ALGORITHMS FOR IMPROVED SINUSOIDAL TRACK ONSET LOCALISATION

Stuart K. Marks, Ruben Gonzalez
School of Information Technology, Griffith University
PMB 50, Gold Coast Mail Centre QLD 9726
Australia
Email: s.marks@griffith.edu.au, r.gonzalez@griffith.edu.au

Abstract

The sinusoidal model is a key component in many speech and audio coding applications, however its reconstruction quality is severely limited by its poor time resolution, and is unable to capture the onset of sinusoidal partials in a signal accurately. This paper proposes two algorithms which improve the onset localisation of sinusoidal tracks.

While significant improvement can be achieved by using multiresolution sinusoidal analysis to manage the time/frequency resolution tradeoff in individual frequency-subbands, multiresolution sinusoidal analysis cannot improve upon the inherent time/frequency resolution of the Discrete Fourier Transform. The proposed algorithms work in collaboration with multiresolution sinusoidal analysis to improve the time resolution of the sinusoidal model.

This paper will show that these algorithms drastically reduce the average onset error, without adjusting the frequency resolution of the sinusoidal model.

Key Words
Audio and Speech Processing, Time/Frequency resolution, Sinusoidal Model.

1. Introduction

The sinusoidal model [1] has become a key component in many speech and audio coding applications. Its popularity is due to its suitability to the nature of musical sound and speech signals and the perception of sound within the human auditory system.

The quality of the reconstructed audio from the sinusoidal model is highly dependent on the ability of the model to capture the time varying nature of sinusoidal partials within the signal [2]. The sinusoidal model's ability to achieve this with adequate accuracy is limited by the time/frequency resolution of the chosen frequency analysis technique. In many implementations of the sinusoidal model the time resolution required to capture the signal onset can't be achieved, while maintaining adequate frequency resolution to avoid the production of audible artefacts.

Recent model coders have used transient models in collaboration with sinusoidal models to capture the onset and offset times of sinusoidal signals [3,4,5,6]. The focus of this paper is to enable the sinusoidal model to improve the accuracy of the onset and offset times autonomously.

The Discrete Fourier Transform (DFT) is a key component in most sinusoidal models as it provides efficient transformation of discrete time domain signals into bases of complex exponentials. The window length of the DFT is the key factor in determining the time/frequency resolution. By decreasing the window length the time resolution is improved, but the frequency resolution is lost, extending the window length has the opposite effect. This is the nature of the time/frequency resolution tradeoff, making the required time and frequency resolution for a given application difficult, if not impossible, to achieve.

With multiresolution sinusoidal analysis the time/frequency resolution tradeoff is handled within frequency subbands to better match the characteristics of the human ear [7,8]. With this approach high frequency subbands can improve time resolution, as the frequency resolution at these high frequencies is less than at lower frequencies. Whereas the lower frequency subbands require improved frequency resolution, so time resolution is sacrificed. For many audio signals this approach is appropriate as high frequency signals are typically short in duration, whereas low frequency signals generally have increase duration.

While multiresolution sinusoidal analysis allows the time/frequency resolution tradeoff to be managed independently within frequency subbands, multiresolution analysis doesn't improve upon the time/frequency resolution of the DFT.

This papers examines algorithms for localising the onset of a sinusoidal tracks produced by the sinusoidal model without adjusting the frequency resolution of the DFT, resulting in improved reconstructed audio quality. While only onset time is discussed in the following sections, these algorithms provide that same benefits for offset localisation of the sinusoidal signal.

The next section in this paper introduces the theory behind the onset localisation algorithms, and the descriptions of the two proposed algorithms. The section following provides the results from experiments measuring the accuracy of the algorithms with elementary signals. The final section presents a conclusive discussion on the performance of these algorithms.

2. Algorithms

The first step in sinusoidal analysis is sinusoidal estimation, a number of sinusoidal estimation techniques have been proposed in the literature including direct DFT estimation [1], quadratic interpolation [9], cross-correlation [10], F-test [11], Iterative Least Squares [12] and Signal Derivative [13]. These techniques all use an overlapped window DFT, known as a Short Time Fourier Transform, as part of the sinusoidal estimation process.

An important parameter for the STFT is the amount of window overlap between successive estimate windows. Often the overlap is expressed as the percentage of signal covered by successive windows, or at times the stride or leap of the window overlap is used. In this paper the overlap will be expressed as the window overlap factor, s , which is equal to the reciprocal of the window stride. E.g. for a STFT with 75% window overlap, 25% window stride, the window overlap factor $s = 4$. The following discussions also assume that the window overlap factor is a positive integer.

As a signal onset is encountered by a STFT the estimates produced by a technique using the STFT will be distorted according to the location of the signal onset in relation to the position of the overlapped windows.

This is shown in Figure 2.1 which shows the sequence of amplitude estimates which are produced using the quadratic interpolation sinusoidal estimation technique with window overlap factors of two, four and eight for a amplitude-modulated signal. The estimates are joined into sinusoidal tracks, which trace the envelope of the amplitude-modulated signal, but at the signal onset the envelope dips below the signal amplitude and the sinusoidal tracks begin before the signal onset.

Increasing the window overlap factor has the advantage of providing a more accurate representation of the evolving sinusoidal signal, but increases the severity of the onset error.

Figure 2.1 demonstrates a number of phenomenon with the estimates at and around the signal onset. These phenomenon will allow for improved onset localisation of the sinusoidal track. Firstly the first s sinusoidal estimates are affected by the proportion of the signal that was covered by their window, where s is the window overlap factor. Secondly the true signal onset will occur in the last $\frac{N}{s}$ samples of the first estimate window, where N is the window length. Finally the first $\frac{s}{2}-1$ estimates are superfluous as they occur significantly before the signal onset.

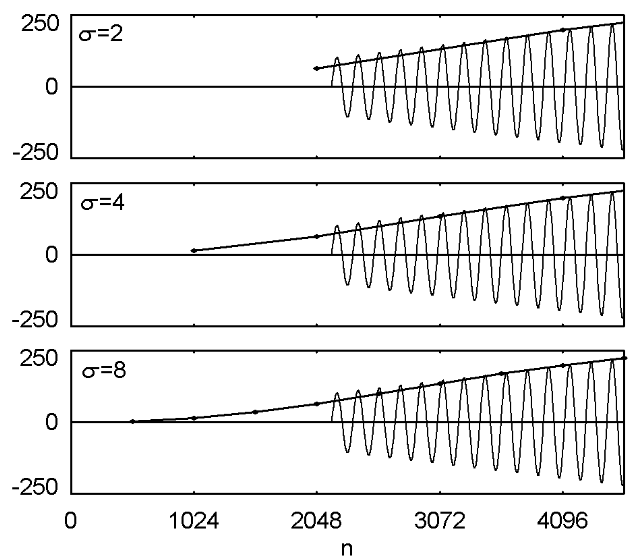

Figure 2.1 Smearing of the onset of a 250Hz amplitude-modulated sinusoidal signal with onset at $n = 2172$ for various window overlap factors and a window length of 4096 samples.

High quality sinusoidal model use sinusoidal tracking to improve reconstruction quality and decrease data rates [14]. The sinusoidal track produced for the amplitude-modulated signal in Figure 2.1 will have an artificial ramp-up at the signal onset due the onset error. When the window overlap is increased the duration of the ramp-up increases, adversely affecting the reconstruction quality at the onset of the signal

The algorithms proposed in the following two subsections utilise these phenomenon at the signal onset to improve the localisation of the track onset. The algorithms are rather aggressive in their application of the phenomenon. Therefore they rely on the sinusoidal track containing a single evolving sinusoidal partial in its entirety. If this assumption isn’t satisfied the algorithms may produce artefacts.

2.1 Estimate-Based Localisation Algorithm

Estimate-based onset localisation attempts to localise the onset timing of the sinusoidal track to the actual onset of a signal by using the information contained within the sinusoidal estimates. The estimate-based approach uses the fact that the onset of a signal will only cover a portion of the first few windows, adversely effecting the accuracy of these estimates. Examining how these estimates have been affected can allow approximation of the onset time of the signal, and correction for the adverse effects on the estimates.

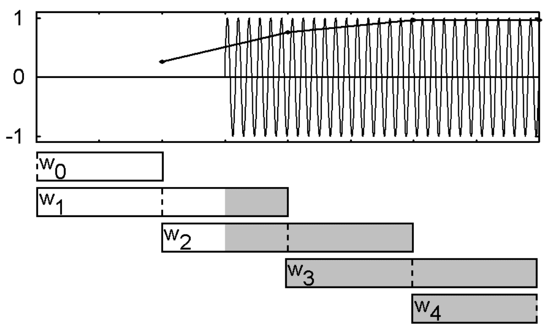

Figure 2.2 Window sequence for sinusoidal modeling using $s = 2$, with the produced amplitude estimates. Shading in windows indicates proportion of the signal within the window.

The effects on the estimates is dependent on the windowing scheme used within the STFT, namely the window function, $w(n)$, and the window overlap factor, s. The estimates in Figure 2.2 were generated from a STFT using a rectangular window function with a window overlap factor of two. The 250Hz steady-sinusoidal onset is halfway between the centres of the first two windows. The first two amplitude estimates are produced from a DFT frame which only contains $\frac{1}{4}$ and $\frac{3}{4}$ of the signal respectively. While the frequency and phase estimates will remain largely unaffected, the amplitude estimate will be scaled directly according to this proportion. In this case the first estimate will be one fourth of the actual amplitude, and the second estimate will be three quarters of the actual amplitude.

This information can be used in collaboration with the known amplitude to reduce the ramp-up of the reconstructed signal by adjusting the amplitude estimates and estimating the onset time of the sinusoidal signal. If a rectangular window function is used, the onset offset is estimated using (2.1) where N is the window length, $\hat{A}$ is the amplitude estimate for the window, and A_{actual} is the actual amplitude of the signal at the centre of the window.

$$\begin{aligned} n_{offset} &= \tfrac{N}{2} - n_{portion} \\ &= \tfrac{N}{2} - \frac{\hat{A}N}{A_{actual}} \end{aligned} \tag{2.1}$$

When the window function is some function other than a rectangular window function $n_{portion}$ is the lowest x which satisfies (2.2), where $w(n)$ is the window function.

$$\sum_{n=0}^{x-1} w(n) \geq \frac{\hat{A}}{A_{actual}} \sum_{n=0}^{N-1} w(n) \tag{2.2}$$

Obviously the actual amplitude is not known, and a predicted amplitude from an established amplitude trend from close-by estimates is used for the onset-offset estimation.

The complete algorithm operates on sinusoidal tracks independently by performing the following steps:

Step 1: Delete superfluous estimates from the sinusoidal track
Step 2: Establish an amplitude trend from close-by unaffected estimates.
Step 3: Calculate the onset offset using the amplitude trend and the first estimated amplitude.
Step 4: Offset the timing of the first estimate by the onset offset.
Step 5: Adjust the amplitude estimates of all affected estimates.

This algorithm has provided drastic improvements with localisation of the sinusoidal track onset time and reconstruction quality. However the accuracy of the estimate is dependent on the accuracy of the amplitude estimates, and the validity of the established amplitude trend.

2.2 Inter-Subband Localisation Algorithm

The Inter-Subband onset localisation method utilises information from multiresolution analysis to estimate the onset of the signal. The algorithm concept is straightforward, as it simply uses the improved time resolution from other frequency-subbands to improve the onset time of a sinusoidal track which begins in a subband with poorer time resolution.

The algorithm finds a track with the closest harmonic match from frequency subbands with improved time-resolution, and uses this signal's improved time resolution to localise the onset of the signal. This approach works well for many real world signals as they contain broadband energy and harmonics at their onset.

Figure 2.3 shows how even a single sinusoid will create noise of low amplitude in higher frequency subbands due to the sudden onset of the signal. The onset range is known from the onset phenomenon and the inter-subband algorithm searches the frequency subbands with improved time-resolution for a sinusoidal track with an onset time within this range and the closest harmonic match.

The inter-subband onset localisation algorithm works in a similar manner to the estimate-based onset localisation algorithm, but this time onset offset in step 3 is calculated from the difference in onset times from the track and the closet harmonically matching track.

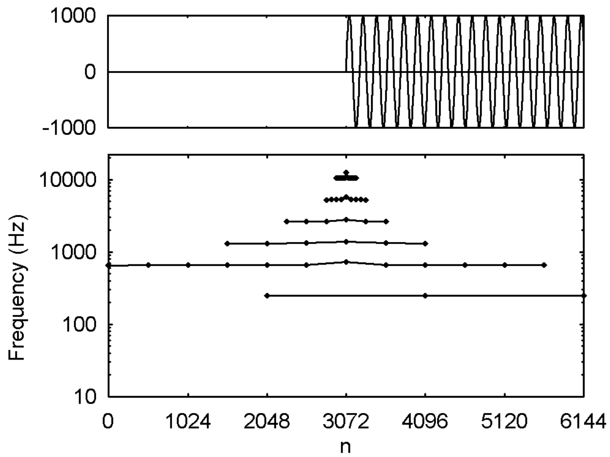

Figure 2.3 Onset noise in higher frequency subbands due to onset of a 250Hz sinusoidal signal at n = 3072 for a six-subband sinusoidal analysis system.

The inter-subband localisation algorithm can perform well, but performance is dependent on the accuracy of the initial frequency of the tracks, which may be inaccurate for when short signal portions are contained within a DFT frame. Also some harmonic matches will not provide a drastic improvement in onset localisation, as can be seen in Figure 2.3 with the fourth subband track.

3. Results

The accuracy of the onset localisation algorithms were compared with no localisation for a number of elementary signals, including steady sinusoids, amplitude-modulated sinusoids and frequency-modulated sinusoids. The sinusoidal analysis utilises quadratic-interpolation sinusoidal estimation and the heuristic sinusoidal tracking algorithm [1] to produce the sinusoidal tracks.

The number of frequency subbands and the window overlap factor were varied, and for each combination the onset error for each algorithm was averaged across a number of frequencies in the audible frequency range (20Hz to 22.1kHz) with various onset delays.

Figure 3.1 shows the onset error for steady sinusoids when no attempts are made to improve the accuracy of the onset time of the sinusoidal track. As the window overlap increases the average onset error increases. Increasing the number of subbands allows improved time resolution in higher subbands, reducing average onset error.

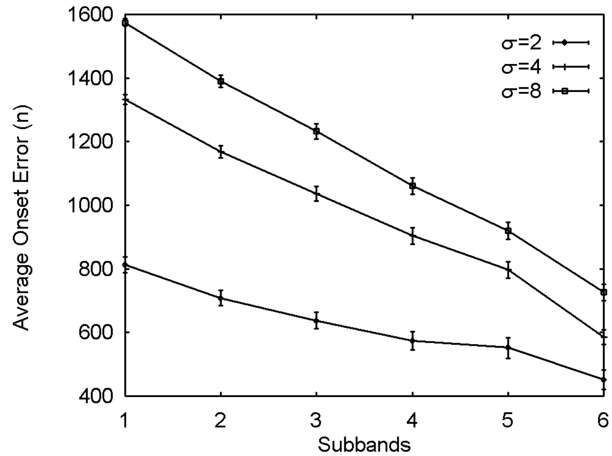

Figure 3.1 Average onset error for a steady-sinusoidal signals without localisation.

When the estimate-based onset localisation algorithm is used the average onset error is drastically decreased. This is highlighted in Figure 3.2 which shows the average onset error for steady-sinusoid signals. There is significant variation with the averages due to the estimation error caused by inaccurate amplitude trends.

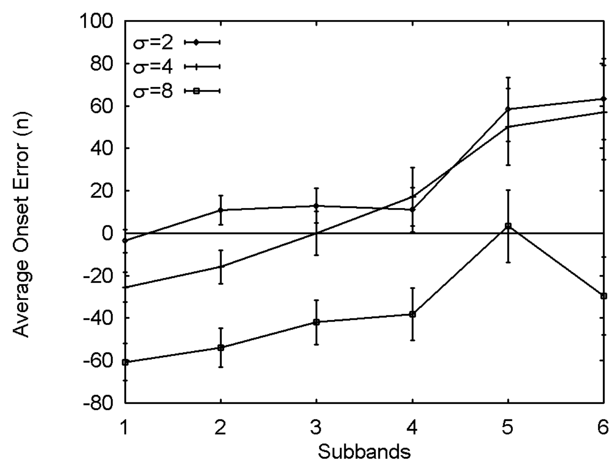

Figure 3.2 Average onset error for the estimate-based onset localisation algorithm with a steady-sinusoidal signal.

The estimation inaccuracy due to an inaccurate amplitude trend is further highlighted in Figure 3.3 which shows the average onset error for the estimate-based localisation algorithm for an amplitude-modulated sinusoidal signal with a quadratic envelope. Due to the variation in the amplitude of the signal, the amplitude trend becomes more inaccurate. This causes the average onset error to increase by around two times in comparison to the steady-sinusoidal results. However the average onset error for the estimate-based localisation algorithms still remains significantly better when no onset localisation is used.

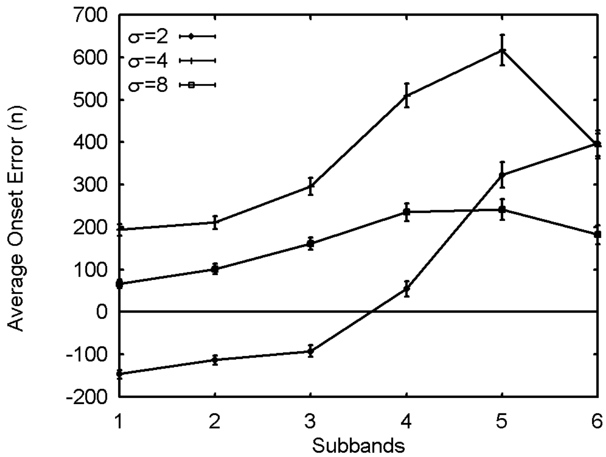

Figure 3.3 Average onset error for the estimate-based onset localisation algorithm with an amplitude-modulated sinusoidal signal.

The inter-subband onset localisation algorithm doesn't perform as well as the estimate-based onset localisation algorithm as is shown in Figure 3.4. This is because for a large number of cases the track with the best harmonic match doesn't provide a large improvement in onset timing, especially for a multiple subband sinusoidal model. This is the reason for the increase in average onset error from the two-subband system to the six-subband system.

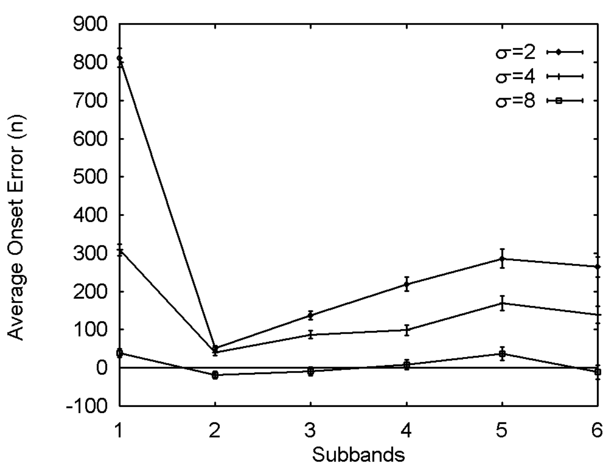

Figure 3.4 Average onset error for the inter-subband onset localisation algorithm with a steady-sinusoidal signal

The inter-subband onset localisation algorithm is able to perform well even when the frequency of the sinusoidal track varies; this is shown in Figure 3.5 which plots the average onset error for the inter-subband algorithm for a linear frequency-modulated sinusoidal signal. The average onset error is comparable with the average onset error for a steady-sinusoidal, allowing for the differences between steady and frequency modulated sinusoidal signals.

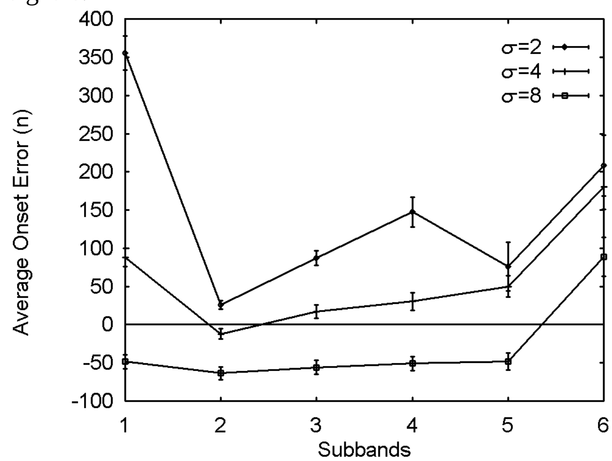

Figure 3.5 Average onset error for the inter-subband onset localisation algorithm with a frequency-modulated sinusoidal signal

4. Conclusion

The benefits of using the onset localisation algorithms for improving the time resolution of sinusoidal analysis have been demonstrated in this paper. The estimate-based algorithm slightly outperforms the inter-subband algorithm, but both greatly improve upon the performance experienced without onset localisation.

The algorithms achieve this improvement by utilising information from the estimates and subbands which up to now has been ignored. The complexity cost is minimal as the algorithms operate on individual sinusoidal tracks after the sinusoidal estimation and tracking algorithms have completed.

5. Acknowledgements

Australian Research Council's Spirt Scheme, and Active Sky Inc.

References

[1] McAulay, R. and Quatieri, T., Speech Analysis/Synthesis Based on a Sinusoidal Representation, *IEEE Transactions on Acoustics, Speech, Signal Processing*, *34*(4), Aug. 1986, 744-754,

[2] Goodwin, M. and Vetterli, M., Time-frequency signal models for music analysis, transformation, and

synthesis, *Presented at IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Paris, France, 1996.

[3] Verma, T. and Meng, T., A 6Kbps to 85Kbps scalable audio coder, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '00, 2, 2000, 877-880.

[4] Painter, T. and Spanias, A, Perceptual segmentation and component selection in compact sinusoidal representations of audio, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 5, 2001, 3289-3292.

[5] Hamdy, K., Ali, A. and Tewfik, A., Low bit rate high quality audio with combined harmonic and wavelet representations, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1996.

[6] Levine, S., *Audio representations for data compression and compressed domain processing*, PhD thesis, Stanford University, 1998.

[7] Levine, S., Verma, T. and Smith, J., Alias-free, multiresolution sinusoidal modeling for polyphonic, wideband audio, *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, 1997.

[8] Boff, K., Kaufman, L. and Thomas, J., *Handbook of perception and human performance, 1*, (John Wiley and Sons, New York, 1986).

[9] Smith, J. and Serra, X., PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation, *Proceedings of the International Computer Music Conference*, 1987. 290-297.

[10] Rodet, X., Musical sound signal analysis/synthesis: Sinusoidal + residual and elementary waveform models, *IEEE Time-Frequency and Time-Scale Workshop*, 1997.

[11] Thomson, David, Spectrum estimation and harmonic analysis, *Proceedings of the IEEE, 70(* 9), 1982, 1055-1096.

[12] Depalle, P. and Helie, T., Extraction of spectral peak parameters using a short-time Fourier transform and no sidelobe windows, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, USA, 1997

[13] Desainte-Catherine, M. and Marchand, S., High precision Fourier analysis of sounds using signal derivatives, *Journal of Acoustic Engineering Society, 48*(7), 2000.

[14] Verma, T., *A perceptually based audio signal model with application to scalable audio compression*, PhD Thesis, Stanford University, 1999.