ERROR COMPENSATION IN MODELING TIME-VARYING SINUSOIDS

Wen Xue, Mark Sandler

Centre for Digital Music, Dept. Elec. Eng.
Queen Mary, Univ. of London, UK
{xue.wen|mark.sandler}@elec.qmul.ac.uk

ABSTRACT

In this article we propose a method to improve the accuracy of sinusoid modeling by introducing parameter variation models into both the analyzer and the synthesizer. Using the least-square-error estimator as an example, we show how the sinusoidal parameters estimated under a stationary assumption relate to the real nonstationary process, and propose a way to reestimate the parameters using some parameter variation model. For the synthesizer, we interpolate the parameters using the same model, with the phase unwrapping process reformulated to adapt to the change. Results show that the method effectively cuts down the systematic error of a conventional system based on a least-square-error estimator and the McAulay-Quatieri synthesizer.

1. INTRODUCTION

Sinusoid modeling expresses a pitched sound as the linear combination of time-varying sinusoids. This involves an analyzer and a synthesizer. The analyzer does sinusoidal parameter estimation and the synthesizer rebuilds the signal from the estimated parameters. However, since parameter estimators rarely consider the in-frame dynamics, most current sinusoid modeling systems carry an inborn error even if the signal being analyzed strictly matches the sinusoid model. This error is often ignorable when a clean residue is not crucial, but becomes significant when one tries to subtract a sinusoid from the original. The systematic error is a combination of an analyzer error and a synthesizer error. In section 2 we give a short review of sinusoid modeling. Section 3 explains how the analyzer error occurs and how it can be mended. Section 4 gives an example on what we can do to cut down the synthesizer error.

2. SPECTRAL MODELING SYNTHESIS

The complete sinusoid modeling was first presented by McAulay and Quatieri [1] and later refined by several. The parameter estimator has been improved by more accurate parameter estimation methods, such as those summarized in [2]. Partial tracking has been improved by using more natural tracking methods to connect spectral peaks in consecutive frames [3, 4]. The model itself has been extended to involve non-stationary noise [5]. On the synthesizer side, the reconstruction process proposed in [1] remains the same. Suppose we have a single time-varying complex sinusoid

$$x(n) = a(n)e^{j\varphi(n)}, \text{where } \varphi(n)-\varphi(n-1) = 2\pi\int_{n-1}^{n} f(t)dt \quad (1)$$

where n is a sampled version of the continuous time, t. The *instantaneous* amplitude $a(n) > 0$ and frequency $f(t) > 0$ are slow-varying functions of time. The complete sinusoid model is constructed by summing up multiple sinusoids (*partials*) in the form of (1). For each sinusoid, the analyzer estimates its amplitudes, frequencies and phase angles at a set of points $n_0, n_1, ..., n_F$, $F+1$ being the total number of measurements. We denote the estimates as $\hat{a}(n_0)$, $\hat{f}(n_0)$, etc. We use the term *estimator error* to refer to the difference between the estimates and their true values, such as between $\hat{a}(n_1)$ and $a(n_1)$. The analyzer also includes a peak tracker which forms sinusoids from local spectral peaks. When the signal has two or more partials, there may also be a peak tracker error.

The McAulay-Quatieri synthesizer connects two consecutive measure points with a sinusoid segment by interpolating the parameters. As the true parameter variation laws usually do not coincide with the interpolation laws, this interpolation introduces a synthesizer error.

Using all the $F+1$ measured parameter sets, we can reconstruct a sinusoid covering the interval $[n_0, n_F]$. We denote the rebuilt signal $\hat{x}(n), n_0 \le n \le n_F$, and define the *relative modeling error* as

$$e = \frac{1}{n_F - n_0 + 1}\sum_{n=n_0}^{n_F}\left\|\frac{\hat{x}(n)-x(n)}{a(n)}\right\|^2 \quad (2)$$

in which the difference of corresponding values is normalized by the instantaneous amplitude. When the signal being analyzed strictly fits the model, this final error (2) is a combined result of estimator, peak tracker, and synthesizer errors.

3. PARAMETER REESTIMATION

Most sinusoid modeling systems use a frame-based spectral analyzer. For each frame, the amplitude and frequency are assumed constant during the whole frame and calculated from the short-time Fourier transform, along with a phase angle [2]. In most cases the results are intuitively assigned to be the instantaneous parameters at the frame centre.

Let N be the frame width. As any estimate is calculated from N data points, it depends on N instantaneous parameter sets rather than equals its instantaneous value at $N/2$, the frame centre. Instead of using the estimates directly, we try to recover the instantaneous values from a sequence of multiple estimates, using parameter variation information derived from the estimates. That is, given a sequence of parameter estimates, we try to find a sinusoid in the form of (1) which, when fed into the analyzer, generates the given estimates. To do this we need to study the quantitative relation between the true parameters and the frame-based estimates. Naturally the relation depends how the parameters are estimated.

In this article we use a least-square-error (LSE) estimator for parameter measurement, which minimizes the square error between the spectrum of a pure sinusoid and that of a narrow-band signal. In short, for any pure sinusoid with parameter set (a, f, φ), we can calculate its spectrum $ae^{j\varphi}H_f(k)$. Given a narrow-band spectrum $X(k)$, the LSE estimator finds the frequency $\hat{f}$ that

maximizes the inner product $\langle X, H_f \rangle$ by the norm, which can be shown to yield the least square error. The amplitude and phase angle are then given by $\hat{a}e^{j\hat{\varphi}} = \langle X, H_{\hat{f}} \rangle / \|H\|^2$, where $\|H\|^2$ is a normalizing factor determined by the window function. In this article the window function is always real, symmetric and lowpass.

Let our data frame be

$$x(n) = a(n)e^{j\left(\varphi_c + 2\pi \int_{N/2}^{n} f(t)dt\right)}, \ 0 \le n < N \tag{3}$$

Let the window function be $w(n)$, $\varphi_{mn} = 2\pi \int_m^n f(t)dt$, $\varphi_n = \varphi_{\frac{N}{2},n}$, then the short-time Fourier transform of $x(n)$ is

$$X_k = \sum_{n=0}^{N-1} w(n)a(n)e^{j\left(\varphi_c + \varphi_n - 2\pi k \frac{n}{N}\right)}, \ 0 \le k < N \tag{4}$$

The Fourier transform of a pure zero-phase unit sinusoid is

$$H_k = \sum_{n=0}^{N-1} w(n)e^{j2\pi\left(f_0\left(n-\frac{N}{2}\right) - \frac{kn}{N}\right)}, \ 0 \le k < N \tag{5}$$

Define $b(n) \equiv w(n)^2 a(n)$, $\Delta\varphi_{mn}(g) = \varphi_{mn} - 2\pi(n-m)g$, we calculate the square norm of $\langle X, H \rangle$

$$\begin{aligned} \|\langle X, H \rangle\|^2 \ = \ & N^2 \left(\sum_{n=0}^{N-1} b(n)^2 \right. \\ & \left. + \ 2 \sum_{n=1}^{N-1} \sum_{m=0}^{n-1} b(n)b(m) \cos \Delta\varphi_{mn}(f_0) \right) \end{aligned} \tag{6}$$

To maximize the above we set its derivative regarding f_0 to 0:

$$\begin{aligned} \frac{d\|\langle X, H \rangle\|^2}{df_0} &= N^2 2\pi \sum_{n=1}^{N-1} \sum_{m=0}^{N-1} (n-m)b(n)b(m) \sin \Delta\varphi_{mn}(\hat{f}) \\ &= 0 \end{aligned} \tag{7}$$

where $\hat{f}$ is the frequency that maximizes $\|\langle X, H \rangle\|^2$, *i.e.* the LSE estimate. Let $w_{mn} = (n-m)w(m)^2 w(n)^2$. After some math we get

$$\hat{f} = \frac{\sum_{l=0}^{N-2} \eta_l(\hat{f}) \int_0^1 f(l+t)dt}{\sum_{l=0}^{N-2} \eta_l(\hat{f})} \tag{8}$$

where $\eta_l(g) = \sum_{n=l+1}^{N-1} \sum_{m=0}^{l} w_{mn} a(n)a(m) \mathrm{sinc} \frac{\Delta\varphi_{mn}(g)}{\pi}$, and the sinc function $\mathrm{sinc}(x) = \frac{\sin \pi x}{\pi x}$. The amplitude and phase estimates $\hat{a}$ and $\hat{\varphi}$ are given by

$$\hat{a}e^{j\hat{\varphi}} = \frac{e^{j\varphi_c} \sum_{n=0}^{N-1} w(n)^2 a(n) e^{j2\pi\left(\int_{N/2}^{n} f(t)dt - \left(n-\frac{N}{2}\right)\hat{f}\right)}}{\sum_{n=0}^{N-1} w(n)^2} \tag{9}$$

Eq. (8) implies that the frequency estimate is a weighted average of the instantaneous frequency over the frame. The weights depend on the window function, instantaneous amplitudes, and the instantaneous frequencies themselves. In particular, if the frequency remains constant, it equals the estimate regardless of the amplitude.

3.1. Pure Amplitude Change

The easiest case of parameter dynamics is amplitude change while the frequency stays constant. The signal can be written as

$$x(n) = a(n)e^{j(2\pi f n + \varphi)} \tag{10}$$

As stated above, the estimated frequency shall equal f. An immediate result is that the phase estimate is accurate as well. The amplitude estimate can be easily expressed using (9):

$$\hat{a} = \frac{\sum_n a(n)w(n)^2}{\sum_n w(n)^2} \tag{11}$$

The symmetry if (11) implies that only the even-symmetric part of $a(n)$, regarding the frame centre, contributes to the estimate. Pure amplitude change happens to stable-pitch sound sources such as piano or oboe. If a sinusoid is assumed to have a constant frequency, we can re-estimate the centre amplitude using (11).

3.2. Frequency and Amplitude Change

Instantaneous frequency change rarely happens without accompanying amplitude change. The general form of such a sinusoid is

$$x(n) = a(n)e^{j\left(\varphi_c + 2\pi \int_{N/2}^{n} f(t)dt\right)} \tag{12}$$

The frequency estimate is a weighted average of the instantaneous frequency during the frame. Calculation shows that when the amplitude is constant, the averaging weights η_l is symmetric and slightly more concentrated towards the frame centre than the square of the window function.

Like the constant frequency case, when the amplitude is constant during a frame, only the even part of $f(t)$ contributes to the frequency estimate. This implies that the frequency measurement is exact for a linear chirp. Considering amplitude change, a more general result is that the even part of $f(t)$ contributes to the frequency estimate when the even part of $a(n)$ is non-zero, and the odd part of $f(t)$ contributes to the frequency estimate when the odd part of $a(n)$ is non-zero.

3.3. Reestimation of Parameters

We formulate the error compensation task as follows: given $F+1$ measure points, say, $n_0, n_1, ..., n_F$, and $F+1$ parameter measurements $(\hat{a}_m, \hat{f}_m, \hat{\varphi}_m)_{m=0,1,...,F}$, find a series of parameters (a_m, f_m, φ_m), $m = 0, 1, .., F$, that generate the estimates through the analyzer.

Our key equation (8) involves $2N-1$ unknowns, *i.e.* $a(n)$, $0 \le n < N$, and $\int_0^1 f(l+t)dt$, $0 \le l < N-1$. This is too many to recover from the given estimates. We introduce parametric models $f(t, \Sigma_f)$ for frequency and $a(n, \Sigma_a)$ for amplitude variation, so that all the unknowns can be calculated from a sequence of $F+1$ parameter sets. Denote such a sequence $\mathbf{P} = \left\{ (a_m, f_m, \varphi_m)^P \,|_{m=0,...,F} \right\}$, and the corresponding variation model parameters $\Sigma(\mathbf{P})$. (8) and (9) relates a parameter set $\mathbf{P}$ to its estimate $\hat{\mathbf{P}}$ in the form of

$$\hat{\mathbf{P}} = \mathcal{P}\left(\Sigma(\mathbf{P})\right) \tag{13}$$

That is, given any $\mathbf{P}$, we can estimate $\Sigma(\mathbf{P}) = \{\Sigma_f(\mathbf{P}), \Sigma_a(\mathbf{P})\}$, then calculate $\hat{\mathbf{P}}$ by (8) and (9). The reestimation task is just the opposite: given $\hat{\mathbf{P}}$ we try to find the original $\mathbf{P}$. We rewrite (13) as

$$\mathbf{P} = \hat{\mathbf{P}} - \mathcal{P}\left(\Sigma(\mathbf{P})\right) + \mathbf{P} \tag{14}$$

A recurrent method for solving (13) is derived from (14) as follows:

- 0° Define distance function $D(\mathbf{P}_1, \mathbf{P}_2)$, convergence threshold ε, and the maximal number of iterations MAX; set $\mathbf{P}_0 = \hat{\mathbf{P}}$;
- 1° for $n = 1, 2, ...,$ MAX, do 2°–5°;
- 2° estimate the variation parameters $\Sigma\left(\mathbf{P}_{n-1}\right)$;
- 3° calculate $\hat{\mathbf{P}}_{n-1} = \mathcal{P}\left(\Sigma\left(\mathbf{P}_{n-1}\right)\right)$ using (8) and (9);
- 4° if $D(\hat{\mathbf{P}}, \hat{\mathbf{P}}_{n-1}) < \varepsilon$, output $\mathbf{P}_{n-1}$, return;
- 5° calculate $\mathbf{P}_n = \hat{\mathbf{P}} - \hat{\mathbf{P}}_{n-1} + \mathbf{P}_{n-1}$;
- 6° output $\mathbf{P}$ with non-convergence tag.

We ignore phase estimates during this stage to avoid phase wrapping problem. The phase angle can be reestimated as

$$\varphi = \hat{\varphi} - \arg \sum_{n=0}^{N-1} w(n)^2 a(n) e^{j2\pi \left(\int_n^{\frac{N}{2}} f(t)dt - \left(n - \frac{N}{2}\right)\hat{f} \right)} \quad (15)$$

which is an immediate result of (9).

3.4. Test Examples

We use cubic splines to model amplitude and frequency variation. A cubic spline is a piecewise trinomial with continuous 1st and 2nd derivatives. Accordingly, the phase angle function is a quartic polynomial.

Let $F+1$ be the total number of estimates (frames). Parameters are estimated at points $0, N/2, ..., NF/2$. The cubic spline fills the gaps between 0 and $NF/2$. However, the reestimation of the first point 0 requires half a frame before zero. In this case we extrapolate the spline half a frame beyond its effective interval. The same is done to the last frame.

The test signals are synthesized sinusoids for which we have the "true" instantaneous parameters at any point. The error between the true parameter set $\mathbf{P}$ and an estimate $\tilde{\mathbf{P}}$ is defined as

$$ERR(\tilde{\mathbf{P}}, \mathbf{P}) = \frac{1}{F+1} \sum_{m=0}^{F} \sum_{n=0}^{N-1} w(n)^2 \Big(\tilde{a}_m \cos(\tilde{\varphi}_m + 2\pi \tilde{f}_m (n - N/2)) - a_m \cos(\varphi_m + 2\pi f_m (n - N/2)) \Big)^2 \Big/ \frac{a_m^2}{2} \sum_{n=0}^{N-1} w(n)^2 \quad (16)$$

We run tests on three types of signals: exponential-decay amplitude with constant frequency, constant amplitude with sinusoid-modulated frequency, and exponential-decay linear chirp. The first two are simplified cases of real sounds, and the last is included to represent combined amplitude-frequency change. Reestimated results are compared with the original. Tests show that the error hardly depends on the absolute signal level, frequency or phase. In all the tests we set central frequencies of three concurrent sinusoids to 0.151, 0.251, 0.351 (the Nyquist frequency being 0.5), and phase angles to 0. 11 frames are extracted with a Hann window of size 1024. So $F = 10$. The maximal iteration count is set at 25.

1. Exponential amplitude The main variable in this test is the rate of amplitude decay, defined as $\lambda_a = \frac{a(0)}{a(N/2)}$, *i.e.* the amplitude drops by a factor of a per hop size. In the test λ_a varies between 1 and 4. The results are given in Figure 1(a), in which line ① is the error calculated for the LSE estimate, and line ② from the reestimates. Line ② lies below ① between $\lambda_a = 1$ and $\lambda_a = 3$, an improvement of 15–25 dB for most of the interval. When $\lambda_a > 3$ the cubic spline can no longer keep up with the global signal dynamics (*e.g.* 95 dB for $\lambda_a = 3$), and the reestimation fails. Line ③ gives the result we get after one iteration. It is shown that most improvement is achieved by the first iteration of up to 25.

2. Sinusoid-modulated frequency The frequency modulator has three parameters: amplitude a_M, frequency f_M and phase angle. We fix the modulator phase to 0 at time 0, f_M to 0.2 and 0.33 per frame, and vary a_M between 0 and 10 bins. The maximal frequency change rate is $2\pi a_M f_M$ bins per frame. Results are given in Figures 1(b) and 1(c). We see that the first iteration is normally enough when the modulation is small, but more are needed when the modulation is high.

3. Exponentially decreasing linear chirp In this example we have two variable parameters: the linear frequency change and exponential amplitude change rates. We measure the frequency change rate Δf in bins per frame, and the amplitude change decay rate λ_a as in the first example. We let Δf be 2, 4, 8, and vary λ_a between 1 and 4. Results are given in Figures 1(d)–1(f). Again we see that the reestimation fails for high decay rates. It is also shown that the more the frequency varies, the less the first iteration contributes to the total improvement.

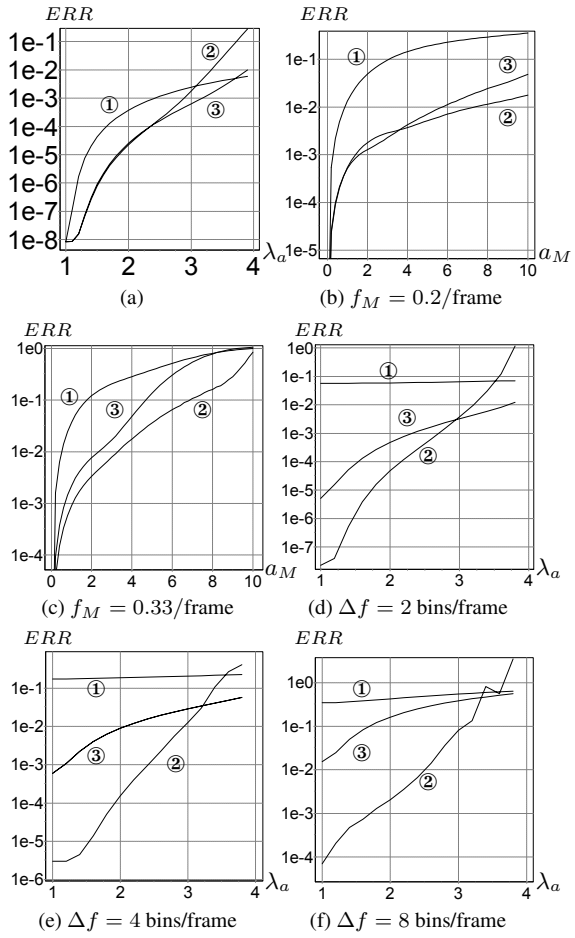

Figure 1: *Testing analyzer error. (a) exponential amplitude; (b)(c) modulated frequency; (d)(e)(f) exponential-decay linear chirp.*

4. RESYNTHESIS

We consider only the resynthesis *with phase*. As the amplitude and frequency variations are modeled with cubic splines, it is natural to use this model in resynthesis instead of the standard interpolation method in [1]. The sample-wise amplitudes can be derived from the cubic spline interpolation. For the phase angles, we can compensate for the model-to-measurement difference as follows.

Let the cubic spline frequency 0 and N be

$$f(t) = a\, t^3 + b\, t^2 + c\, t + d \quad (17)$$

and the phase estimate be $\varphi(0) = \varphi_0$, $\varphi(N) = \varphi_N$. The model phase function can be written as

$$\tilde{\varphi}_{(t)} = 2\pi \left(\frac{a}{4}\, t^4 + \frac{b}{3}\, t^3 + \frac{c}{2}\, t^2 + d\, t \right) + \varphi_0 \quad (18)$$

We write the final interpolation function φ as

$$\varphi(t) = \tilde{\varphi}(t) + \theta(t) \quad (19)$$

that satisfies

$$\begin{array}{l}\varphi(0)=\varphi_0, \varphi(N)=\varphi_N+2k\pi,\\ \varphi'(0)=\tilde{\varphi}'(0), \varphi'(N)=\tilde{\varphi}'(N)\end{array} \tag{20}$$

where $k \in \mathbb{Z}$ is chosen to minimize $\theta(t)$. We write (20) in term of $\theta(t)$:

$$\begin{array}{ll}\theta(0)=0, \theta(N)=\varphi_N-\tilde{\varphi}(N)+2k\pi,\\ \theta'(0)=0, \theta'(N)=0\end{array} \tag{21}$$

The four conditions of (21) suggests the use of a trinomial for $\theta(t)$, *i.e.* $\theta(t)=pt^3+qt^2+rt+s$. By solving (21) we get

$$\begin{array}{l}p=\frac{-2}{N^3}d(k), q=\frac{3}{N^2}d(k), r=s=0,\\ d(k)\equiv\varphi_N-\tilde{\varphi}(N)+2k\pi\end{array} \tag{22}$$

To minimize $\theta(t)$ we choose the integer k that minimizes $d(k)$, which is simply the integer nearest to $(\tilde{\varphi}(N)-\varphi_N)/2\pi$. This is very similar to the phase unwrapping process in [1], which can be regarded as a *linear spline* version.

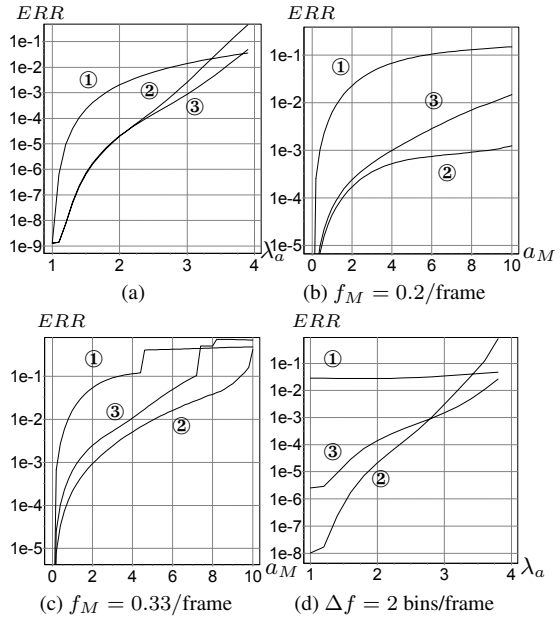

Figure 2: *Testing synthesizer error. (a) exponential amplitude; (b)(c) modulated frequency; (d) exponential-decay linear chirp.*

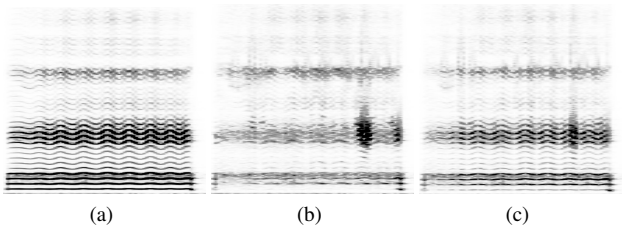

Figure 3: *Testing vocal vibrato. (a) original; (b) cubic spline modeling; (c) LSE-MQ modeling.*

4.1. Test Examples

In this section we compare the synthesizer error to the original one in [1] for signals used in section 3.4, plus a real recorded excerpt. For the synthesized sounds an extra result is given for using the [1] resynthesizer the true parameters to give some idea of pure synthesizer error. Errors are evaluated using equation (2).

1. Tests on synthesized signals The results are given in Figure 2(a)–2(d) for the first four test settings in Figures 1. The synthesizer errors show similar development trends to the analyzer errors in Figures 1. Both the absolute value and the measured improvements are slightly better than the analyzer error values, thanks to the use of interpolation.

2. Test on recording For this example we take a recording of soprano vibrato from the RWC database [6]. The spectrogram of the original is given in Figure 3(a). Figure 3(b) is the residue we get by subtracting the resynthesized result using cubic spline modeling, while Figure 3(c) is the result derived from the old system. The new system outperforms the old one except for the part where the residue shows some transient. Averaging over frames, the new modeling achieves 22 dB SNR, compared to 16 dB for the old one. We also see that the new residue is less sinusoidal, or more noise-like, than the old one.

5. CONCLUSION

In this article we have addressed the problem of estimating parameters from non-stationary sinusoids. Rather than directly using the results obtained using a stationary assumption, we propose to reestimate the parameters incorporating frequency and amplitude variation models. Although we have chosen an LSE estimator and a cubic spline model for our system, this reestimation framework is open to many other methods. For the resynthesizer part we have reformulated the phase unwrapping process in the context of cubic spline interpolation. Tests show the improvement in the accuracy of parameters, as well as in the resynthesized signals.

6. ACKNOWLEDGEMENTS

This work was supported by EU-FP6-IST-507142 project SIMAC (Semantic Interaction with Music Audio Contents) and Centre for Digital Music.

7. REFERENCES

[1] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 34, no. 4, pp. 744–754, 1986.

[2] F. Keiler and S. Marchand, "Survey on extraction of sinusoids in stationary sounds," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-02),* Hamburg, Germany, 2002, pp. 51–58.

[3] P. Depalle, G. Garcia, and X. Rodet, "Tracking of partials for additive sound synthesis using Hidden Markov Models," in *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc. (ICASSP'93),* Minneapolis, USA, 1993, pp. 225–228.

[4] M. Lagrange, S. Marchand, M. Raspaud, and J.-B. Rault, "Enhanced partial tracking using linear prediction," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-03),* London, UK, 2003, pp. 141–146.

[5] X. Serra and J. O. Smith, "Spectral modeling synthesis:a sound analysis/synthesis based on a deterministic plus stochastic decomposition," *Computer Music J.*, vol. 14, no. 4, pp. 14–24, 1990.

[6] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: Popular, classical, and jazz music databases," in *Proc. Int. Conf. Music Information Retrieval (ISMIR'02),* Paris, France, 2002, pp. 287–288.