ADDITIVE AND MULTIPLICATIVE REESTIMATION SCHEMES FOR THE SINUSOID MODELING OF AUDIO

Wen Xue and Mark Sandler

Centre for Digital Music, School of Electronic Engineering and Computer Science,
Queen Mary, University of London, London, United Kingdom
phone: +44 2078825528, fax: +44 2078827997,
email: {xue.wen;mark.sandler}@elec.qmul.ac.uk
web: www.elec.qmul.ac.uk/digitalmusic/

ABSTRACT

This paper discusses an approach to accurate sinusoid modeling of audio. We propose an iterative framework which functions as a "wrapper" that works with arbitrary sinusoid modeling systems to boost their accuracy. It involves one or more reestimation steps. In each step the parameter estimates are updated by combining a second set of parameters evaluated from the latest modeling error signal. An additive scheme and a multiplicative scheme are proposed for this reestimation step. On a limited test set the framework is shown to offer 2dB to 40.5dB (average 14.6dB) improvement in signal-to-residue ratio within 5 updates, which is 56.3% to 98.6% (average 79.5%) of the largest possible improvement, in dB, obtained by interpolating exact parameters.

1. INTRODUCTION

Sinusoid modeling uses slow-varying sinusoids to represent deterministic components in speech and audio [1][2]. A slow-varying complex sinusoid, say x, is defined as

$$x(t) = a(t)e^{j\varphi(t)},\ \varphi(t) = \varphi(0) + 2\pi\int_0^t f(\tau)d\tau \qquad (1)$$

where f, a and φ are the instantaneous frequency, amplitude and phase angle. In reverse, given the complex sinusoid x, the instantaneous frequency, amplitude and phase angle can be expressed as

$$f = \frac{\varphi'}{2\pi} = \mathrm{Im}\frac{x'}{2\pi x},\ a = |x| = e^{\mathrm{Re}\log x},\ \varphi = \mathrm{Arg}x = \mathrm{ImLog}x, \qquad (2)$$

where Arg is the phase angle unwrapped modular 2π, and Log is the complex logarithm function unwrapped modular $j2\pi$. Parameters of a complex sinusoid is unique up to a phase wrap of $2k\pi$, $k\in$Z. If both f and a are slow-varying, then x is said to be slow-varying. In practice we always deal with real-valued sinusoids. Unlike complex ones, real time-varying sinusoids generally do not have unique parameter sets [3]. A real sinusoid is said to be slow-varying if any of its parameter sets is.

A complete sinusoid modeling system includes an analyzer and a synthesizer. The analyzer finds sinusoidal components and evaluates their instantaneous frequencies, amplitudes and phase angles from the waveform at sparsely distributed measurement points. The synthesizer rebuilds these components as time-varying sinusoids by interpolating the estimated parameters. Figure 1 gives a brief outline of a sinusoid modeling system, in which the single lines carry waveform representation, and the double line carries sinusoid representation. A module that converts between these two representations is either an analyzer (marked by "A") or a synthesizer (marked by "S"), depending on the direction of conversion.

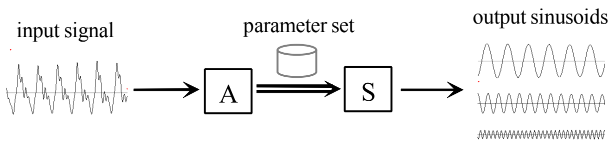

Figure 1. Sinusoid modeling data flow

Sinusoid modeling is most useful when each sinusoid accurately appoximates a deterministic component, known as a partial, in the analyzed audio, so that we may remove a partial by direct subtraction and obtaining a clean residue. In practice it is often difficult, and a phantom partial often remains in the residue, which preserves much perceptual characteristics of the removed sinusoid. We attribute this to three types of errors: model error, analysis error, and synthesis error. Model error occurs if the target partial can not be represented as a slow-varying sinusoid, such as in a vibrato with heavy reverb. Analysis error occurs if the parameters are not accurately estimated, and synthesis error occurs if the amplitude and frequency interpolation laws used in the synthesizer do not coincide with the true ones. In this paper we focus on the analysis error of real slow-varying sinusoids. Due to the non-uniqueness of parameters of real time-varying sinusoids, it is difficult to separate analysis and synthesis errors in any strict sense. We always address the total error of the two stages, which can be measured by comparing two waveforms, and which we try to reduce without altering the synthesis settings. Analysis and synthesis errors are most prominent for time-varying sinusoids (also known as non-stationary sinusoids), mostly due to the lack of a priori knowledge on their parameter variations.

Many efforts have been reported for the accurate modeling of sinusoids. On the analyzer part, methods have been proposed for the accurate estimation of parameters. Early methods [1][2][4] assume short-time stationarity and estimate parameters without considering parameter variations. Later methods take parameter variation into consideration by assuming certain parametric laws: linear frequency and amplitude law is used in [5]-[7], linear frequency and log amplitude law is used in [8][9], while [10] suggests using a dictionary from which a best-matching law can be selected. These methods easily fail if the presumed laws do not coincide with those of the signal. A reestimation-based method has been proposed to cope with arbitrary parameter variations [11], which also lays the foundation of the multiplicative reestimation scheme to be discussed in this paper. Improvements on the synthesizer have been limited, partially due to the fact that synthesis errors are comparatively less prominent, and can be easily controlled by using smaller hops between measurement points. Several interpolation models smoother than the original MQ synthesizer [1] have been developed, including polynomial interpolation with pre-evaluated derivatives [12] and cubic splines [11].

In this paper we generalize the method in [11] into an iterative reestimation framework, which updates the parameter estimates to compensate for modeling errors. This framework neither presumes any specific parameter variation law nor relies on any specific analyzer or synthesizer, but wraps up arbitrary analyzer/synthesizer to boost modeling accuracy of sinusoids with arbitrary parameter variations. In the following section we propose two reestimation schemes, an additive scheme and a multiplicative one, within this framework. The effectiveness of the reestimation framework will be tested in section 3.

2. THE REESTIMATION METHOD

The common framework of the two reestimation schemes is illustrated in Figure 2, where single and double lines carry waveform and sinusoid representations respectively. Let x be the original signal and x_1 be its estimate with parameter set $(a_1,\varphi_1,\varphi_1')$. We compare x and x_1 to obtain an error signal y_1, which captures the information of x that is lost in x_1. We further evaluate sinusoid parameters (b_1,θ_1,θ_1') from y_1 and combine it with $(a_1,\varphi_1,\varphi_1')$ into $(a_2,\varphi_2,\varphi_2')$, from which a sinusoid x_2 can be constructed. If (b_1,θ_1,θ_1') is a good model of y_1, then a large part of the information of x lost in x_1 can be restored in x_2.

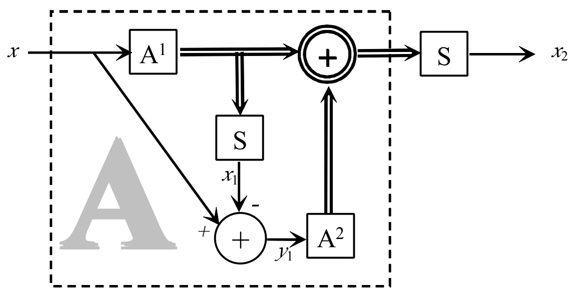

Figure 2. The reestimation framework

Apart from analyzer and synthesizer modules that already exist in any sinusoid modeling system, Figure 2 also contains a comparison module that calculates the error signal y_1 by comparing two waveforms, and a combination module that combines two sinusoid parameter sets into one. The latter is the exact inverse of the former in sinusoid domain, so that if no error is incurred at A^2 then the combined parameter set is accurate for x. In other words, the analysis error in x_2 depends solely on how well y_1 is estimated, regardless of x_1. Each pair of comparison and combination modules make up a reestimation scheme within the framework. Two conditions must be met for a reestimation scheme to be useful: that y_1 be a slow-varying sinusoid itself, and that y_1 be more accurately estimated than x has been. In this paper we discuss the additive and multiplicative schemes, named after the operation taken in the combination module.

2.1 Additive reestimation

The following proposition states that if x and y are complex sinusoids with slow-varying amplitudes and their frequencies are close, then $z=x+y$ is a complex sinusoid with a slow-varying amplitude and a frequency close to theirs, unless x and y cancel each other when added. In addition, if x and y have smooth amplitude variations and slow frequency variations, so has z.

Proposition 1. Let $x=ae^{j\varphi}$ and $y=be^{j\theta}$ be complex sinusoids, a, b, φ, $\theta\in$R, $\frac{|a'|}{a},\frac{|b'|}{b}\leq\varepsilon_1$, $|\varphi'-\theta'|<2\varepsilon_2$, $\frac{|a''|}{a},\frac{|b''|}{b}\leq\varepsilon_3$, $|\varphi''|,|\theta''|<\varepsilon_4$, and let $z=ce^{j\psi}=x+y$, c, $\psi\in$R, then

$$\frac{|c'|}{c}\leq\beta(\varepsilon_1+\varepsilon_2),\ |\psi'-\varphi'|+|\psi'-\theta'|\leq 2\beta(\varepsilon_1+\varepsilon_2), \tag{3a}$$

$$\frac{|c''|}{c}\leq\beta(\varepsilon_3+\varepsilon_4)+\beta^2(\varepsilon_1^2+\varepsilon_2^2+0.5\varepsilon_1(\varphi'+\theta')+\varepsilon_1\max(\varphi',\theta')) \tag{3b}$$

$$|\psi''|\leq\beta(\varepsilon_3+\varepsilon_4)+\beta^2(\varepsilon_1^2+\varepsilon_2\max(\varphi',\theta')+\varepsilon_1(\varphi'+\theta')) \tag{3c}$$

where

$$\beta=\frac{|x|^2+|y|^2+2|\mathrm{Re}\,x^*y|}{|x|^2+|y|^2+2\mathrm{Re}\,x^*y}=1+2\frac{|\mathrm{Re}\,x^*y|-\mathrm{Re}\,x^*y}{|x|^2+|y|^2+2\mathrm{Re}\,x^*y}. \tag{4}$$

This is proved by substituting the following

$$x'=x(\frac{a'}{a}+j\varphi'),\ x''=x(\frac{a''}{a}-(\varphi')^2+j\varphi''+j\frac{a'}{a}\varphi'), \tag{5a}$$

$$y'=y(\frac{b'}{b}+j\theta'),\ y''=y(\frac{b''}{b}-(\theta')^2+j\theta''+j\frac{b'}{b}\theta') \tag{5b}$$

into

$$\frac{c'}{c}=\mathrm{Re}\frac{z'}{z},\ \psi'=\mathrm{Im}\frac{z'}{z}, \tag{5c}$$

$$\frac{c''}{c}=\mathrm{Re}\frac{z''}{z}+(\mathrm{Im}\frac{z'}{z})^2,\ \psi''=\mathrm{Im}\frac{z''}{z}-\mathrm{Re}\frac{z'}{z}\mathrm{Im}\frac{z'}{z}. \tag{5d}$$

The value of β remains at 1 when Re(x^*y) is positive, i.e. x and y are *in phase*. When Re(x^*y)<0, β grows above 1 and is upper-bounded by $1+4\alpha(1-\alpha)^{-2}$, where α is the ratio of a and b. $\beta\to\infty$ as $y\to -x$, indicating that if $-y$ is a good approximation of x, then z tends to be noise-like.

Based on Proposition 1, we propose an additive reestimation scheme in which the error signal y_1 is taken as the modeling residue of (a,φ,φ'), i.e.

$$y_1 = x - x_1. \quad (6)$$

If x_1 does not accurately model x, then y_1 must be a slow-varying sinusoid. The combination module combines the models by

$$a_2 e^{j\varphi_2} = a_1 e^{j\varphi_1} + b_1 e^{j\theta_1}, \quad (7a)$$

$$\varphi_2' = \frac{a_1\varphi_1'(a_1 + b_1\cos\delta) + b_1\theta_1'(b_1 + a_1\cos\delta) + (a_1'b_1 - a_1b_1')\sin\delta}{a_1^2 + b_1^2 + 2a_1b_1\cos\delta} \quad (7b)$$

where $\delta = \varphi_1 - \theta_1$. (7b) is derived during the proof of (3b). Apart from the standard parameters, to use (7b) we need amplitude derivatives of x_1, which can be obtained from the synthesis module.

The concept of additive reestimation has previously been explored in [13], where two sinusoids are "fused" without considering parameter variations. The additive scheme improves modeling accuracy if the modeling residue of y_1 is smaller than y_1 itself, which is usually true if y_1 is a slow-varying sinusoid.

2.2 Multiplicative reestimation

The following proposition states that if x and y are complex sinusoids with slow-varying amplitudes and frequencies, then so is $z=xy$.

Proposition 2. Let $x=ae^{j\varphi}$ and $y=be^{j\theta}$ be complex sinusoids, a, b, φ, $\theta\in$R, $\frac{|a'|}{a}, \frac{|b'|}{b} \le \varepsilon_1$, $|\varphi''|, |\theta''| < \varepsilon_2$, and let $z=ce^{j\psi}=xy$, c, $\psi\in$R, then

$$\frac{|c'|}{c} \le 2\varepsilon_1, \ \psi'' \le 2\varepsilon_2, \quad (8)$$

This is proved by differentiating $c=ab$ and $\psi=\varphi+\theta$.

Based on Proposition 2, we propose a multiplicative reestimation scheme in which the error signal y_1 is taken as complex ratio of x and x_1, i.e.

$$y_1 = \frac{x}{x_1} \quad (9)$$

The combination module combines the models by

$$a_2 = a_1 b_1, \ \varphi_2 = \varphi_1 + \theta_1, \ \varphi_2' = \varphi_1' + \theta_1' \quad (10)$$

The multiplicative scheme was previously developed in the context of demodulation [11]. It improves sinusoid modeling if parameters of y_1 are more accurately evaluated than those of x_1. This is usually true if $(a_1, \varphi_1, \varphi_1')$ has good or mediocre accuracy, so that the division by x_1 in (9) removes a large part of parameter dynamics from x.

2.3 Cascade of reestimation updates

Comparing Figures 1 and 2, it is obvious that the components within the dashed box in Figure 2 compose an analyzer module. By using this analyzer in the place of A^1, (or equivalently, using x_2 in the place of x_1,) we are able to repeat the reestimation process to further improve accuracy. This cascading of the reestimation framework leads to an iterative procedure. The choice of analyzer, synthesizer and reestimation scheme is free to vary from one iterate to the next. The process stops either after a preset number of iterates, or when no substantial accuracy improvement is gained in the latest iterate.

3. TESTS

3.1 Test set

We run tests on three groups of synthesized sinusoids, including linear chirps, amplitude modulated, and amplitude-and-frequency modulated sinusoids. All samples are 8192 points long. A fixed window size 1024 and frame hop 512 is used, giving 15 basic measure points (15 frames) per sample. The sinusoids have a base amplitude of 1 and are quantized to the precision of 2^{-10}.

Group 1 contains 120 samples, with 20 central frequencies uniformly sampled from 255.00bin to 255.95bin (1bin=1/1024), combined with 6 frequency slopes $2f_1$ at 0, 0.25, 0.5, 1, 2, 4 bins per frame (i.e. per 512 points). Results are given as functions of f_1, averaged over f_0.

Group 2 contains 220 samples, with the same 20 f_0's as above, 6 modulation depths A_M from 0.15 to 0.9 with modulation period T_M fixed at 5 frames, and 6 T_M's from 5 to 15 frames with A_M fixed at 0.9. The modulation phases φ_M are selected at random. Results are given as functions of A_M and T_M in two separate tests, averaged over f_0.

Group 3 contains 220 samples, with the same 20 f_0's as above, 6 frequency modulation extents A_M from 1 bin to 32 bins with modulation period T_M fixed at 5 frames, and 6 T_M's from 5 to 15 frames with A_M fixed at 8 bins. The modulation phases φ_M are selected at random. The amplitudes are taken as quadratic functions of frequency so that the peak frequency has twice the amplitude as f_0. Results are given as functions of A_M and T_M in two separate tests, averaged over f_0.

The test set is summarized in Table 1, where all frequency parameters are in bins, and T_M is in frames.

Group	Control variables	Parameters (n=-4096, …, 4095, N=1024)
1	f_0, f_1	$a_n = 1$, $f_n = f_0 + 2nf_1/N$, $\varphi_n = 2\pi n(f_0 + nf_1/N)/N$
2	f_0, A_M, T_M, φ_M	$a_n = 1 + A_M\cos(\varphi_M + 4\pi n/NT_M)$, $f_n = f_0$, $\varphi_n = 2\pi n f_0/N$
3	f_0, A_M, T_M, φ_M	$a_n = 1 + (f_n - f_0)^2/A_M^2$, $f_n = f_0 + A_M\cos(\varphi_M + 4\pi n/NT_M)$, $\varphi_n = \frac{2\pi}{N} f_0 n + \frac{T_M A_M}{2}\sin(\varphi_M + 4\pi n/NT_M)$

Table 1. Test set

3.2 Test settings

We use signal-to-residue ratio (SRR) for evaluating sinusoid modeling accuracy. Least square (LS) [3] and reassignment (RA) [7] estimators are tested within the framework as basic analyzers. LS is engaged as a specimen of estimators that do not consider parameter variation. RA is engaged as one of the start-of-the-art estimators which explicitly address the non-stationarity issue. In this test RA is implemented in an enhanced version following [11]. Cubic spline interpolator is used as the basic synthesizer.

For each basic analyzer four system settings are tested, including the basic system alone, basic system with one additive update (A), basic system with one multiplicative update (M), and basic system with up to 5 better-of-the-two updates (I5). In the last setting both schemes are tried at every iterate, but only the better reestimate is preserved to start the next iterate, until neither is better than the original estimate, or 5 iterates are finished.

Apart from the above eight settings, we also construct a "reference" sinusoid from the true parameters and measure its SRR. Despite the non-uniqueness issue, this SRR provides a good reference on the best possible accuracy given the interpolator. The gap between the basic SRR and this reference SRR measures how much space is available for improvement. We call it *improvement room* for convenience.

3.3 Results

Figure 3 shows the results tested on linear chirps, depicted as SRR (y axis) against chirp rate (x axis) for LS in all four settings. We also include the result of RA for comparison. For linear chirps RA in the enhanced implementation provides accurate results for all parameters. In Figure 1 the SRR of RA is capped at about 68dB by the limited input precision. LS also provides accurate frequency estimates for linear chirps, but is unable to estimate amplitude and phase angle accurately under its native stationarity assumption. However, Figure 1 shows that by wrapping LS in the iterative framework we can achieve the accuracy almost as good as that of RA within 5 iterates. We have further counted that the average number of iterates used is 3.45, out of which 1.45 (42%) are additive updates and 2 are multiplicative. Of all the first iterates 12.5% are additive, 83.3% are multiplicative, while the rest 4.2% do not see improvement in overall accuracy by either scheme. These observations agree with the direct comparison in Figure 1, which shows the multiplicative scheme more effective than the additive scheme. RA does not benefit from the reestimation framework, as the basic RA analyzer is already perfect for linear chirps. The reference SRR result is not drawn in Figure 3, as it overlaps that of RA. In this test LS-I5 offers an average of 40.5dB improvement in SRR over LS, covering 98.6% of the improvement room.

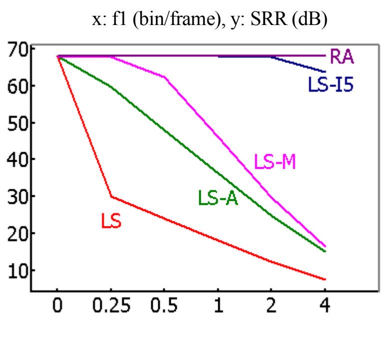

Figure 3. Results for linear chirps

Results for amplitude modulated sinusoids are given in Figure 4. In the upper half we depict SRR against modulation depth A_M, while the modulation T_M period is fixed at 5 frames; in the lower half we depict SRR against modulation period while the modulation depth is fixed at 0.9. Results of LS in four settings are found in the left column, of RA are found in the right column. The reference SRR is included as dotted lines.

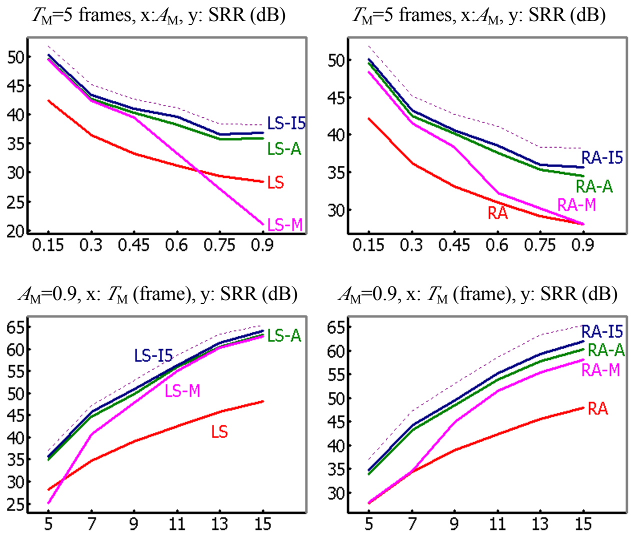

Figure 4. Results for AM sinusoids

For these signals LS and RA are similar in accuracy and behaviour. The effect of reestimation is clearly visible by comparing the distance of these curves from the dotted reference line. Unlike for linear chirps, for amplitude modulated sinusoids additive reestimation is shown to be the more effective of the two schemes, taking up 79.6% of all first iterates. In all settings most improvements are obtained from the first update. LS-I5 offers an average of 10.2dB improvement in SRR, covering 84.9% of the improvement room. With RA the two numbers are 9.4dB and 77.1%.

Results for frequency modulated sinusoids with accompanying amplitude modulation are given in Figure 5. In the upper half we depict SRR against frequency modulation extent A_M, while the modulation T_M period is fixed at 5 frames; in the lower half we depict SRR against modulation period while the frequency modulation extent is fixed at 8 bins. The left column compares results of LS in four settings together with basic RA, the right column compares results of RA in four settings together with LS-I5. The reference SRR is included as dotted lines.

For these signals basic RA consistently outperforms basic LS, but is easily beaten by LS with no more than 5 updates. The reestimation schemes improve the accuracy of both basic analyzers alike. At the end of 5 updates LS and RA show very similar results. Improvements appear to be more gradual with varying frequencies, therefore better accuracy may be expected beyond 5 iterates. Of the two reestimation schemes the multiplicative is shown to be the more effective, taking up 81.9% of all first iterates, while the additive scheme does better for small frequency modulation extents. This can be explained by Proposition 1: the larger the frequency variation, the larger difference to be expected between true and estimated frequencies, thus the less likely the residue be a slow-varying sinusoid. On the average LS-I5 offers 17.8dB improvement in SRR, covering 79.5% of the improvement room. With RA these two numbers are 8.1dB and 68.1%.

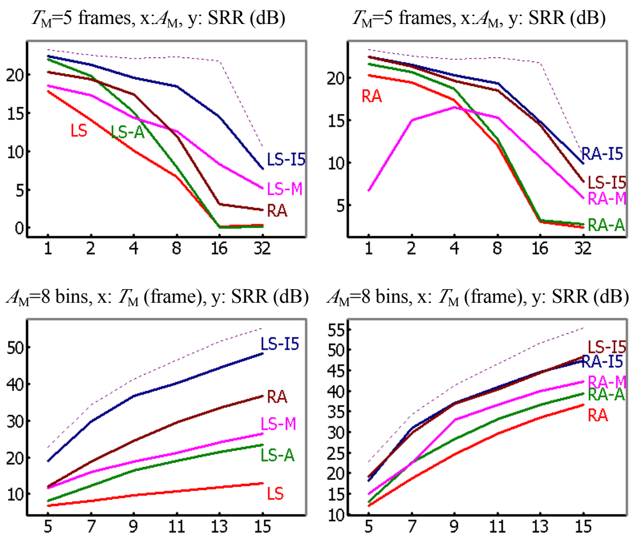

Figure 5. Results for AM-FM sinusoids

4. CONCLUSION

In this paper we have proposed a reestimation framework for improving the accuracy of sinusoid modeling. By updating parameter estimates incorporating information embedded in the modeling error, we achieve higher accuracy without expanding the data structure. The framework does not depend on any specific signal model, analyzer or synthesizer, but works with any analyzer-synthesizer combo to provide accuracy boost for arbitrary slow-varying sinusoids. It is flexible, robust and very easy to implement. By maintaining the data structure they also preserve component integrity, avoiding expressing a single sinusoid in multiple parts, which is common in decomposition-based approximation techniques such as wavelets and pursuits.

Within the framework we have implemeted an additive and a multiplicative scheme for calculating the error signal and updating parameters. Each scheme has its own advantage and we have shown that they can be easily combined to make up for each's disadvantage. The framework, moreover, is open to further exploration in search of more powerful reestimation schemes.

5. ACKNOWLEDGEMENT

This work was supported by the EPSRC EP/E017614/1 project OMRAS2 (Online Music Recognition and Searching).

6. REFERENCES

[1] R.J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Tran. Acous. Sp. Sig. Proc.*, vol.34, no.4, 1986, pp. 744-754.

[2] X. Serra, "Musical sound modeling with sinusoids plus noise," G. D. Poli, A Picialli, S. T. Pope, and C. Roads ed., *Musical signal processing*, Swets & Zeitlinger Publishers. 1997.

[3] Wen X., *Harmonic sinusoid modeling of tonal music events*, PhD Thesis, University of London, 2007.

[4] F. Keiler and S. Marchand, "Survey on extraction of sinusoids in stationary sounds," in *Proc. DAFx'02*, Hamburg, 2002.

[5] P. Masri and N. Canagarajah, "Extracting more detail from the spectrum with phase distortion analysis," in *Proc. DAFx'98*, Barcelona, 1998.

[6] G. Peeters and X. Rodet, "SINOLA: a new analysis/synthesis method using spectrum peak shape distortion, phase and reassigned spectrum," in *Proc. ICMC'99*, Beijing, 1999.

[7] A. Röbel, "Estimation of partial parameters for non stationary sinusoids," in *Proc. ICMC'-06*, New Orleans, 2006.

[8] M. Abe and J. O. Smith III, "AM/FM rate estimation for time-varying sinusoidal modeling," in *Proc. ICASSP'05*, Philadelphia, 2005.

[9] S. Marchand and P. Depalle, "Generalization of the derivative analysis method to non-stationary sinusoidal modeling," in *Proc. DAFx'08*, Espoo, 2008.

[10] H. K. Kwok and D. L. Jones, "Improved instantaneous frequency estimation using an adaptive short-time Fourier transform," *IEEE Tran. Sig. Proc.*, vol.48 no.10, 2000.

[11] Wen X. and M. Sandler, "Evaluating parameters of time-varying sinusoids by demodulation," in *Proc. DAFx'08*, Espoo, 2008.

[12] L. Girin, S. Marchand, J. Martino, A. Röbel and G. Peeters, "Comparing the order of a polynomial phase model for the synthesis of quasi-harmonic audio signals," in *Proc. WASPAA'03*, New Paltz, 2003.

[13] T. Virtanen, "Accurate sinusoidal model analysis and parameter reduction by fusion of components," in *Proc. AES 110th Convention*, Amsterdam, 2001.