EVALUATING PARAMETERS OF TIME-VARYING SINUSOIDS BY DEMODULATION

Wen Xue and Mark Sandler

Centre for Digital Music, Dept. of Electronic Engineering,
Queen Mary, University of London
London, United Kingdom
{xue.wen;mark.sandler}@elec.qmul.ac.uk

ABSTRACT

In this paper we propose a method for reestimating the instantaneous parameters of time-varying sinusoids based on demodulation. The method uses rough primitive parameter estimates to construct a sinusoidal multiplier. By applying this multiplier to the original sinusoid we get a third sinusoid with slower-varying parameters than the latter. In this way the analysis of a fast-varying sinusoid is reduced to that of a slow-varying one, whose parameters can be more reliably evaluated. We also propose a front-end reestimator using variable window sizes to provide inputs to the demodulation-based reestimator, so that the primitive estimates is reliable even for fast-varying parts. This reestimation method is non-parametric, stable, easy to implement, and do not require the use of any specific estimator. Its effectiveness is shown by results on various test sets.

1. INTRODUCTION

This paper discusses parameter estimation of time-varying sinusoids for sinusoid modeling of audio [1], [2]. In particular, we focus on the correction of estimating errors due to the variation of parameters within the duration over which they are evaluated. While sinusoid models generally allow arbitrary parameter variations as long as they are "slow", most current estimators assume stationarity in parameters [1]-[3] or parametric parameter variation models [4]-[8]. Estimators assuming stationarity have good performance if the variations are very slow, but easily fail when they become moderately fast. Reassignment-based methods [9]-[10], taking advantage of the accurate localization of linear chirps by time-frequency reassignments, are able to provide accurate estimates for these chirps, but not for other signals. More complicated parametric models have been explored in [4]-[8], which still lack the flexibility to generalize. Large errors are likely to occur when the signal behaviour departs from the estimator assumptions.

However, most of these estimators are designed to give accurate results for constant sinusoids. Accordingly, as long as the parameter variations remain very slow, relatively high accuracy may be expected even if the assumptions are not fully satisfied. Therefore if we find for a fast-varying sinusoid x a slow-varying equivalent y, so that there exists a known bijection between their parameters, then we can improve the estimation by applying the estimator to the slow-varying y instead of to the fast-varying x. In the following we propose a demodulation method based on this idea. The method requires the parameters be roughly known for multiple frames, therefore it functions as a reestimation scheme to refine primitive results in a post-tracking stage. We also propose a window size selection method as a front-end, which directly evaluates parameter dynamics from the primitive results, and choose a "good" window size accordingly. These methods do not require the use of any specific estimator, but act as "boosters" to work with arbitrary estimators.

There are fundamental connections between these methods and an adaptive STFT scheme proposed in [11], wherein the authors choose a window size and multiplier that "matches the signal" from a pre-defined library, in a similar manner as the matching pursuit [12]. Accordingly, its use is limited by the features of pursuits, such as the trade-off between adaptation capacity and limited library size.[1] Our methods calculate the multiplier and window size directly from primitive estimates.

In this paper we use the FFT-based least-square-error (LSE) estimator [13] to work with the reestimation schemes. LSE estimator gives accurate frequency estimation for linear chirps. Its frequency estimation error for other sinusoids can be roughly outlined by [14, §3.4.2]

$$\hat{f} - f_0 \cong \alpha D \tau^2\ ,\ D = \kappa f_2 + a_1 f_1\ , \tag{1}$$

where f_0 is the frequency, $\hat{f}$ its estimate, f_1 and $2f_2$ its 1st- and 2nd-order derivatives, a_1 the normalized amplitude derivative, τ the window size, and α and κ are positive values that depend on the window type only. (1) clearly claims that the frequency error increases with parameter dynamics and window size, on which we address respectively in the next two sections.

2. PARAMETER DEMODULATION

Let x be a time-varying sinusoid,

$$x_n = a_n e^{j\varphi_n}\ , \tag{2}$$

and $\tilde{a}_n$ and $2\pi\tilde{f}(t)$ be two sequences that approximates a_n and $d\varphi/dt$, respectively. We construct a second time-varying sinusoid

$$y_n = \frac{x_n}{\tilde{a}_n} e^{-j2\pi\int_0^n \tilde{f}(t)dt} \tag{3a}$$

Let the parameter estimates of y be $\hat{f}^y$, $\hat{a}^y$ and $\hat{\varphi}^y$, then the parameters of x can be estimated using

$$\hat{f}^x = \tilde{f} + \hat{f}^y\ ,\ \hat{a}^x = \tilde{a}\cdot\hat{a}^y\ ,\ \hat{\varphi}^x = \hat{\varphi}^y + 2\pi\int\tilde{f}dt\ . \tag{3b}$$

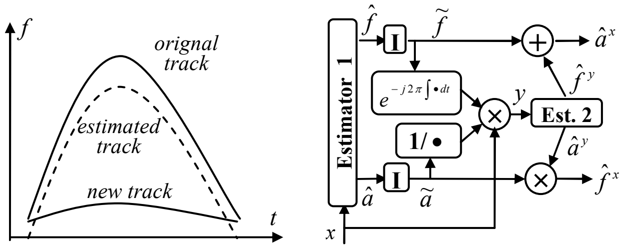
We call the above the demodulation method, since the multiplier

[1] Actually in [11] linear chirps are finally chosen to match all signals, in which case it is no more advantages than the parametric methods involving chirps.

$e^{-j2\pi\int_0^n \tilde{f}(t)dt}$ removes part of the frequency modulation, and $1/\tilde{a}_n$ part of the amplitude modulation, from sinusoid x. Compared to x, y is a sinusoid with less frequency and amplitude variations, and according to (1), can be more reliably evaluated.

$\tilde{a}$ and $\tilde{f}$ can be constructed from primitive estimates obtained with an estimator. If the estimator fits in a parametric model, such as in [4][6], the model itself can provide $\tilde{a}$ and $\tilde{f}$. If not, we interpolate the primitive estimates between frames. Following [13], in this paper a cubic spline interpolation is used.

Figure 1a illustrates the frequency demodulation, with the true frequency track in a solid line on top, and its estimate in dashed line. The demodulation step subtracts the dashed track from the solid one to give the nearly flat track at the bottom, which has a much slower variation. Figure 1b gives the block diagram of the demodulation-based reestimator, where the blocks marked "I" stand for interpolators.



(a) removing frequency dynamics (b) block diagram

Figure 1: *Demodulation method*

New multipliers can be constructed from the reestimates to further remove parameter dynamics in y, leading to an iterative procedure. Frequency and amplitude demodulation can be applied separately. We put more favour in frequency demodulation by skipping amplitude demodulation at some iterates. The iteration stops when the dynamics measure D (see (1)) fails to decrease, or a maximal number of iterates is met.

3. VARIABLE-WINDOW-SIZE METHOD

A pre-requisite of the demodulation method is that the primitive estimates approximate the true parameters, which is usually satisfied for small and moderate parameter dynamics. In case of large dynamics, errors in the primitive estimates are often so large that no valid multiplier can be found, causing the demodulation method to fail, as shown by test results in section 4.

However, (1) suggests that the error due to large dynamics, measured by D, can be compensated by using shorter windows. Moreover, (1) provides $D\tau^2$ as an indicator of frequency error, and a way to directly estimate D by parabolic fitting, so that the window size can be chosen by limiting $D\tau^2$ under a threshold Th.

In numerical computation it is convenient to let all window sizes be powers of 2. We look for primitive estimates that are calculated from frames with large dynamics, i.e. $D\tau^2>Th$, and reestimate them using shorter window sizes chosen as $2^{-k}\cdot 2\tau$, so that $D\cdot 2^{-2k}\tau^2\leq Th$. Considering harmonics and noise, it is not practical to let τ grow arbitrarily small. We always set a minimal allowed window size for this method.

Tests show that the variable-window-size method greatly improves the estimates for large parameter dynamics, but has little effect for small ones. Since the demodulation method is effective for small dynamics but not for large ones, it is natural to combine the two to build a reestimator that is robust for both large and small dynamics. This is implemented by concatenating the two reestimators directly: the variable-window-size method being used as a front-end to the demodulation method to provide for the latter refined primitive estimates.

4. TESTS

We run tests on four groups of synthesized signals, including linear chirps, amplitude modulated, frequency modulated, and amplitude-and-frequency modulated sinusoids. All samples are 16384 points long. A fixed window size 1024 and hop size of 512 are used (except in the variable-window-size method, where window sizes 1024, 512, 256 and 128 are allowed). This gives 31 parameter sets per sample. As this paper focuses on errors due to parameter variations, harmonics and noise are not considered in the test. However, since the demodulation step changes neither the noise level nor the frequency interval between partials, its effectiveness is unlikely to be sensitive to noise or harmonics.

Group 1 contains 160 samples, with 20 central frequencies f_0 from 255.00bin to 255.95bin (1bin=1/1024), combined with 8 frequency slopes f_1 ay 0, 0.125, 0.25, 0.5, 1, 2, 4, 8 bins per frame (i.e. per 512 points). Results are given as functions of f_1, averaged over f_0.

Group 2 contains 1800 samples, with 5 f_0's from 99.00bin to 99.80bin, 6 modulation depths A_M from 0.15 to 0.9, 6 modulation periods T_M from 2 to 12 frames, and 10 modulating phases φ_M from 0 to 0.45π. Results are given as functions of A_M and T_M, averaged over f_0 and φ_M.

Group 3 contains 1800 samples, with the same 5 f_0's, 6 modulation amplitudes A_M from 1bin to 32bin, the same 6 T_M's and 10 φ_M's. Results are given as functions of A_M and T_M, averaged over f_0 and φ_M.

Group 4 contains 1800 samples, with the frequencies arranged the same way as in Group 3, and the amplitudes 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, averaged over f_0 and φ_M.

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

Table 1: *Test set*

Group	Control variables	Parameters (n=-8192, …, 8191, 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	$a_n=1+A_M\cos(\varphi_M+4\pi n/NT_M)$, $f_n=f_0$, $\varphi_n=2\pi nf_0/N$
3	f_0, A_M, T_M, φ_M	$a_n=1$, $f_n=f_0+A_M\cos(\varphi_M+4\pi n/NT_M)$, $\varphi_n=\frac{2\pi}{N}f_0n+\frac{T_MA_M}{2}\sin(\varphi_M+4\pi n/NT_M)$
4	f_0, A_M, T_M, φ_M	$a_n=1+(f_n-f_0)^2/A_M^2$, $f_n=f_0+A_M\cos(\varphi_M+4\pi n/NT_M)$, $\varphi_n=\frac{2\pi}{N}f_0n+\frac{T_MA_M}{2}\sin(\varphi_M+4\pi n/NT_M)$

The performance of parameter estimation is evaluated using a synthetic approach. Two constant sinusoids are synthesized from the true and estimated parameter sets respectively, then the square error between the two is compared to the former to produce a noise-to-signal ratio (NSR). One NSR is calculated per frame. The NSR of a test sample is obtained by averaging the NSR's calculated from all its frames. The final results are given as SNR's (SNR=1/NSR), in dB, so as to make them positive.

The LSE method is used as the estimator module to work with the reestimation schemes (i.e. estimators I and II in Figure 2). The plain LSE estimator (LSE), reassignment estimator (RA) [10] and an enhanced Abe-Smith (AS) estimator [6] are tested along with the proposed reestimators, i.e. LSE with demodulation (DV), with variable window size (VW) and their direct concatenation (VWDV). The plain LSE estimator calculates the parameters as in [13] without considering parameter dynamics. It is tested as the baseline estimator to demonstrate the improvements brought by the reestimation schemes.

The reassignment estimator calculates a pair of time and frequency estimates as

$$\hat{t} = -\mathrm{Im}\frac{\partial X_k/\partial\omega}{X_k},\ \hat{f} = \frac{k}{N} + \mathrm{Im}\frac{\partial X_k/\partial t}{2\pi X_k} \tag{4a}$$

where X_k is the spectral peak at the sinusoid and

$$\frac{\partial X_k}{\partial\omega} = -j\sum_n n x_n w_{-n} e^{-j\frac{2\pi kn}{N}},\ \frac{\partial X_k}{\partial t} = \sum_n x_n w'_{-n} e^{-j\frac{2\pi kn}{N}}. \tag{4b}$$

w' is the derivative of the window function w. The reassignment method does not provide a straightforward way for evaluating the amplitude. However, it is possible to estimate the frequency derivative of a linear chirp by

$$\hat{f}_1 = \frac{\partial\hat{f}}{\partial t} = \frac{1}{2\pi}\frac{\mathrm{Im}\dfrac{\partial^2 X_k/\partial t^2}{X_k} - \mathrm{Im}\dfrac{(\partial X_k/\partial t)^2}{X_k^{\,2}}}{-\mathrm{Im}\dfrac{\partial^2 X_k/\partial t\partial\omega}{X_k} + \mathrm{Im}\dfrac{(\partial X_k/\partial\omega)(\partial X_k/\partial t)}{X_k^{\,2}}}, \tag{5a}$$

where the second-order derivatives are calculated with

$$\frac{\partial^2 X_k}{\partial t^2} = \sum_n x_n w''_{-n} e^{-j\frac{2\pi kn}{N}},\ \frac{\partial^2 X_k}{\partial t\partial\omega} = -j\sum_n n x_n w'_{-n} e^{-j\frac{2\pi kn}{N}}. \tag{5b}$$

w'' is the derivative of w'. Using $\hat{f}$ and $\hat{f}_1$, the amplitude and phase angle can be estimated with

$$\hat{a}e^{j\hat{\varphi}} = \frac{\sum w_n^2 x_n e^{-j2\pi(n\hat{f}+n^2\hat{f}_1)}}{\sum w_n^2}, \tag{6}$$

which involves a demodulation step removing the chirp.

The Abe-Smith estimator is implemented following [6]. This method requires the use of a quadratic interpolation estimator. Given a set of primitive estimates $\hat{f}$, $\log\hat{a}$ and $\hat{\varphi}$, as well as two functions $A(f)$ and $\Phi(f)$ derived by interpolating the log-amplitude spectrum and unwrapped phase spectrum respectively, the Abe-Smith estimator calculates

$$p = -\frac{\pi A''(\hat{f})}{A''^2(\hat{f}) + \Phi''^2(\hat{f})},\ \alpha = -2p\Phi'(f),\ \beta = p\frac{\Phi''(\hat{f})}{A''(\hat{f})}, \tag{7a}$$

where α is interpreted as the exponential amplitude decay rate and 2β the linear frequency chirp rate. After estimating these rates the frequency is reestimated using

$$\hat{\hat{f}} = \hat{f} - \alpha\beta/2\pi p. \tag{7b}$$

[6] also proposes reestimation methods for amplitude and phase angle. However, instead of using the original Abe-Smith methods for all parameters, we only use the frequency and frequency slope estimates, then evaluate the amplitude and phase angle using (6). This modification has also been suggested in [8], and shows consistent improvement over the original Abe-Smith method in our tests. We call it an enhanced Abe-Smith method.

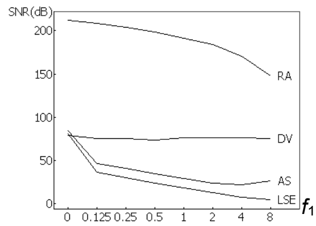
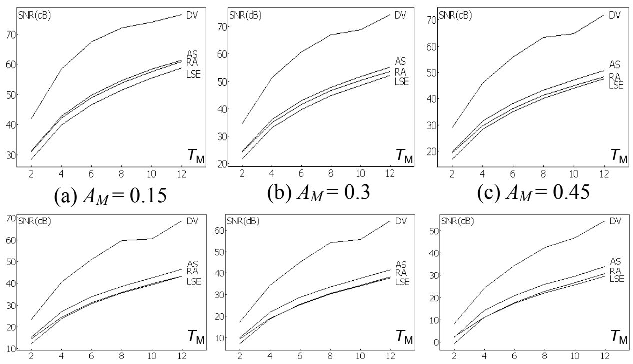


Figure 2: *Result for linear chirps*

Figure 2 compares test results of LSE, RA, AS and DV methods on linear chirps. For these signals RA works perfectly, with SNR between 140dB and 220dB. DV is not as accurate as the former, but also gains significant improvements, with SNR between 70dB and 80dB. AS only moderately outperforms the plain LSE. For linear chirps VW has the same results as LSE, as the LSE estimates from linear chirps have D, defined in (1), close to zero, therefore no shorter window size will be chosen. Likewise VWDV has the same results as DV.



(a) $A_M = 0.15$ (b) $A_M = 0.3$ (c) $A_M = 0.45$

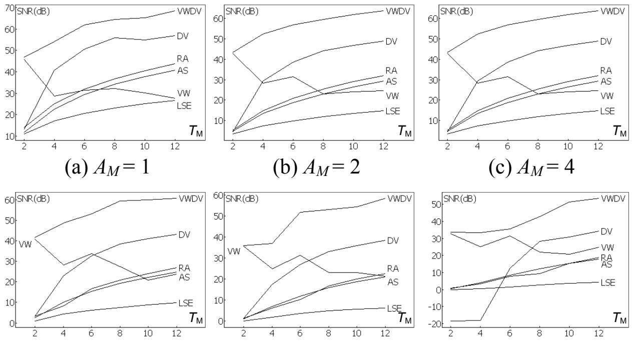
(d) $A_M = 0.6$ (e) $A_M = 0.75$ (f) $A_M = 0.9$

Figure 3: *Result for pure tremolos*

Results of LSE, AS, RA and DV methods are given in Figure 3 for amplitude modulated sinusoids. For these signals RA gains little or no improvement over LSE. AS works slightly better than RA, yet the improvement is less than 5dB. DV works better than both of them. Again, VW and VWDV have the same results as LSE and DV respectively, since D is close to zero.

Results of all 6 methods are given in Figure 4 for frequency modulated sinusoids. All other methods consistently outperform LSE. Again RA and AS methods give similar results, this time the former being slightly better. DV outperforms the two, unless for very fast or very large modulation, in which case all three work poorly. VW appears to be highly effective for fast modulations, but not for slow ones. The nonmonotonicity of the VW results is due to the discontinuous jumps of the window size between powers of 2. The VWDV method, combining DV and

VW schemes, consistently outperforms all other methods, even when DV or VW work poorly by themselves.

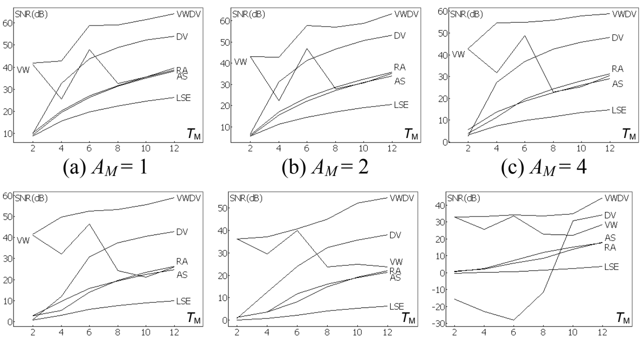


(a) $A_M = 1$ (b) $A_M = 2$ (c) $A_M = 4$

(d) $A_M = 8$ (e) $A_M = 16$ (f) $A_M = 32$

Figure 4: *Result for pure vibratos*

Results of all 6 methods are given in Figure 5 for frequency-and-amplitude modulated sinusoids. They bear much similarity to those in Figure 4.



(a) $A_M = 1$ (b) $A_M = 2$ (c) $A_M = 4$

(d) $A_M = 8$ (e) $A_M = 16$ (f) $A_M = 32$

Figure 5: *Result for vibrato and accompanying tremolo*

5. CONCLUSION

In this paper we have proposed a demodulation method for re-estimating parameters of time-varying sinusoids. It captures the dynamics information embedded in primitive inaccurate estimates, and uses this information to cancel the dynamics of the original sinusoid. The idea of using rough estimates to refine sinusoid analysis was previously explored by us in [13]. However, like many other methods for analyzing time-varying sinusoids, [13] is based on fitting the sinusoid onto a presumed parametric model (cubic spline), and works only with a specific parameter estimator (LSE). The method here, on the other hand, relies only on two loose assumptions: that we are able to find approximate parameters, and that the estimator we use has better accuracy for slower-varying sinusoids. Reestimation using the demodulation scheme not only bypasses potential overfitting problem in model-based approaches, but also allows the use of arbitrary estimators. The cost is a moderately higher computation load than some quick fitting methods such as [6] and [10].

The demodulation method is robust and numerically stable for small parameter dynamics, but works poorly for large ones. We attribute this to the large error in the primitive estimates, which breaches the first assumption above. To deal with this we have proposed a variable-window-size method, also making use of the primitive estimates, that effectively improves the estimation of fast-varying sinusoids. We have also shown that these two independent methods can be easily connected together to make up for the demodulation method's incapability of handling very fast frequency variations, and the variable-window-size method's limited accuracy.

6. ACKNOWLEDGMENTS

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

7. 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(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] F. Keiler and S. Marchand, "Survey on extraction of sinusoids in stationary sounds," in *Proc. DAFx'02*, Hamburg, 2002, pp. 51-58.

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

[5] 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, pp. 153-156.

[6] M. Abe and J. O. Smith III, "AM/FM rate estimation for time-varying sinusoidal modeling," in *Proc. ICASSP'05*, vol.3, Philadelphia, 2005, pp. 201-204.

[7] R. Badeau, B. David and G. Richard, "High-resolution spectral analysis of mixtures of complex exponentials modulated by polynomials," *IEEE Tran. Sig. Proc.*, vol.54(4), 2006, pp.1341-1350.

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

[9] F. Auger and P. Flandrin, "Improving the readibility of time-frequency and time-scale representations by the reassignment method," *IEEE Tran. Sig. Proc.*, vol.43(5), 1995, pp.1068-1089.

[10] A. Röbel, "Estimating partial frequency and frequency slope using reassignment operators," in *Proc. ICMC'02*, Göteborg, 2002, pp. 122-125.

[11] H. K. Kwok and D. L. Jones, "Improved instantaneous frequency estimation using an adaptive short-time Fourier transform," *IEEE Tran. Sig. Proc.*, vol.48(10), 2000, pp. 2964-2972.

[12] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Tran. Sig. Proc.*, vol.51(1), 2003, pp. 101-111.

[13] Wen X. and M. Sandler, "Error compensation in modeling time-varying sinusoids," in *Proc. DAFx'06*, Montreal, 2006, pp. 173-176.

[14] Wen X., *Harmonic Sinusoid Modeling of Tonal Music Events*, PhD Thesis, University of London, 2007.