Techniques for Improving the Accuracy of Sinusoidal Tracking

Author

Marks, Stuart, Gonzalez, Ruben

Published

2005

Conference Title

Internet and Multimedia Systems and Applications ~EuroIMSA 2005~

TECHNIQUES FOR IMPROVING THE ACCURACY OF SINUSOIDAL TRACKING

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

Abstract

This paper proposes three novel techniques for improving the accuracy of the heuristic sinusoidal tracking algorithm proposed in [1]. When applied to audio coding these techniques extend the traditionally speech coding approach into true wideband audio coding.

These techniques provide proper multiresolution sinusoidal tracking, matching across a number of variables instead of frequency alone, and global optimization of these variables across sinusoidal tracks.

When these techniques are used together a heuristic tracking algorithm is created which has many of the benefits of tracking algorithms formulated as a Hidden Markov Model (HMM) problem, but at a far reduced computational and implementation complexity.

Key Words

Multimedia information systems, sinusoidal, tracking, sinusoidal model, quality, compression.

1. Introduction

Sinusoidal tracking, also referred to as frequency line tracking in the literature, is the process of creating sinusoidal tracks from a sequence of estimates. Sinusoidal tracking is a crucial component in many areas of signal processing, including: radar, seismology, sonar, and radio astronomy, and audio. The focus for the remainder of this paper is with audio signal processing, in particular audio coding and auditory scene analysis.

A heuristic algorithm for performing sinusoidal tracking was introduced by McAulay and Quatieri 1986 [1] and has remained a popular tracking algorithm with sinusoidal audio coders and auditory scene analysis. Despite this popularity little literature has been devoted to improving the performance of the heuristic sinusoidal algorithm.

There have been a number of different tracking algorithms developed, including reformulations into a Hidden Markov Model (HMM) framework [2,3,4,5], reformulation into a bipartite matching problem [6] and Neural Network approaches [7]. These algorithms haven't become popular in audio signal processing due to a number of reasons, including: the computational complexity for multiple sinusoidal tracks is much larger than the heuristic algorithm, frequency resolution is often reduced, and increased implementation complexity.

For audio signal processing sinusoidal tracking takes estimates of windowed sinusoidal components. Usually the estimates are produced from Discrete Fourier Transform (DFT) coefficients from a Short Time Fourier Transform (STFT). The tracking algorithm produces sinusoidal tracks which model the evolution of particular sinusoidal partials. This allows for improved reconstruction quality as it enables interpolation between the sinusoidal estimates, improved compression rates [8] due to reduced entropy, and improved high-level modeling due to the creation of meaningful audio objects [9].

Obviously the benefits of sinusoidal tracking are dependent on the performance of the sinusoidal tracking algorithm. If the algorithm incorrectly joins estimated sinusoidal components into tracks, then the tracks loose their significance.

While the original heuristic algorithm has adequate performance for it's application to speech coding, it has a number of limitations which impact its performance with polyphonic audio; namely inability to track multiple harmonic structures, effects from non-harmonic partials, crossing sinusoidal partials, and large frequency variations [10].

Some literature has addressed a few of these limitations [11,12], but there still remains a number of shortcomings with the heuristic tracking algorithm which limits performance. Such as true multiresolution sinusoidal tracking, matching across a number of variables instead of frequency alone, and global optimization of the sinusoidal tracks. These limitations are discussed and addressed in this paper.

This paper begins by dividing the heuristic tracking algorithm into its separate operations. Then the proposed improvements to each operation are discussed in detail. The following section will demonstrate the performance of improved algorithm with simulated signals. The final section will outline the conclusions of this work.

2. Heuristic Tracking Algorithm

The heuristic tracking algorithm processes the estimates in an iterative fashion, typically progressing from beginning to end. For each iteration the algorithm can be broken into two key operations: fetch and match. The fetch operation selects which estimates and tracks should be considered for tracking in the current iteration. While the match operation selects which estimates should be appended to each evolving sinusoidal track. To do this the match operation uses a “matching criterion”, a concept taken from [5], which produces a coefficient indicating the suitability of a match between a particular estimate and track.

The remainder of this section will discuss each of these operations and the matching criterion individually, beginning with existing techniques and moving into the improved techniques proposed by this paper.

2.1 Fetching

The fetching of estimates and tracks in a single-subband tracking system is straight forward, as the windows of the STFT are uniform. Estimates come from the current STFT window, while tracks will be from the previous STFT window. However this is not the case for multiresolution sinusoidal analysis.

It has been found that multiresolution sinusoidal analysis drastically improves performance due to the time/frequency tradeoff being managed in individual subbands [13,14,15]. Historically multiresolution sinusoidal models have performed sinusoidal tracking independently within the individual subbands [8,11,16], as the window lengths of the STFT between subbands are no longer uniform. This artificially splits tracks as they attempt to cross subband boundaries.

This paper proposes a technique for fetching which allows the sinusoidal tracking algorithm to operate across subband boundaries. Essentially it works by extending the notion that at a given point in time an evolving track will expect a new estimate to be added, otherwise the track will die. The time, usually in samples, can be calculated from the last estimate in the track and the window overlap, as shown in (2.1) where L is the window length, r is the percentage of window overlap.

$$n = n_{i-1} + \frac{rL}{100} \tag{2.1}$$

For a given time instant, multiresolution fetching simply select tracks which expect a new estimate at this instant, and select estimates from all subbands with a time equal to the instant, as is demonstrated in Figure 2.1 below.

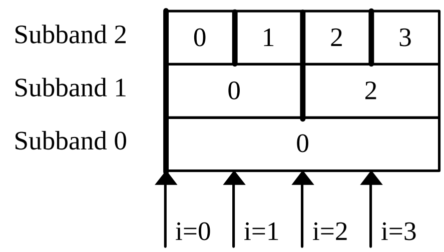

Figure 2.1 Multiresolution fetching of estimates, numbers in the blocks indicate the iteration the block is fetched.

An added advantage of performing multiresolution fetching is that the components which follow no longer need to consider subbands independently, reducing the complexity of their design and implementation.

2.2 Matching

Traditional heuristic matching iterates through each track according to increasing frequency, where the frequency of a track is determined from the last estimate in the track. For each track the most appropriate estimate, as judged by the matching criterion, is selecting as the “candidate match”. If no such candidate match exists then the track is completed.

The second step is to check if the candidate estimate is better matched to the next track, if this is not the case then the “definitive match” is declared from the “candidate match” and the track and estimate are removed from further consideration. If the candidate estimate is better matched to the next track, then the algorithm checks if the previous estimate can be matched to the track, if so this it is declared the “definitive match”.

Once all tracks have been considered, any unmatched tracks are completed, and any unmatched estimates create new tracks.

This approach provides a matching operation which is localised in frequency, with only a limited number of cases tested. This corresponds to incorrect matching of sinusoidal tracks with similar frequencies under certain circumstances.

A global matching operation has been devised which considers all cases globally across all variables considered by the matching criterion. It is an elegant solution with little increase in complexity than the traditional approach.

The first step of the global algorithm is to create a matrix which contains the coefficient produced by the matching criterion for each track and estimate combination. As will be discussed in the following section, the coefficient measures the suitability of a match between a given track and estimate, with a lower value representing increased suitability.

The global algorithm iteratively finds the minimum coefficient from the matrix, which is declared the "definitive match" and the row and column which represent the track and estimate are removed from the matrix. This process continues till either the matrix is empty or until the minimum coefficient is larger than a threshold set by the matching criterion. Any unmatched tracks are completed, and unmatched estimates become new tracks.

2.3 Matching Criterion

The matching criterion is used by the matching operation to determine how "close" or how suitable a match is between a track and estimate pair. The matching criterion produces a coefficient, k, which is the measure of this suitability. Lower coefficients are more suitable than higher coefficients, while a threshold value, k_{MAX}, is used to indicate when a match isn't suitable.

The matching criterion used in [1] uses frequency values from the estimate and the track to measure matching suitability. A set matching interval, Δ, is used to determine if the match is suitable (2.2).

$$\begin{aligned} k_{MAX} &= \Delta \\ k &= \left| f_e - f_{t,n} \right| \\ m(t,e) &= \begin{cases} k & \text{if } k < \Delta \\ k_{MAX} & \text{otherwise} \end{cases} \end{aligned} \tag{2.2}$$

The fixed matching interval worked sufficiently for narrow frequency band applications such as speech coding. However, for wideband audio a fixed matching interval isn't appropriate as a large matching interval is required to track high frequency tracks, but this increases the probability of matching sinusoidal tracks with noise estimates at lower frequencies.

To overcome this limitation it was proposed [11] that the matching interval should vary proportionally to the current frequency (2.3), accommodating the logarithmic nature of frequency. The threshold coefficient is set to some number larger than the product of the maximum frequency and the matching coefficient, b.

$$\begin{aligned} k_{MAX} &\geq \frac{bf_s}{2} \\ k &= \left| f_e - f_{t,n} \right| \\ m(t,e) &= \begin{cases} k & \text{if } k < bf_{t,n} \\ k_{MAX} & \text{otherwise} \end{cases} \end{aligned} \tag{2.3}$$

While this improves the performance of the matching criterion, an approach based off frequency values is still highly limited, due to the following: track trajectories cannot cross each other, tracking is still highly susceptible to the effects of noise estimates, and the tracks have a horizontal and erratic nature which is not consistent for slowly varying sinusoidal signals.

It is stated in [5] that the continuity of slopes is preferable to the continuity of values for sinusoidal tracking, and that a matching criterion which considers the slope, or the derivative, of the frequency values is beneficial. Such a technique was used in [12]. It has been found that a matching criterion which uses derivatives instead of values allows track trajectories to cross, produces smoother track trajectories, and is less susceptible to the effects of noise estimates.

The slope is estimated from successive frequency estimates, and then the matching coefficient can be calculated from the absolute value of the difference between the last slope of the track, and slope between the last frequency estimate of the track and the new estimate (2.4). The matching coefficient threshold must be larger than the maximum product of the maximum frequency difference and the matching interval coefficient, b.

$$\begin{aligned} k_{MAX} &\geq \frac{bf_s}{2} \\ k &= \left| \left(f_{t,n} - f_{t,n-1} \right) - \left(f_e - f_{t,n} \right) \right| \\ m(t,e) &= \begin{cases} k & \text{if } k < b\left| f_{t,n} - f_{t,n-1} \right| \\ k_{MAX} & \text{otherwise} \end{cases} \end{aligned} \tag{2.4}$$

Using frequency information only in the matching criterion makes matching susceptible to noise estimates which are close in frequency. To overcome this, the matching criterion needs to measure the suitability of a particular matching using more information than frequency alone.

A multivariable matching criterion may use amplitude and harmonicity information in addition to frequency to measure the suitability of a particular match. Amplitude information can be considered much the same way as frequency, with either a value or derivative approach. Harmonicity can utilize the progress of harmonically related tracks to determine which would be the most suitable estimate. Purnhagen has utilized phase to enable phase-locked sinusoidal tracking [17]. An investigation into the benefits of considering harmonicity information has not been conducted for this work as the combination of amplitude and frequency information proved to be beneficial.

Some implementations of the heuristic algorithm use multivariable matching in a post-processing fashion [18], The approach taken here is to use this information during sinusoidal tracking phase, as it is far more powerful due to the larger number of cases considered. This is similar to the approach taken with HMM/A tracking algorithms [2,4,5].

$$\begin{aligned} \Delta A_{diff}(t,e) &= \left|\left(A_{t,n} - A_{t,n-1}\right) - \left(A_e - A_{t,n}\right)\right| \\ \Delta f_{diff}(t,e) &= \left|\left(f_{t,n} - f_{t,n-1}\right) - \left(f_e - f_{t,n}\right)\right| \end{aligned} \quad (2.5)$$

The difficulty with a multivariable matching criterion is how to combine the information generated to create a sensible matching coefficient. It has been found that simple relationships in the A_{diff} - f_{diff} plane produce high performance matching criteria for both value and derivative information. The remainder of this section will only consider derivative information; it should be noted that value information can be used in a similar fashion.

A linear relationship was investigated, but was found to give too much emphasis to large amplitude and frequency mismatches, instead an inverse relationship (2.6) has proven effective as it penalizes these mismatches.

$$\begin{aligned} \Delta A_{diff} &= \frac{k}{\Delta f_{diff}} \\ k &= \Delta A_{diff} \times \Delta f_{diff} \\ k_{MAX} &= \infty \\ \Delta &= b\left|f_{t,n} - f_{t,n-1}\right| \\ m(t,e) &= \begin{cases} k & \Delta f_{diff}(t,e) < \Delta \\ k_{MAX} & otherwise \end{cases} \end{aligned} \quad (2.6)$$

3. Simulation Results

To measure the tracking algorithm accuracy in isolation a simulated environment was created. A sinusoidal track simulator was used to create a sequence of sinusoidal estimates and the expected tracks that these estimates should produce.

The track simulator begins by creating a number of random tracks using a "random walk" framework. Then "noisy" estimates are added at a given proportion, with the random frequency, amplitude and phase values.

The accuracy of a given tracking algorithm can then be determined by checking if all the tracks generated by the tracking algorithm from the estimates match the expected tracks from the simulator. If the generated and expected tracks match, then a correct match is declared, otherwise an incorrect match occurs. To keep the results comprehensible there is no distinction or grading of incorrect matches, just the boolean result is used. A number of iterations for a given tracking algorithm are used to determine the tracking algorithm's probability of correctly tracking the estimates.

The first set of experiments measured the accuracy of various matching criteria in a single subband system with independent fetching and local matching. The number of sinusoidal tracks created by the simulator was varied, as well as the matching coefficient. One thousand iterations were performed for each combination.

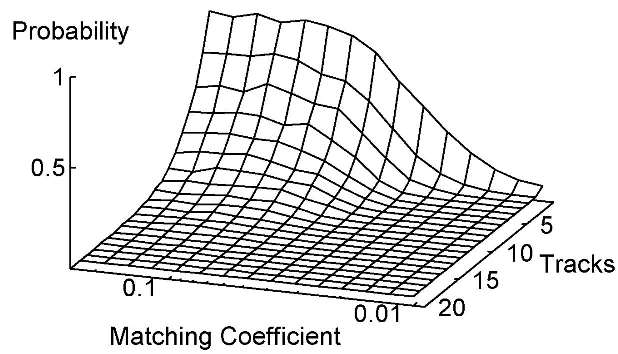

Figure 3.1 Tracking accuracy of the tracking algorithm using the frequency-matching criterion.

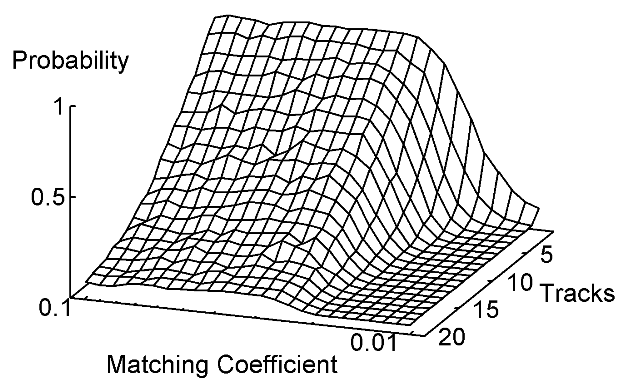

Figure 3.2 Tracking accuracy of the tracking algorithm using the frequency derivative matching criterion.

The results in Figures 3.1, 3.2, and 3.3 show that the multivariable matching criterion has better and more consistent accuracy than the other approaches for a large number of sinusoidal tracks. This is the case as the multivariable matching criterion has reduced susceptibility to noise estimates, is able to work with more tracks as it allows trajectories to cross and is more able to distinguish between tracks.

The next set of experiments measured the accuracy of the local and global matching operations. This was done within a single-subband system using the independent fetching and multivariable matching criterion, with $b = 0.03$.

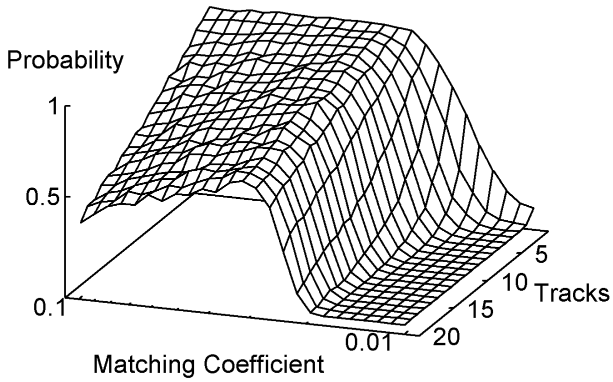

Figure 3.3 Tracking accuracy of the tracking algorithm using the Multivariable matching criterion.

Local matching works surprisingly well in most cases, but performance drops for high-density signals. To highlight the difference in accuracy between the two matching operations, the density of the simulated signal was increased by having a larger proportion of noise estimates and using a greater number of tracks.

The results plotted in Figure 3.4 show that as the density of the signal increases the global matching has increased accuracy compared to the local matching. This occurs because the global matching finds the true best match, as defined by the matching criterion, for each iteration.

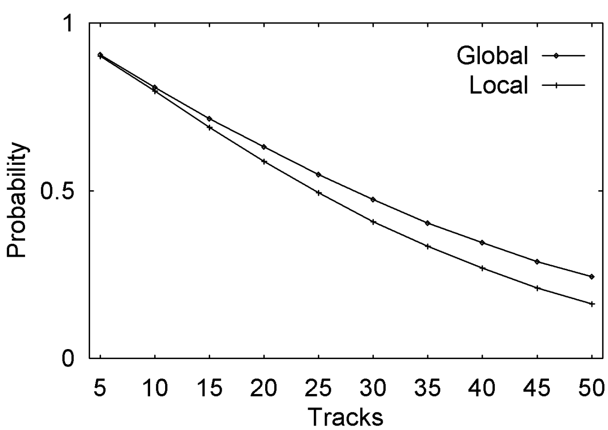

Figure 3.4 Tracking accuracy of the tracking algorithm using global matching versus local matching

The final set of simulation experiments measure the accuracy of the independent and multiresolution fetching operations. This was done with multiple-subband systems using the global matching and multivariable matching criterion, with $b = 0.03$.

The difference in tracking accuracy can clearly be seen by comparing Figure 3.5 to Figure 3.6. As the number of subbands increase the likelihood of a sinusoidal track crossing a subband boundary. With independent fetching this will not occur, whereas multiresolution can allow this boundary crossing to occur. This is why multiresolution fetching can provide a more consistent tracking accuracy as the number of subbands is increased. However the multiresolution fetching does not allow all subband crossings to occur, so there is still a slight trailing off as the number of subbands increases.

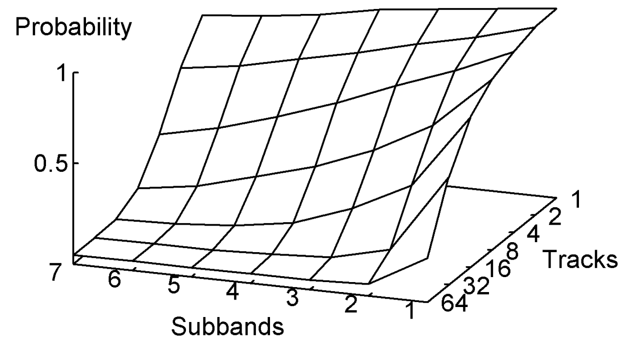

Figure 3.5 Tracking accuracy of the tracking algorithm using independent-subband fetching.

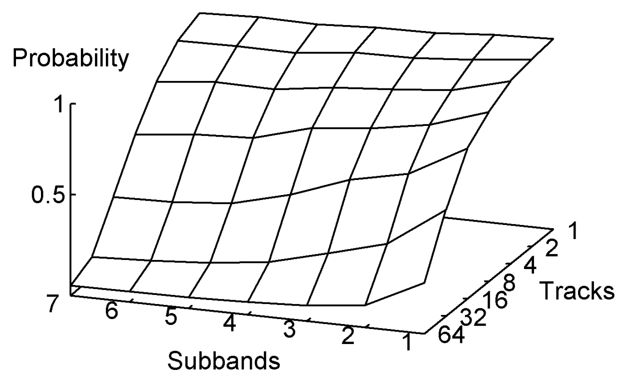

Figure 3.6 Tracking accuracy of the tracking algorithm using multiresolution fetching.

4. Conclusion

The results presented in this paper show that the techniques introduced in this paper increase the accuracy of the heuristic sinusoidal tracking algorithm substantially.

By improving the accuracy of sinusoidal tracking less sinusoidal tracks will be produced, enabling improved compression due to lower entropy, and improved auditory scene analysis due to the more precise modeling of audio objects within sinusoidal signals.

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, vol. 34, no. 4, pp. 744-754, Aug. 1986

[2] Steele, A., Streit, R. and Barrett, R., "Non-linear frequency line tracking algorithms", ASSP '89, Signal Processing Theories, Implementation and Applications, Apr. 1989.

[3] Streit, R. and Barrett, R., "Frequency line tracking using Hidden Markov Models", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 38, is. 4, pp. 586-598, Apr. 1990.

[4] Xie, X and Evans, R., "Multiple target tracking and multiple frequency line tracking using Hidden Markov Models", IEEE Transactions on Signal Processing, vol. 39, is. 12, pp. 2659-2676, Dec. 1991.

[5] Depalle, P., Garcia, G. and Rodet, X., "Tracking of partials for additive sound synthesis using hidden markov models", IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-93, vol. 1, pp. 225-238, Apr. 1993.

[6] Jensen, J., Heusdens, R. and Veenman, C., "Optimal time-differential encoding of sinusoidal model parameters".

[7] Adams, G. and Evans, R., "Neural networks for frequency line tracking", IEEE Transactions on Signal Processing, vol. 1, pp. 225-228, Apr. 1993.

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

[9] Ellis, D. and Rosenthal, D., "Mid-Level Representations for Computational Auditory Scene Analysis", International Joint Conference on Artificial Intelligence - Workshop on Computational Auditory Scene Analysis, Montreal, Canada, Aug. 1995.

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

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

[12] Virtanen, T., "Audio Signal Modeling with Sinusoids Plus Noise", Masters of Science Thesis, Tempere University of Technology, Finland, Aug. 2000.

[13] Goodwin, M., "Multiresolution sinusoidal modeling using adaptive segmentation", Proceedings of IEEE International Conference on Acoustics, Speech, and Signal, ICASSP '98, vol. 3, pp. 1525-1528, May 1998.

[14] Goodwin, M., "Adaptive Signal Models - Theory, Algorithms, and Audio Applications", PhD Thesis, University of California, Berkeley, 1997.

[15] Anderson, D., "Speech Analysis And Coding Using A Multi-Resolution Sinusoidal Transform", Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Atlanta, USA, pp. 1037-1040, 1996.

[16] 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.

[17] Purnhagen, H., "Parameter estimation and tracking for time-varying sinusoids", Proceedings of IEEE Benelux Workshop on Model based Processing and coding of Audio, Leuven, Belgium (MPCA-2002), Nov. 2002.

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