(i)}\right\}_{i=1\ldots N}$$

where $\omega_t^{(i)}$ is the weight, which approximate this posterior as follows:

$$\hat{p}(k_{0:t}, \boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) = \sum_{i=1}^{N} \omega_t^{(i)} \delta_{(k_{0:t}^{(i)}, \boldsymbol{\theta}_{0:t}^{(i)})}(k_{0:t}, \boldsymbol{\theta}_{0:t}) \quad (17)$$

Assuming such a particles set is available at time $t-1$, the PF algorithm updates this set from time $t-1$ to time t and thus yields an approximation of the posterior at time t. This update procedure has two main steps: first, each particle $k_{0:t-1}^{(i)}, \boldsymbol{\theta}_{0:t-1}^{(i)}$ is extended with a new state $k_t^{(i)}, \boldsymbol{\theta}_t^{(i)}$ proposed by the proposal density $q_t\big(k_t, \boldsymbol{\theta}_t|k_{t-1}^{(i)}, \boldsymbol{\theta}_{t-1}^{(i)}\big)$. Then, the weights are updated:

$$\omega_t^{(i)} \propto \omega_{t-1}^{(i)} \times p\big(\mathbf{y}_t|k_t^{(i)}, \boldsymbol{\theta}_t^{(i)}\big) \frac{p\big(k_t^{(i)}, \boldsymbol{\theta}_t^{(i)}|k_{t-1}^{(i)}, \boldsymbol{\theta}_{t-1}^{(i)}\big)}{q_t\big(k_t^{(i)}, \boldsymbol{\theta}_t^{(i)}|k_{t-1}^{(i)}, \boldsymbol{\theta}_{t-1}^{(i)}\big)} \quad (18)$$

Table 2. Outline of the particle filtering implemented for our problem, at time t.

- At time $t = 0$
  - ○ For $i = 1 \ldots N$
    - **Initialization**
      - – Sample $\big(k_0^{(i)}, \boldsymbol{\theta}_0^{(i)}\big)$ according to $p(k_0, \boldsymbol{\theta}_0)$
      - – Set $\bar{\boldsymbol{\theta}}_0^{(i)} = \mathbb{E}[\boldsymbol{\theta}_0^{(i)}]$
- At time $t \geq 1$
  - ○ For $i = 1 \ldots N$
    - **Update the particles**
      - – Sample $k_t^{(i)}$ according to the probabilities given in Table 1.
      - – Sample $\boldsymbol{\theta}_t^{(i)} \sim q_t(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(i)}) = \mathcal{N}(\bar{\boldsymbol{\theta}}_t^{(i)}, \mathbf{P}_{tt}^{(i)})$ where $\bar{\boldsymbol{\theta}}_t^{(i)}$ and $\mathbf{P}_{tt}^{(i)}$ are computed by the Unscented Tranform and Kalman filter equations.
      - – Set $\big(k_{0:t}^{(i)}, \boldsymbol{\theta}_{0:t}^{(i)}\big) = \big(k_t^{(i)}, \boldsymbol{\theta}_t^{(i)}, k_{0:t-1}^{(i)}, \boldsymbol{\theta}_{0:t-1}^{(i)}\big)$
    - **Update the hyperparameters**
    - **Evaluate the importance weights**
  - ○ **Normalization:** the importance weights are normalized so that the sum equal to 1.
  - ○ **Selection step:** multiply/suppress particles according to their weights.
  - ○ **Output:** from the particle approximation of the posterior density, compute the conditional marginal mean: $\mathbb{E}[h_t(\boldsymbol{\theta}_{0:t})] \approx \frac{1}{N}\sum_{i=1}^{N} h_t(\boldsymbol{\theta}_{0:t}^{(i)})$ with $h_t(\boldsymbol{\theta}_{0:t}) = \boldsymbol{\theta}_t$.

where $\propto$ denotes "proportional to". The weights are computed up to a normalizing constant. A normalization step will insure the sum equal to 1. The likelihood $p\big(\mathbf{y}_t|k_t^{(i)}, \boldsymbol{\theta}_t^{(i)}\big)$ and the prior $p\big(k_t^{(i)}, \boldsymbol{\theta}_t^{(i)}|k_{t-1}^{(i)}, \boldsymbol{\theta}_{t-1}^{(i)}\big)$ are defined in Sections 2.1 and 2.2 respectively. The design of the proposal distribution is of paramount importance in sequential importance sampling algorithms. Any densities q_t are possible, provided its support is wider than that of the posterior. The proposal density must be as close as possible of the posterior and it must be easy to sample from it. It has been shown [10, 11] that particle filters with a proposal distribution obtained using the Unscented Kalman Filter (UKF) are particulary suitable if the model of observation is nonlinear, which is the case here. Another critical point in SMC methods is that of weights degeneracy. After a few number of iterations, one particle emerges and has a weight equal to 1 whereas all the other particles are rejected with a weight equal to 0. The fallout of this is a loss of variability. To avoid this, a selection step is introduced [6] which multiply particles with high weights and suppress particles with low weights. After this step, the particles are still distributed according to the posterior with all the weights equal to $\frac{1}{N}$.

We present in Table 2 the outline of the particle filtering algorithm implemented here.

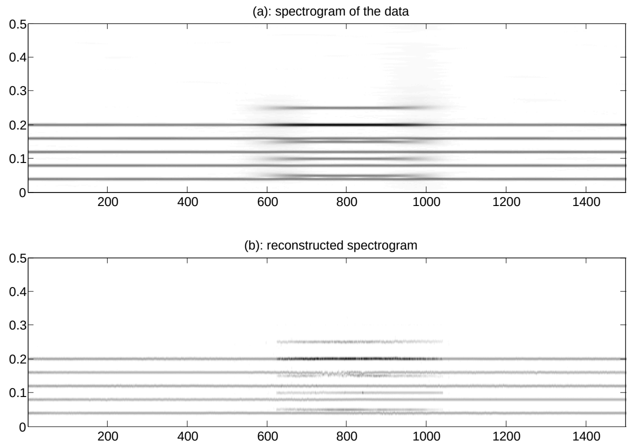

Fig. 1. (a): Spectrogram of the data. The column of this spectrogram at time t is the observation vector $\mathbf{y}_t$. (b): Reconstructed spectrogram. The column of this spectrogram at time t is given by $\sum_{i=1}^{N} \omega_t^{(i)} g_t(k_t^{(i)}, \boldsymbol{\theta}_t^{(i)})$.

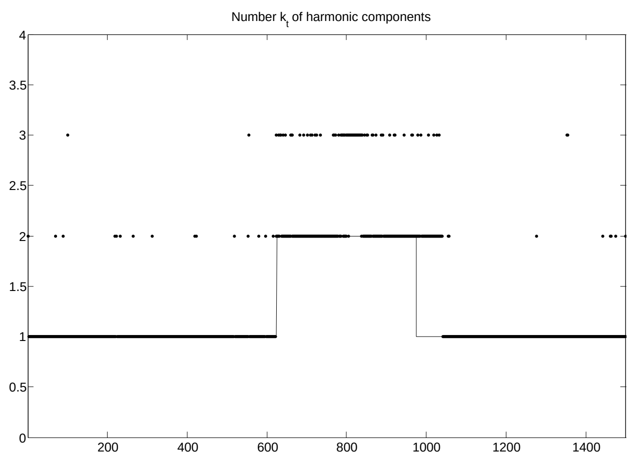

Fig. 2. Number k_t of harmonic components. The doted line is the estimation of k_t and the solid line indicates the simulated value of k_t.

4. RESULTS

We obtained good results with our previous model [1] which addressed the issue of individual time-varying frequencies detection and tracking. This paper involves improvements over it in terms of harmonic components modelling. We apply this new model implemented by the algorithm given in Table 2 to synthetic signals to test the efficiency of our method. Experiments on synthetic signals also allow the comparaison between the estimated harmonic components and the ones used to synthesize the signal. In [1] we have shown that our method is able to track time-varying frequencies. Here, tests are mainly aimed to verify its capability to detect (birth and death) harmonic components. An example of this study is represented on Fig.'s 1-2-3. The simulated signal is composed of one stationary harmonic component with (normalized) fundamental frequency equals to 0.04 with 4 overtone partials. A second harmonic component appears at time sample 624 and disappears at time sample 973. Its fundamental frequency is 0.05 also with 4 overtone partials. The two fundamental frequencies are chosen such that the two harmonic components have one common overtone partial frequency. A Gaussian white noise of variance 0.05 is added to the data. The following set of parameters was chosen: $N = 500$, $H = 6$, $k_{min} = 1$, $k_{max} = 3$, $\sigma_\varphi^2 = \sigma_\alpha^2 = 0.01$, $\sigma_\epsilon^2 = 0$ (in our example, the variance $r_t^{\mathbf{y}}$ do not change with time and is fixed $r_t^{\mathbf{y}} = r^{\mathbf{y}} = 0.25$) and $\mathbf{w}$ is a Hamming window of length 255.

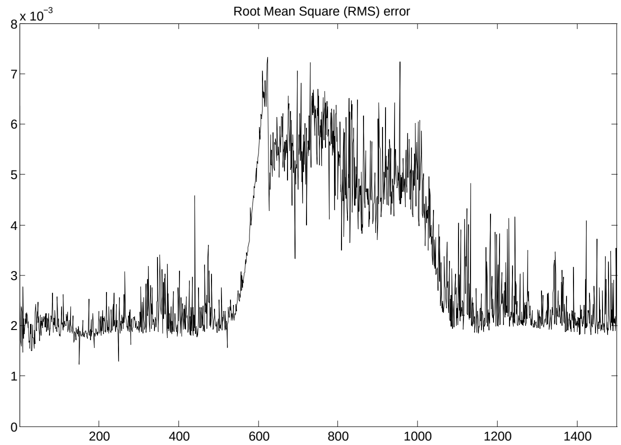

Fig. 3. Root Mean Square (RMS) error. At time t, the RMS error is computed by $\sqrt{\frac{\sum_{i=1}^{N} \omega_t^{(i)} \|\mathbf{y}_t - g_t(k_t^{(i)}, \boldsymbol{\theta}_t^{(i)})\|^2}{L_{\mathbf{w}}}}$.

In Fig. 1, we can see that the harmonic components frequencies are estimated with a good accuracy whereas estimated amplitudes are less precise. This is due to the value of the observation variance $r^{\mathbf{y}}$ which roughly determines if the likelihood is peaked or not. If the likelihood is very peaked, amplitudes estimation will be very accurate but the algorithm will have some troubles to manage changes of harmonic components number and inversely. The chosen value of $r^{\mathbf{y}}$ has to address this compromise. Fig. 2 shows the estimated number of harmonic components. This estimation is performed as follows: at time t, particles can be gather in $k_{max} - k_{min}$ sets according to the value of $k_t^{(i)}$, $i = 1 \ldots N$. Within each set, all particles have the same value of k_t. We sum the weights of the particles of each set and the estimated value of k_t is the one corresponding to the set with the higher weight. The comparaison with Fig. 1, shows that when the estimated value of k_t is not equal to the simulated one, the extra harmonic components have very weak amplitudes. Another estimation procedure taking into account the harmonic components amplitudes may give a better estimated value. We also note that when the second harmonic component disappears, there is a time delay before k_t decreases. This is due to the windowing effect, which affects both the time and the frequency resolutions of the spectrogram.

5. CONCLUSION

Our previous work shows that detection and tracking of several instantaneous frequencies can be performed from the spectrogram using particle filtering. Based on this work, we have modified our model to take into account possibly harmonic structures (fundamental and overtone partials). Futhermore, we do not assume any parametric model for partials amplitude envelop. We obtained promising results on synthetic signals and experiments on real data will follow this paper.

6. REFERENCES

[1] C. Dubois, M. Davy, and J. Idier, "Tracking of time-frequency components using particle filtering," in *Proc. IEEE ICASSP*, Philadelphia, PA, USA, Mar. 2005.

[2] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and Quadratic Time-Frequency Signal Representations," *IEEE SP Magazine*, pp. 21–67, Apr. 1992.

[3] P. Flandrin, *Temps-fréquence*, Hermès, 1993.

[4] C. Andrieu, M. Davy, and A. Doucet, "Efficient Particle Filtering for Jump Markov Systems. Application to Time-Varying Autoregressions," *IEEE trans. on Signal Processing*, July 2003.

[5] Z. Chen, "Bayesian filtering: From kalman filters to particle filters, and beyond," *On line: http://soma.crl.mcmaster.ca/∼zhechen/homepage.htm*, 2003.

[6] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 2001.

[7] M. Davy, S. J. Godsill, and J. Idier, "Bayesian Estimation and Analysis of Polyphonic Harmonic Music," *Journal of the Acoustical Society of America*, Aug. 2004, under second revision.

[8] A. Doucet, "On sequential simulation-based methods for bayesian filtering," Tech. Rep., Signal Processing Group, Department of engineering, University of Cambridge CB2 1PZ Cambridge, 1998.

[9] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE trans. on Signal Processing*, vol. 50, Feb. 2002.

[10] S. J. Julier and J. K. Uhlmann, "A new extension of the kalman filter to nonlinear systems," in *Proc. SPIE*, 1997, vol. 3068, pp. 182–193.

[11] R. van der Merwe, A. Doucet, N. de Freitas, and E. Wan, "The unscented particle filter," Tech. Rep., Department of engineering, University of Cambridge CB2 1PZ Cambridge, 2000.