TRACKING OF PARTIALS FOR ADDITIVE SOUND SYNTHESIS USING HIDDEN MARKOV MODELS

Ph. Depalle, G. García & X. Rodet

IRCAM, 31 rue Saint Merri, 75004 Paris, France
Tel: (33-1) 44-78-48-45
email:phd@ircam.fr & garcia@ircam.fr

A B S T R A C T

In this paper, we present a sinusoidal partial tracking method for additive synthesis of sound. Partials are tracked by identifying time functions of parameters as underlying trajectories in successsive set of spectral peaks. This is done by a purely combinatorial Hidden Markov Model. We consider a partial trajectory as a time-sequence of peaks which satisfies continuity constraints on parameter slopes. Our method allows frequency line crossing and can be used for formant tracking.

1. INTRODUCTION

This paper presents a sinusoidal partial tracking method, based on Hidden Markov Models (HMM), for additive synthesis of sound. Additive synthesis, one of the highest quality synthesis methods in musical applications, is based on a model which represents a sound signal $s[n]$ at a sampling rate F_e as a sum of J sinusoids $c_j[n]$ (called "partials"), with time-varying frequency f_j, amplitude a_j and phase φ_j , $1 \le j \le J$:

$$s[n] = \sum_{j=1}^{J} c_j[n] = \sum_{j=1}^{J} a_j[n] \cos(\Phi_j[n])$$

with $\Phi_j[n] =$

$$\Phi_j[n-1] + \frac{2\pi}{F_e} f_j[n] + \varphi_j[n] - \varphi_j[n-1]$$

This signal model is good for precise and independent control of the time evolution of each partial. This is the main reason why this model has been widely used, despite certain drawbacks such as: high computation cost, absence of noise components and difficult user control. It should be noted here that synthesis by inverse Fourier transform [1] overcomes many of these problems.

When simulating a natural sound, the time functions, defining the frequency, amplitude and phase evolution of each partial, are generally obtained by performing an automatic analysis of this sound. The complete procedure of additive analysis/synthesis is shown in the diagram 1.

The FFT block computes Short Term Fourier Transforms of successive signal frames which overlap in time. For each frame, the peak detection block extracts the set of spectral peaks defined by their frequencies, amplitudes and phases. Some of these peaks belong to partial lines while others are spurious peaks (due to noise or analysis window lobes). The peak matching block has to track the partials by identifying time functions of parameters as underlying trajectories in the whole set of peaks.

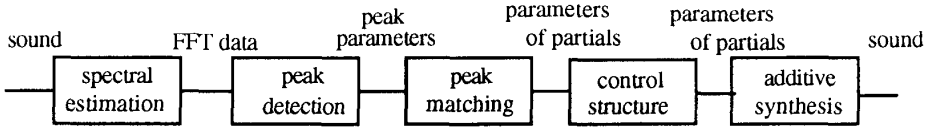
This paper focuses on the peak matching procedure, which is the most difficult part of the analysis process. The following section give an overview of the existing techniques. Section 3 presents the proposed peak matching method and Section 4 shows some experimental results.

2. OVERVIEW OF TECHNIQUES

Two categories of techniques are briefly discussed. The first one includes techniques designed for additive synthesis of sound which do not use HMM. The second category includes techniques based on HMM which are not designed for synthesis of sounds.

The technique presented in [2] is developed for speech synthesis applications; it does not consider spurious

Diagram 1



0-7803-0946-4/93 $3.00

spectral peaks and is not well suited to identify time varying frequency partials. In [3] the method, designed for musical applications, solves some of the problems of the preceding technique, but peak matching is done by optimising each trajectory independently and locally in time. Consequently, the resulting trajectories are not optimal in a global sense. Another approach, which is useful when dealing with harmonic signals, is based on a logical filtering of the whole set of spectral peaks. It selects the greatest amplitude peak in each frequency band centred on a multiple of the fundamental frequency. Evidently this method fails when used on inharmonic sounds.

The techniques presented in [4] and [5] use HMM for frequency line tracking. The use of HMM allows us to globally optimise the trajectories, but they are not well-suited to our problem here because *"there is no explicit notion of data association and the tracks are not considered separately" [5]* and also because [4], [5] use constraints which are too rigorous on the signal.

3. METHOD DESCRIPTION

3.1. Principle [6]

First, let us consider the intuitive model of trajectory that we use: a trajectory is a time-sequence of peaks which satisfies continuity constraints on parameter slopes. Consequently, our method tends to identify trajectories whose amplitude and frequency slopes evolve smoothly in time. This criteria, deduced from our observations, is different from standard criteria: we retain the continuity of slopes as preferable to the continuity of values.

Secondly, let us describe the type of Hidden Markov Model [7] that we formalised:

At time k, there are h_k peaks $P_k[j]$, $0 \le j < h_k$ ordered by growing frequency. Each partial or trajectory is labelled by an index greater than zero. The problem is to associate an index $I_k[j]$, $0 \le j < h_k$ to each peak $P_k[j]$. When a peak $P_k[j]$ is detected as a spurious one, it is associated with a null index $I_k[j] = 0$.

At time k, a state S_k is defined by an ordered pair of vectors (I_{k-1}, I_k) and the observation is defined by an ordered pair of integers (h_{k-1}, h_k). Notice that we only retain the combinatorial aspect and that frequency and amplitude parameters of peaks are not taken as observations. In fact they are considered parameters of the Markov Model and are used to compute the transition probability between two states. Also notice that the probability of observation of (m,n) is always equal to unity for the states defined by ordered pairs of vectors of size m and n, and is zero for all others.

Determining the set of trajectories is equivalent to finding the optimal sequence of states S_k from which we derive the corresponding sequence of vectors I_k. In practice, we find the optimal sequence of states by means of the Viterbi algorithm, which maximises the joint probability of state and observation sequences leading to a globally optimal solution.

3.2. State transition probabilities calculation

The only model parameters we must compute are the state transition probabilities. Let us consider a transition from a state S_{k-1} into a state S_k. It is clear that three frames of peaks are concerned, corresponding to times k-2, k-1 and k, as the states S_{k-1} and S_k are defined by the vectors I_{k-2}, I_{k-1}, and I_k. Let $f_k(j)$ and $a_k(j)$ be the frequency and amplitude of the peak j in the frame k. For each peak j of frame k, $0 \le j < h_k$, we evaluate a "matching criterion" $\theta_k(j)$ which depends on two other peaks t and r of frames k-2 and k-1 respectively, such that $I_{k-2}(t)=I_{k-1}(r)=I_k(j)$. This matching criterion is defined by equation 1

where $\Delta a_k(j,r) = a_k(j) - a_{k-1}(r)$, $\Delta f_k(j,r) = f_k(j) - f_{k-1}(r)$ and μ, σ_f, σ_a, σ_{fn}, and σ_{an} are free parameters which may vary in time. For simplicity's sake, the time interval between two frames is assumed to be constant.

It can be seen that this matching criterion penalises the fact that spurious peaks are matched (i.e. assigned to partials) and that peaks corresponding to partials are not matched. In other words, it favours the continuity of parameter slopes when $I_k(j)>0$ (i.e. peaks t, r and j are assigned to partials) and penalises this continuity when $I_k(j)=0$ (i.e. peaks t, r and j are spurious ones).

In the case $I_k(j)=0$, the peaks t and r are chosen among all the spurious peaks of frames k-2 and k-1 so that they minimise the quantity

$$\Delta f_k(j,r,t) = \Delta f_k(j,r) - \Delta f_{k-1}(r,t).$$

Equation 1

$$\theta_k(j) = \begin{cases} \exp\left\{ -\dfrac{\left[\Delta f_k(j,r) - \Delta f_{k-1}(r,t)\right]^2}{\sigma_f^2} - \dfrac{\left[\Delta a_k(j,r) - \Delta a_{k-1}(r,t)\right]^2}{a_k^2(j)\sigma_a^2} \right\} & \text{if } I_k(j) > 0 \\ \left\{1-(1-\mu)\exp\left\{-\dfrac{\left[\Delta f_k(j,r) - \Delta f_{k-1}(r,t)\right]^2}{\sigma_{fn}^2}\right\}\right\} \times & \\ \left\{1-(1-\mu)\exp\left\{-\dfrac{\left[\Delta a_k(j,r) - \Delta a_{k-1}(r,t)\right]^2}{a_k^2(j)\sigma_{an}^2}\right\}\right\} & \text{if } I_k(j) = 0 \end{cases}$$

Let the states currently considered at times k-1 and k be $S_{k-1}=x$ and $S_k=y$ respectively. We calculate a global score $g_{xy}(k)$ over all the peaks of frame k as

$$g_{xy}(k) = \prod_{j=0}^{n_k - 1} \theta_k(j).$$

As the transition probabilities $\alpha_{xs_k}(k)$ from the state $S_{k-1}=x$ into all possible states S_k must satisfy the condition

$$\sum_{s_k} \alpha_{xs_k}(k) = 1$$

we obtain this probabilities by normalising the global scores $g_{xy}(k)$ in the following manner:

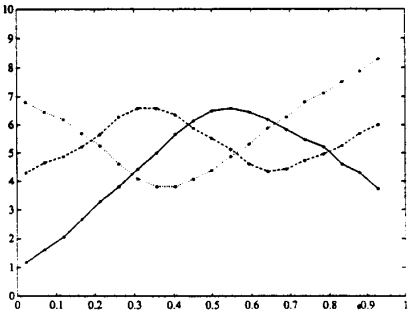
$$\alpha_{xy}(k) = g_{xy}(k) \Big/ \sum_{s_k} g_{xs_k}(k)$$

3.3. Algorithm

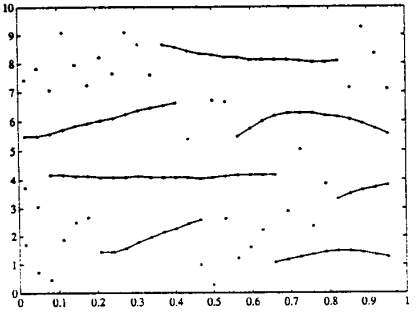
The computational complexity of the algorithm depends strongly on the number of combinations of indexes in the vectors I_k, which determines the number of states. It also depends on the number of frames on which frequency lines have to be optimised. In order to limit this complexity (otherwise the method is not practicable) we use the Viterbi algorithm on a window of T frames length which slides frame by frame and we introduce some constraints on the index combinations. For instance, for a given window "births" and "deaths" of partials are disallowed to reduce the number of possible states, forcing the state sequences to have a constant number of trajectories. Consequently, there always exist two integers t and r which satisfy the condition $I_{k-2}(t)=I_{k-1}(r)=I_k(j)$. To take into account births and deaths of partials on the whole set of frames, we use a higher level procedure. Births are detected by looking for partials in the current window which did not exist in the preceding one; deaths are found by looking for partials in the preceding window which disappear in the current one. Consequently, the length T fixes the shortest duration of a partial. Other constraints can be eventually introduced if necessary such as: defining a maximum number of indexes greater than zero in vectors I_k, selecting a reduced set of possible number of indexes, constraining the number of partial crossings to be under a given value (possibly zero), or limiting the value of the frequency slopes $\Delta f_k(j,r)$.

4. EXPERIMENTAL RESULTS

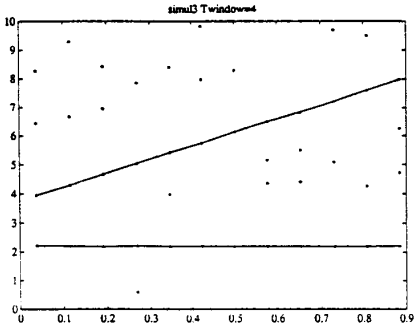
The following figures show some experimental results obtained by applying our method to simulated and real data. In the two first examples peaks are set by hand. In the third one we process real data extracted from a synthetic signal. In the fourth one we track the partials of an inharmonic sound. In the fifth one we try on a unvoiced/voiced transition of speech to test the ability to detect births. The last example shows that the method can be applied without any modification to formant tracking applications.



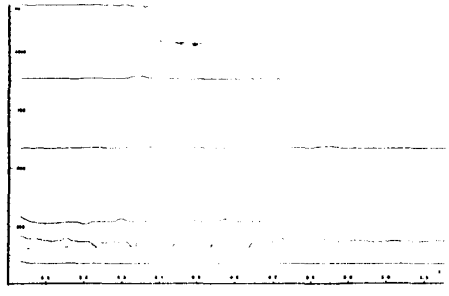
Crossing of partials (simulated data).



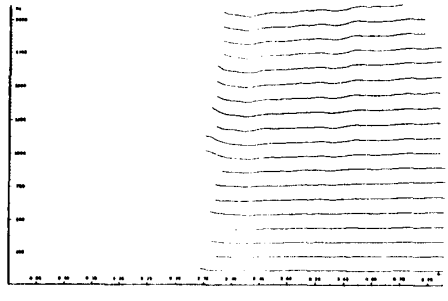
Detection of births and deaths of partials (simulated data).



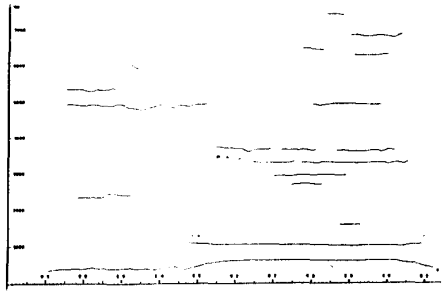
Tracking of two sinusoids embedded in white Gaussian noise (simulated data).



Tracking of partials of a percussive sound (real data).



Tracking of partials of speech signal (diphone /sa/).



Tracking of formants of the same speech signal (diphone /sa/).

5. CONCLUSION

We have shown that a purely combinatorial Hidden Markov Model applied to frequency line tracking leads to very promising results. In addition, the use of parameter slopes instead of parameter values allows for tracking variable frequency lines and solves the problem of frequency line crossing. Our Hidden Markov Model application has been implemented in a modular way so that constraints can be chosen before running it. We are currently working on the reduction of he computational cost of the algorithm by refining the set of constraints.

6. REFERENCES

[1] Xavier Rodet & Philippe Depalle, "A new additive synthesis method using inverse Fourier transform and spectral envelopes", ICMC, San Jose, October 1992.

[2] Robert McAulay & Thomas Quatieri, "Speech analysis/synthesis based on a sinusoïdal representation", IEEE Trans. on Acoust., Speech, and Signal Proc., vol. ASSP-34, August 1986.

[3] Xavier Serra, "A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition", Philosophy Dissertation, Stanford University, October 1989.

[4] Roy Streit & Ross Barrett, "Frequency line tracking using Hidden Markov Models", IEEE Trans. on Acoust., Speech, and Signal Proc., vol. ASSP-38, April 1990.

[5] Xianya Xie & Robin Evans, "Multiple target tracking and multiple frequency line tracking using Hidden Markov Models", IEEE Trans. on Acoust., Speech, and Signal Proc., vol. ASSP-39, December 1991.

[6] G. Garcia, "Analyse des signaux sonores en termes de partiels et de bruit. Extraction automatique des trajets fréquentiels par des modèles de Markov cachés", Mémoire de DEA en automatique et traitement de signal, Orsay, 1992.

[7] Lawrence Rabiner & Biing-Hwang Juang, "An introduction to Hidden Markov Models", IEEE ASSP Magazine, January 1986.