ON SINUSOIDAL PARAMETER ESTIMATION

Stephen Hainsworth

Cambridge University Engineering Dept.
Cambridge, CB2 1PZ, UK
swh21@eng.cam.ac.uk

Malcolm Macleod

QinetiQ
Malvern, WR14 3PS, Uk.
m.macleod@signal.qinetiq.com

ABSTRACT

This paper contains a review of the issues surrounding sinusoidal parameter estimation which is a vital part of many audio manipulation algorithms. A number of algorithms which use the phase of the Fourier transform for estimation (e.g. [1]) are explored and shown to be identical. Their performance against a classical interpolation estimator [2] and comparison with the Cramer Rao Bound (CRB) is presented. Component detection is also considered and various methods of improving these algorithms are discussed.

1. INTRODUCTION

Sinusoidal parameter estimation is a widely explored subject area and many methods have been proposed to achieve this goal. The short time Fourier transform (STFT), often implemented by the fast Fourier Transform (FFT) serves as a starting point for many of these. One class of methods takes the FFT bins around a potential frequency component and performs regression of a known model order to fit a line or curve, the centre of symmetry of which gives the sinusoidal parameters. Examples of such interpolation methods are Grandke [2] and Macleod [3].

Another group of algorithms explicitly use the phase of the FFT to estimate the instantaneous parameters of the signal. The reassignment method [4, 1] is one of these, though other methods such as the that proposed by Marchand [5] will also be discussed.

2. STATISTICAL PERFORMANCE

There are a number of issues surrounding sinusoidal estimation and the performance of algorithms designed for this task. The problem can be broken down into two stages: detection and estimation; these are separate but inextricably linked. Also, early in the consideration of any method should be its intended application: many algorithms are tested upon limited synthetic data, which is fine if the signal for which the algorithm is ultimately intended closely matches this. However, in the real world, many signals are both non-stationary (i.e. time varying), have a high density of components and/or exhibit non-sinusoidal elements. Musical audio is a classic example of this. Thus we shall first examine the case of a single complex sinusoid in Gaussian noise of known variance σ^2 and build up complexity from there.

2.1. Single Complex Sinusoid

A suitable framework for the statistical comparison of algorithms is the Cramer-Rao Bound (e.g. [6]) which describes the best possible lower bound for estimation error given a signal vector and associated noise statistics. If correct detection of a signal component is assumed for now, it remains to estimate the exact frequency ω, phase θ and amplitude b of that component (note, a complex sinusoid):

$$c[n] = b\exp[j(\omega n + \theta)] \tag{1}$$

Two estimation error effects exist; bias and variance. Bias can be considered the mean estimation error over a number of trials while variance is the variation around this mean. The CRB gives theoretical limits on the variance of an *unbiased* estimate.

Some algorithms exhibit inherent bias in the estimate. The most obvious example is estimation direct from the FFT without any further processing: in this case, if the exact frequency falls between two bins, ie $\omega = (k + \Delta) \times 2\pi F_s/N$ where k is the exact bin and Δ is the offset, then by selecting the maximum amplitude bin in the FFT and using that frequency as the estimate, there will be a bias of Δ.

Reassignment [1] has been shown to exhibit this bias effect (albeit of a tiny order compared to the raw FFT) but for a single sinusoid, it can be removed [7]. The triangle algorithm of Keiler [8] also exhibits this inherent bias though they fail to remove it and hence do not gain maximum performance from their algorithm. Figure 1 gives an example of the effects of bias and detection for the reassignment algorithm.

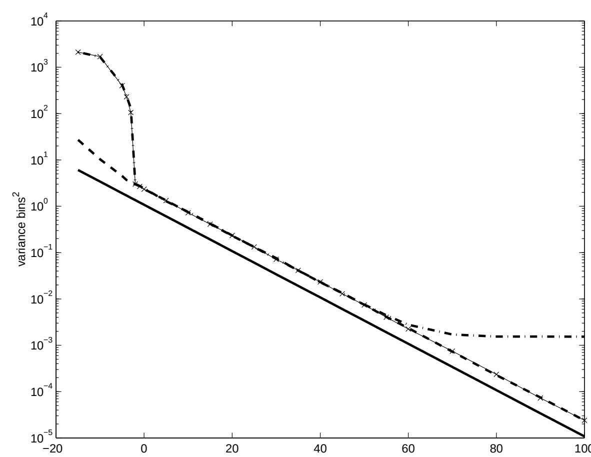


Figure 1: Plot showing the various effects of detection and bias for the reassignment algorithm. - CRB; ·- without inherent bias removed; - - with detection assumed; x- with simple detection algorithm taken into account. The 4dB offset from the CRB is due to the use of windows as described in section 2.2.

2.2. Multiple Components

With multiple components, there is a second bias effect: that of one component upon another. At this point, the issue of windowing becomes important. While it has been shown [6] that windowing increases the variance of an estimator (intuitively, by windowing the outlying data points, the amount of data is being reduced), the rectangular window (i.e. no window) is the worst for this multiple component bias because its side-lobes contain significant power.

In the case of multiple components which are closely spaced in frequency, this bias effect will dominate the estimation unless it is removed. A potential method of bias removal is iterative analysis, e.g. [3], so long as the tones are separated in frequency by several bins. Windowing can reduce the initial bias and speed convergence. For example, while the 3-sample unwindowed interpolator of [3] can get much closer to the CRB (1.2dB), it suffers from high bias leakage and is a very poor starting point for iterative analysis.

It is worth noting at this point that a real sinusoid consists of two complex sinusoids and these will have bias effects on each other, though unless the sinusoid's frequency is very low, they are so far removed from each other in the spectrum that this is small.

2.3. Real Signals

Many real-world signals consist of multiple real sinusoids which are closely spaced. Here, the between-component bias effects are multifarious and very hard to remove completely. Also, the inherent bias due to the method is often still present as well as variance caused by the noise in the signal. The final ingredient of the problem which has been assumed thus far is detection of components. Various methods exist for sinusoidal detection (and hence rejection of peaks which are purely due to noise). Examples are phase consistency [8], noise floor estimation [3] and measures based upon reassignment [9]. Windowing is crucial in this context as well in that it is desirable to have a narrow main lobe (in the frequency domain) to allow closely spaced components to be separated. Finally, most real signals are not stationary and this causes many further effects.

2.4. Summary

Sinusoidal estimation errors consist of three factors: bias inherent in the algorithm (which can sometimes be removed), variance from the noise in the signal and bias from multiple tones. Windowing will increase the variance but reduce the bias. Iterative analysis, as described further in section 6, can be used to help reduce the bias further so long as the components are reasonably separated in frequency. The effects of all of these should be analysed and accounted for when choosing one algorithm over another or selecting settings. A final consideration is that of computational cost which may be an issue for some applications.

3. DESCRIPTION OF ALGORITHMS

3.1. Estimators

Four estimator algorithms are compared in this survey. They all use the discrete Fourier transform as a basis for further processing. The first is that of Grandke [2] which uses the two largest magnitude bins in the Hanning windowed short time Fourier transform to perform an interpolation estimate.

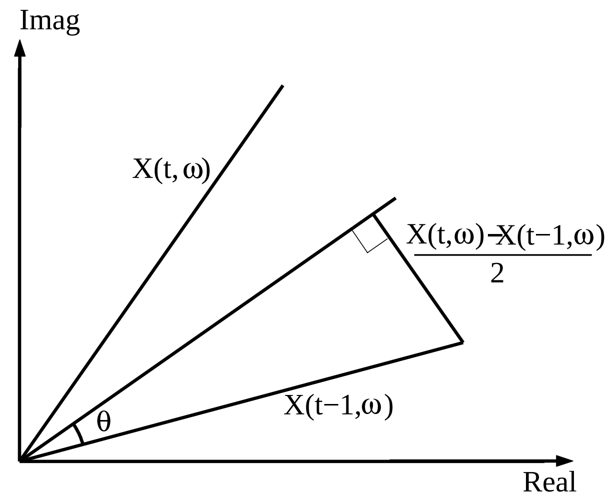


Figure 2: Figure showing vector relationships for Marchand's method of finding instantaneous frequency.

The next method is so-called reassignment [1][1] where the phase of the signal is used, via a set of related windows to produce an improved estimate of frequency. This paper focuses on frequency reassignment which is given by

$$\hat{\omega}(t,\omega) = \omega + \Im\left\{\frac{X_{dh}(t,\omega).X_h^*(t,\omega)}{|X_h(t,\omega)|^2}\right\} \quad (2)$$

where $X_h(t,\omega)$ is the STFT utilising a given window (e.g. Hanning, Blackman) and $X_{dh}(t,\omega)$ is the STFT using the derivative of the original window function as the new window. Reassignment can intuitively be though of as finding the "centre of energy" for each bin and placing the energy there, rather than at the centre of the analysis window. Time reassignment, which is the converse of frequency reassignment, is described by

$$\hat{t}(t,\omega) = t - \Re\left\{\frac{X_{th}(t,\omega).X_h^*(t,\omega)}{|X_h(t,\omega)|^2}\right\} \quad (3)$$

where $X_{th}(t,\omega)$ uses a time ramped version of the original window. It should be noted at this point that frequency reassignment is often identical to many of the instantaneous frequency methods of the past [11, 12] as is proved in [10]. This study uses the reassignment equations with the Hanning window which gives the least intrinsic bias and smallest spectral width of main lobe.

A third approach, which also uses the phase to perform frequency estimation, takes two different windowed data sets, a small number of samples apart and estimates the change in phase directly. This is given by

$$\hat{\omega}(t,\omega) = \frac{\arg(X_h(t,\omega)) - \arg(X_h(t-\delta,\omega))}{\delta} \quad (4)$$

where δ is the separation of the two frames. Dixon [13] uses this method and Arfib et al [14] also describe it. $\delta = 1$ is used in this study. This method is equivalent to reassignment in using phase change to estimate frequency but uses a brute force approach as

[1] A full analysis and review can be found in Hainsworth & Macleod [10]

opposed to the exact instantaneous frequency method of reassignment.

Lastly, Marchand [5] has produced an algorithm based upon signal derivatives. Intuitively, it would seem that this is similar to reassignment where a derivative window is used. However, the implementation given in [15] shows that a difference method is used to generate the signal derivatives. The formulation of Marchand is given by

$$\widehat{\omega}(t,\omega) = 2\arcsin\left(\frac{1}{2}\frac{|X^1(t,\omega)|}{|X(t,\omega)|}\right) \tag{5}$$

where $X^1(t,\omega)$ is the windowed STFT using a differenced signal $d(t) = x(t) - x(t-1)$. It can trivially be proved that $X^1(t,\omega) = X(t,\omega) - X(t-1,\omega)$ which leads to a similar result to the phase differencing approach above. If it is assumed that over a change of one sample, the magnitude of the complex STFT bin values stays constant, an isosceles triangle can be formed from $X(t-1,\omega)$ and $X(t,\omega)$. The unique angle is bisected and this is then described by the arcsin function. Figure 2 shows this graphically. Thus, this method is identical to the phase differencing method of (4), bar numerical issues at frequencies around the Nyquist limit using Marchand's method.

3.2. Detectors

The simplest detector is to make an assumption on the number of signal components present and to then pick the P highest peaks in the STFT. With low amplitude peaks, there is, however, a significant chance that sinusoids falling close to halfway between two bins will be missed in the detection process. Zero padding is a common solution to this, though it has its own downside in that the side-lobes of high amplitude components become "visible" and are likely to be detected.

Therefore, for complex tasks, it is desirable to have more reliable detectors. A survey of the literature gives a number of methods, many of which exhibit undesirable properties. A popular method is to use a measure of fit to the ideal window function for a sinusoid, often by a least squares method [16, 17] but these methods perform badly due to normalisation issues. Keiler [8] describes a method based upon phase consistency where the bins either side of the maximal bin should have a deterministic relationship to each other and the maximal bin. The error between the ideal relationship and the actual phases can give a measure of sinusoidality. Macleod [3] uses a noise floor estimation process to detect components - height above the noise floor determines whether a component is found or ignored. Hainsworth [18] presents a method using reassignment to produce a variety of statistics from which a decision is made as to whether a peak is a sinusoid, a transient or noise. Performance of these algorithms is discussed in section 5.3.

4. CRAMER-RAO BOUND THEORY

The Cramer-Rao lower bound is a measure for evaluating the performance of estimators and the CRB analysis of sinusoid-in-noise estimation algorithms is a well documented area [6, 3]. It is defined to be a limit on the best possible performance achievable for parameter estimation given a dataset. Maximum likelihood estimation will achieve this bound and to increase performance beyond this, prior information must be employed (e.g. [19, 20]).

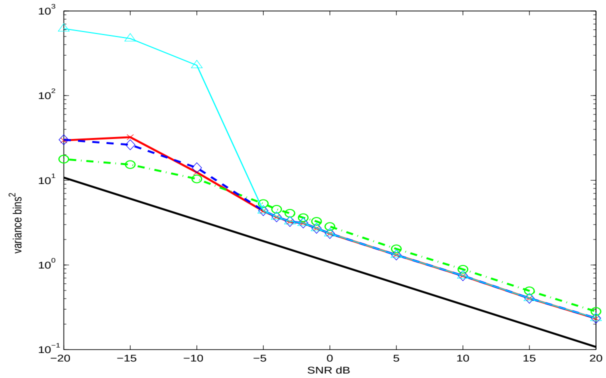


Figure 3: CRB comparison for a single complex tone. (a) - CRB of ML estimator; (b) x, Marchand's estimator; (c) ◊, reassignment using Hanning window; (d) △, difference method; (e) o, Grandke interpolator.

For a single tone embedded in Gaussian noise, the ML estimation algorithm is the one which minimises

$$\sum_{n=0}^{N-1} \hat{v}^2[n] = \sum_{n=0}^{N-1} \left(c[n] - \hat{b}_1 \exp[j(\hat{\omega}_1 n + \hat{\theta}_1)]\right)^2 \tag{6}$$

for the signal

$$c[n] = b_1 \exp[j(\omega_1 n + \theta_1)] + v[n] \tag{7}$$

with parameters

$$\alpha = [\ \omega_1 \quad b_1 \quad \theta_1\ ]^T. \tag{8}$$

and i.i.d. Gaussian noise $v[n]$ with variance σ^2. ML algorithms will often do a coarse and then an increasingly fine grid based search for the parameter set which minimises the error.

Following the derivation in [6], the lower bound, or best possible performance that can be achieved for frequency estimation is

$$\sigma_{\hat{\omega}_1} \geq \frac{6\sigma^2}{b_1^2 N(N^2-1)} \tag{9}$$

This can be seen to be inversely proportional to the signal to noise ratio (SNR) [3].

$$SNR = \frac{b_1^2}{\sigma^2} \tag{10}$$

As mentioned above, there are two effects which can be distinguished for an estimator: bias and variance. The CRB gives a bound on the variance, for an *unbiased* estimate.

5. RESULTS

5.1. Single complex tone

The first comparison to be made is between raw estimation performance without the added complication of detection. Thus, figure 3 shows the performance of the various estimators as they break down due to noise. As the SNR increases, an asymptotic tendency with respect to the CRB[2] is found, with an offset which is due to

[2] All plots show the CRB for unwindowed data.

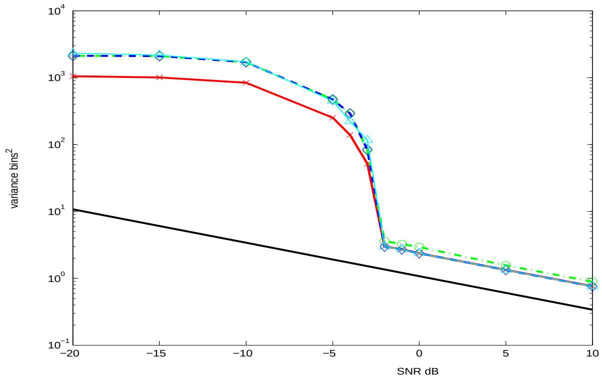


Figure 4: CRB comparison for a single complex tone with detection. (a) - CRB of ML estimator; (b) x, Marchand's estimator; (c) ◊, reassignment using Hanning window; (d) △, difference method; (e) o, Grandke interpolator.

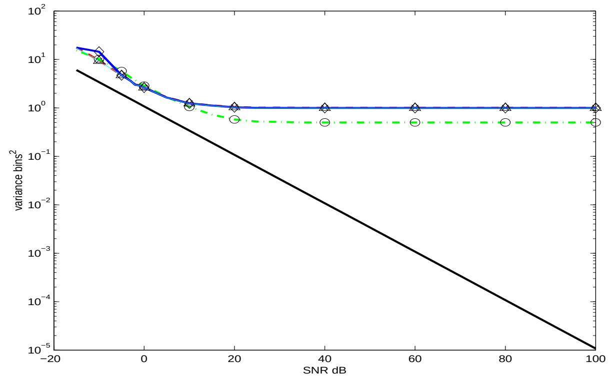
the particular method used (related to the windowing). The results here are independent of window length and sample frequency used though they were generated using a window length, $N = 128$ samples and sample rate $F_s = 4$kHz. They were also averaged over a range of frequencies from 1Hz to $F_s/2$ and inter-bin offsets Δ, though again, the estimator performance was independent of signal frequency and Δ. The three methods which use phase all perform almost equivalently, which is not surprising, given the underlying similarity. They all achieve slightly better performance than the Grandke interpolator by 0.8dB.

Figure 4 shows the performance of each estimator when detection is taken into consideration. For this purpose, the simplest detection scheme of picking the $P = 1$ highest peak was used. It can clearly be seen where the sinusoid is overwhelmed by the noise and the detection becomes random, eventually achieving the second asymptote which is equivalent to picking a random frequency. The performance of Marchand's algorithm in this failure region is artificial: numerical issues with the algorithm mean that it is impossible to pick a value around the Nyquist frequency so the results are being averaged over a smaller range.

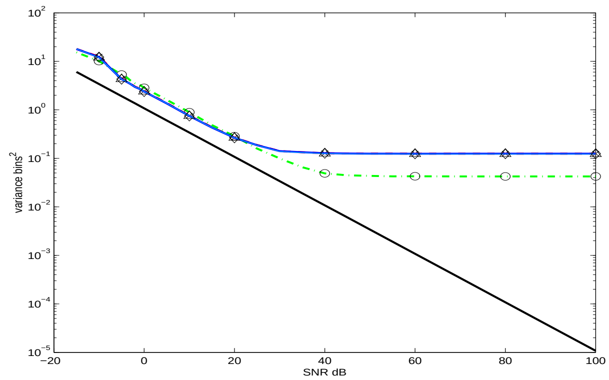
5.2. Multiple complex tones

With multiple complex tones, intercomponent bias is now an issue and this is proportional to the frequency separation between the components. To explore this additional effect, trials were performed for a number of frequency separations between two complex tones. Again with $F_s = 4$kHz and $N = 128$, the lower tone was held constant at around 200Hz while three frequency separations were used: 90Hz (equivalent to 2.88 bins separation), 200Hz (6.4 bins) and 700Hz (22.4 bins). The trials show results averaged over a range of Δ inter-bin offsets and phase relationships. These form a deterministic pattern of varying intercomponent bias.

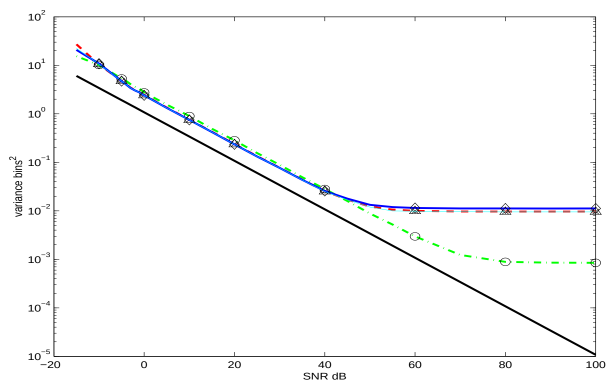
The results for these can be found in figure 5 for the higher of the two tones, plotted against the CRB for a single tone as comparison. Detection issues are ignored and estimation performance is the all that is shown. The interesting points to note are that for high noise (low SNR), the three phase methods all perform approximately equivalently and better than Grandke. However, for high SNR, where the tone is clear from the noise, the bias effect for Grandke is lower than it is for the phase methods. The other



(a) 90Hz separation



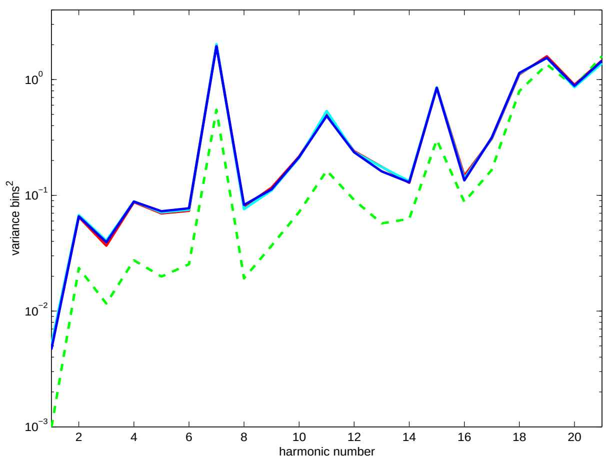
(b) 200Hz separation



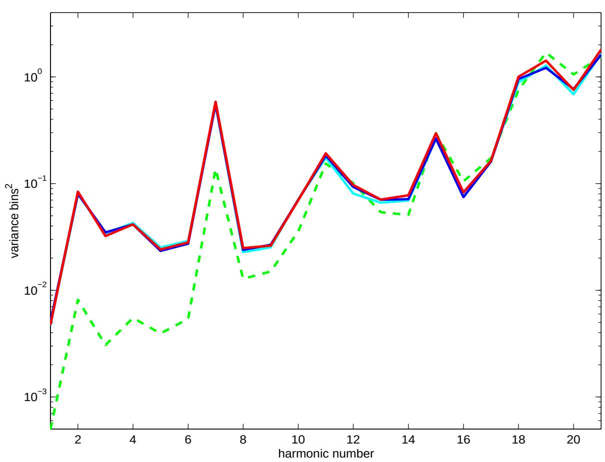
(c) 700Hz separation

Figure 5: Frequency estimate variance averaged over Δ and a range of phase for the higher of two complex tones. (a) - CRB of ML estimator (single tone); (b) x, Marchand's estimator; (c) ◊, reassignment using Hanning window; (d) △, difference method; (e) o, Grandke interpolator.

effect is that the threshold at which the bias becomes the limiting factor changes, as expected, with frequency separation of the two tones.



(a) f_0 =100Hz



(b) f_0 =350Hz

Figure 6: Real tone estimation performance for two f_0 values. Solid lines are the three phase estimation methods; dashed is Grandke.

To further explore this effect, a real piano signal from the McGill Master Samples database [21] was analysed for spectral content and then synthetically reproduced so as to have a ground truth with known frequencies and noise statistics. $F_s = 44.1$kHz and $N = 2048$ samples were used and 21 harmonics were reproduced with varying amplitudes from 54dB to only 4dB above the noise floor. Random phase relationships were chosen and two values for the fundamental frequency, f_0 were tested - 100Hz and 350Hz. Results, averaged over 100 trials can be seen in figure 6 again with detection assumed. For the smaller inter-tone separation, the Granke interpolator is better than the phase methods for a larger range of SNR whereas for f_0 =350Hz, the phase methods, as expected, perform marginally better for a larger range of SNR values.

	Macleod		Keiler	
	N_d	N_f	N_d	N_f
Pseudo $f_0 = 100$Hz	14.07	0.24	20.41	87.24
Pseudo $f_0 = 300$Hz	21	0.28	20.66	75.86
Real guitar	15	0	16	100

Table 1: Results for detection algorithms. N_d is average number of correctly detected tones. N_f is average number of falsely detected tones.

5.3. Detection performance

Raw detection is best tested upon synthetic signals where the exact number and nature of the components is known. The pseudo-real piano sound with 21 harmonics from the last section was therefore used for test purposes with two detection methods: Macleod's noise floor method and Keiler's phase consistency test. Table 1 gives results of this for two different fundamental frequencies, with results averaged over 100 trials with random phase relationships. Also included are results from a real signal consisting of a guitar tone with 16 identifiable harmonics (only one single trial was used for this).

It can be seen that the noise floor method has a very low false alarm rate but in high tone density, has a tendency to miss components while the Keiler phase consistency test has an over-detection problem in that it classifies noise as sinusoidal. All schemes tested by the authors suffered from one or other of these behaviours; the trade-off between false alarm rate and missed detections is a fundamental problem with all detectors.

6. ITERATIVE ANALYSIS

Macleod [3] describes a method for overcoming some of the problems above: if iterative analysis is used, whereby components are detected, estimated and removed from the spectrum before moving on to the next then detection of low energy components becomes more possible. Also, inter-component bias effects are reduced by this. The only problems occur with very closely spaced tones when they become inseparable, and also with real signals with non-sinusoidal components. These leave residuals after removal which can in turn be detected as components if the fit is especially bad.

Figure 7 gives a comparison of the iterative detection/estimation performance for the Grandke estimator [2] with Macleod's noise floor detector [3]. Significant estimation improvement can be seen at high SNR (the first 5 harmonics). Also, for the pseudo-piano tone with $f_0 = 100$Hz, the detection algorithm now finds all 21 tones, though the number of false alarms is also increased slightly.

7. CONCLUSIONS

A number of different tests have been described and comparison with the CRB given. From the results a number of conclusions can be drawn.

- The choice of estimation algorithm is dependent upon the data: for sparse, high noise data, the reassignment algorithm works best, while in low noise or high component density, Grandke is a better estimator.
- Reliable detection of sinusoids is difficult but a sensible combination of noise floor comparison and sinusoidality

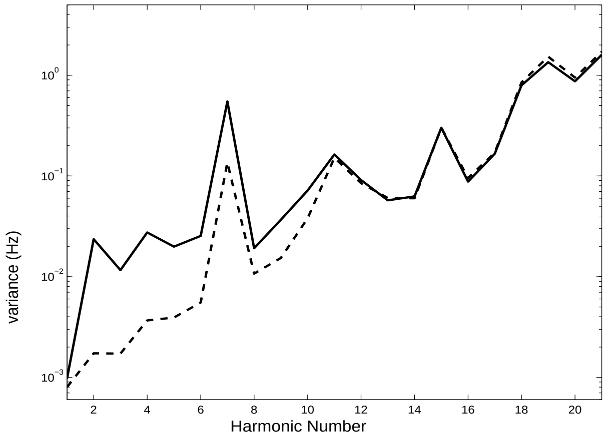


Figure 7: Plot showing iterative detection/estimation (dashed) verses normal detect all then estimate (solid) for Macleod detection and Grandke estimation using pseudo-piano tone with $f_0 = 100$Hz.

testing seems to offer the most likely best solution. [18] describes an attempt to do this.

- The comparison of phase methods for frequency estimation shows that in normal conditions, the three tested methods all performed approximately equivalently. However, Marchand's algorithm performs poorly at high frequency and the phase differencing method breaks down in noisy conditions. Therefore it is suggested that reassignment is the better formulation, given that they are all of the same computational order.
- Reassignment has the added advantage, which has been overlooked in this study, of giving improved time localisation. This can be utilised for transient analysis [10].
- Iterative detection/estimation gives improved performance both for estimation and detection. This improvement could be applied to reassignment, though it has not explicitly been considered in this study.

8. REFERENCES

[1] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the method of reassigment," *IEEE Trans. Signal Processing*, vol. 43, no. 5, pp. 1068–89, May 1995.

[2] T. Grandke, "Interpolation algorithms for discrete Fourier transforms of weighted signals," *IEEE Trans. Instr. and Meas.*, vol. IM-32, no. 2, pp. 350–5, June 1983.

[3] M.D. Macleod, "Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones," *IEEE Trans. Signal Processing*, vol. 46, no. 1, pp. 141–148, Jan. 1998.

[4] K. Kodera, R. Gendrin, and C. de Villedary, "Analysis of time-varying signals with small BT values," *IEEE Trans. ASSP*, vol. 26, no. 1, pp. 64–76, Feb. 1978.

[5] S. Marchand, "Improving spectral precision with an enhanced phase vocoder using signal derivatives," in *Proc. Digital Audio Effects Workshop (DAFx)*, 1998, available on-line: http://www.iua.upf.es/dafx98/papers/.

[6] D.C. Rife and R.R. Boorstyn, "Multiple tone parameter estimation from discrete-time observation," *Bell System Technical Journal*, vol. 55, no. 9, pp. 1389–1410, Nov. 1976.

[7] S.W. Hainsworth and M.D. Macleod, "Time-frequency reassignment: Measures and uses," in *Proc. Cambridge Music Processing Colloquium*, 2003, pp. 36–9.

[8] F. Keiler and U. Zölzer, "Extracting sinusoids from harmonic signals," *J. of New Music Research*, vol. 30, no. 3, pp. 243–58, 2001.

[9] G. Peeters and X. Rodet, "SINOLA: a new analysis/synthesis method using spectrum peak shape distortion, phase and reassigned spectrum," in *Proc. International Computer Music Conference*, 1999.

[10] S.W. Hainsworth and M.D. Macleod, "Time-frequency reassignment: a review and analysis," Tech. Rep. CUED/F-INFENG/TR.459, Cambridge University Engineering Department, 2003.

[11] T. Abe, T. Kobayashi, and S. Imai, "Robust pitch estimation with harmonics enhancement in noisy environments based on instantaneous frequency," in *Proc. ISCLP*, 1996, pp. 1277–80.

[12] D.H. Friedman, "Instantaneous-frequency distribution vs time: an interpretation of the phase structure of speech," in *Proc. ICASSP*, 1985, pp. 1121–4.

[13] S. Dixon, "Extraction of musical performance parameters from audio data," in *Proc. IEEE Pacific-Rim Conf. on Multimedia*, Sydney, Australia, Dec. 2000, pp. 42–45.

[14] D. Arfib, F. Keiler, and U. Zölzer, "Source-filter processing," in *DAFX - Digial Audio Effects*, U. Zölzer, Ed., chapter 9, pp. 299–372. J. Wiley and Sons, 2002.

[15] F. Keiler and S. Marchand, "Survey on extraction of sinusoids in stationary sounds," in *Proc. Digital Audio Effects Workshop (DAFx)*, 2002, pp. 51–8.

[16] G. Peeters and X. Rodet, "Signal characterization in terms of sinusoidal and non-sinusoidal components," in *Proc. Digital Audio Effects Workshop (DAFx)*, 1998.

[17] P. Fernándes-Cid and F.J. Casajús-Quirós, "Multi-pitch estimation of polyphonic music signals," in *Proc. ICASSP*, 1998, vol. 6, pp. 3565–8.

[18] S.W. Hainsworth, M.D. Macleod, and P.J. Wolfe, "Analysis of reassigned spectrograms for musical transcription," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, NY, Oct. 2001.

[19] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2667–76, Oct. 1999.

[20] S.J. Godsill and M. Davy, "Bayesian harmonic models for musical signal analysis," in *7th Valencia Int. Meeting on Bayesian Statistics*, 2002.

[21] McGill University, "Mcgill university master samples collection," http://www.music.mcgill.ca/resources/mums/html/mums.html.