EXPLICIT ONSET MODELING OF SINUSOIDS USING TIME REASSIGNMENT

Aaron S. Master and Kyogu Lee

Stanford University
Center for Computer Research in Music and Acoustics
Stanford, CA 94305-8180 USA

ABSTRACT

We introduce a system that explicitly models onsets of sinusoidal signal components. To do this, the system uses time reassignment data to detect probable onsets. When an onset is detected, the reassigned data is used to estimate the precise location in time of the original onset, allowing synthesis of corresponding output. This is advantageous over conventional time reassignment which implicitly smears onsets. We demonstrate the efficacy of our system on synthetic and real test signals with sudden onsets.

1. INTRODUCTION

In signal processing for speech and music, the accurate modeling of sinusoidal and quasi-sinusoidal onsets is of great importance. Such onsets may occur when a voiced speech sound begins, a note is struck on a piano, or a choir sings a new pitch. This problem has received attention in the context of sinusoidal modeling [1, 2], a fundamentally important speech and music signal representation.

Historically, sinusoidal modeling has used frequency components parametrized by their frequency, amplitude, and sometimes phase. The traditional model has been expanded to include non-sinusoidal parameters such as bandlimited noise [3]. Such parameters are conventionally limited in time resolution by the frame rate and frame length of the short time Fourier transform (STFT) front end to the system.

Note onsets (and other events such as drum hits) require a representation with finer time resolution in order for synthesized outputs to sound psychoacoustically convincing. Transient modeling and time-frequency reassignment have partially met this need. Transients have been modeled in a variety of ways, including as time domain "peaks" detected as residual frequency domain oscillations [4], transform coded bandlimited time domain waveforms [5], and overlap-added residual time domain waveforms [6]. When onsets occur in input to these systems, they are implicitly modeled within the transient frameworks noted, or as a combination of transients, noise, and sinusoids.

Time-frequency reassignment [7] (see also [8] for an in-depth theoretical and historical review tracing back to [9]) may also provide a better framework for modeling onsets. This approach allows energy in one STFT frame to be reassigned to its local center of mass in time-frequency space, rather than to the fixed spectrogram grid. This produces greater time and frequency resolution in sinusoidal modeling parameter estimation than that dictated by the frame rate and frequency bin spacing. In [10], reassigned spectrogram data is used to more accurately model onsets. When the system detects an onset, it uses a parameter pruning technique to reduce the onset smearing that occurs in the time-frequency reassigned representation (which is itself already reducing the smearing of the conventional spectrogram). The authors point out that using accurate sinusoidal onset parameters allows a more consistent model than transient modeling.

Presently, we consider a further step towards accurate onset modeling using time reassignment data. In section two, we discuss this data specifically, reviewing how it may be understood as the center of time-mass of a signal component in time-frequency space. In section three, we introduce a model that uses time reassigned data to accurately estimate the time and amplitude of a sinusoidal signal onset. In section four, we apply the algorithm to several test signals and show that the system more accurately models onsets than do other systems. We conclude with a summary and brief discussion of future work.

2. TIME-FREQUENCY REASSIGNMENT

Time frequency reassignment, which has been used to model onsets more accurately, may be understood as finding the center of mass for a given point in time frequency space. Consider a spectrogram at a frame corresponding to time n and frequency bin k. That bin's magnitude is conventionally plotted at a specific place on a spectrogram. However, most of the energy in bin k at frame n for example may actually have occurred later in the frame than the center of the frame. And, for an example of the dual case, the frequency content of the signal may have been at a slightly higher frequency than that occupied by bin k. We could then say that the local center of time-frequency mass for the spectrogram point in frame n at bin k was higher in frequency and later in time. Visually, this would appear higher and to the right on a conventional (default Matlab) spectrogram plot.

It may be shown [7, 8] that for a signal $x(n)$, respective time $\hat{t}_{k,n}$ and frequency $\hat{\omega}_{k,n}$ reassigned locations are estimated relative to their original locations t_n and ω_k on the spectrogram grid as

$$\hat{t}_{k,n} = t_n - \Re\left\{\frac{X_{th;k,n}X^*_{h;k,n}}{|X_{h;k,n}|}\right\} \quad (1)$$

$$\hat{\omega}_{k,n} = \omega_k + \Im\left\{\frac{X_{dh;k,n}X^*_{h;k,n}}{|X_{h;k,n}|}\right\} \quad (2)$$

where $X_{h;k,n}$ represents the conventional STFT obtained using time domain window h, $X_{dh;k,n}$ represents the STFT obtained when windowing $x(n)$ with the time derivative of the window h, $X_{th;k,n}$ represents the STFT obtained when the window h is multiplied by the time variable t, and * indicates a complex conjugate.

Presently we are concerned specifically with the time reassigned data points $\hat{t}_{k,n}$ in the case of a sinusoidal onset. When a step-function style onset of a sinusoidal component occurs during the middle of a frame, its center of mass in time will be shifted towards future time. (In an analogous fashion, when a sinusoid "turns off," the center of mass of that bin with respect to time will be earlier in the frame.) As an introductory example, consider a step function onset of a sinusoid, such as that seen in figure 1. Consider a spectrogram analysis where a rectangular window of length 10 and hop size 4 is used. Let the onset occur such that the first nonzero sample is at sample 10 of a 10 point frame. There, the center of mass will be at sample 10. In the next frame centered 4 samples later, this same onset occurs at sample 6 of a 10 point frame. Hence, the center of mass for that frame will be at sample 8. If we use conventional time reassignment to relocate the frame mass at sample 8, we have a significant data representation problem: the onset occurs at sample 6, not sample 8. (Also, the amount of mass will be much greater in the second frame than the first, because more of the onset and thus more energy is contained in the second frame.) Similar problems exist as subsequent frames cover more and more of the onset signal.

However, it is a necessary consequence of the onset that the pattern of future time reassignments occurs. In the next section, we introduce a model that exploits this fact to selectively and explicitly model onsets.

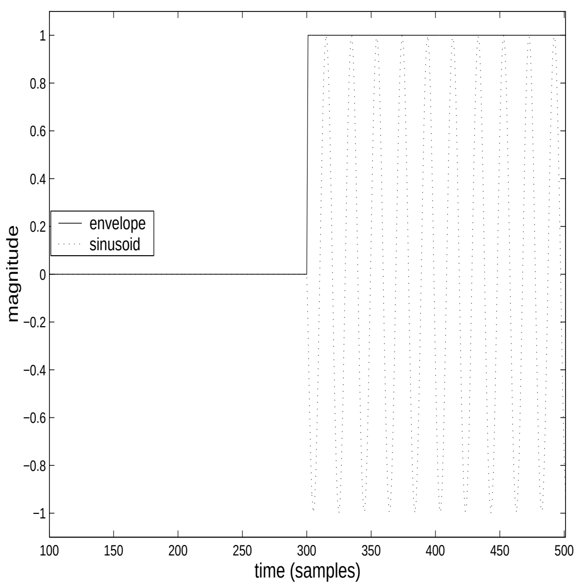

Fig. 1. Sinusoidal test signal and its envelope.

3. EXPLICIT ONSET MODEL AND PARAMETER ESTIMATION

Considering onset cases such as in the previous section's example, we propose the explicit onset model, which treats patterns of repeated future assignments as being indicative of a step-function style onset. When an onset is detected this way, we backsolve for the onset location and amplitude that creates the center of mass data revealed by the time reassignment data. To ensure reliability of the estimates, we employ a heuristic measuring consistency between the data generated by frames containing the supposed onset.

To backsolve for the onset time and amplitude given the time reassignment data, we use a table lookup method. To create the table, we synthesize onset signals (one at each possible sample in the window) under the time window h used in the system, and record their respective centers of mass in time as determined by time reassignment. Then, when a possible onset is detected, we can look up the nearest time reassigned data to estimate an onset time to the nearest time domain sample. Because the system operates in a Fourier space, only one table is needed for all frequencies.

The following outlines how we apply the model and heuristic.

1. Apply spectrogram and time reassignment analysis to an input signal, where the frame hop size is (M/P), M is the length of the window, and $P \geq 4$.
2. When $P - 1$ future reassignments in a row are found for a particular frequency bin, consider a potential onset to be detected.
3. For each of the P reassignments into the future, use table lookup to estimate the location in time of the onset.
4. Determine if the standard deviation of the P onset time estimates (in absolute time units) is below a threshold T_1.
   (a) If so, consider the data to indicate an onset, and use the explicit onset estimates as the detected envelope function of the sinusoid. (Optionally, use the mean of the onset estimates as the onset time, setting the envelope to zero before the onset.)
   (b) If not, use conventional time-frequency reassignment to represent the data.

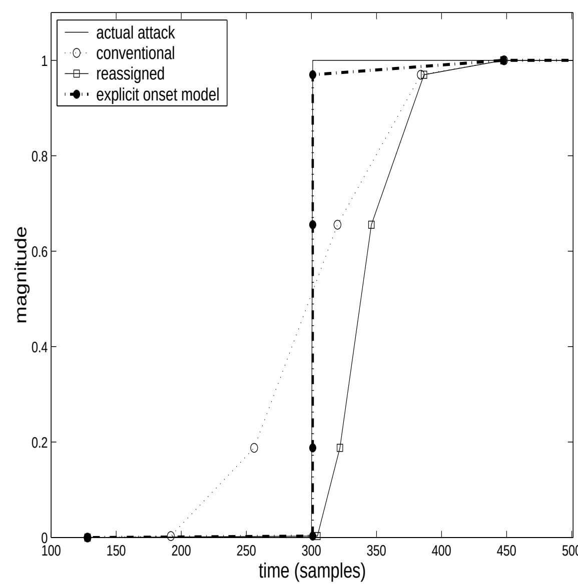

Fig. 2. Sinusoidal test signal: original envelope and detected envelopes.

4. RESULTS AND DISCUSSION

We now apply this model to a variety of test signals with onsets. We begin with the original signal in figure 1, a sinusoid of frequency 416 Hz assuming a sampling rate of 8192 Hz. We use a Hann window of length $M = 256$ and a hop size of $M/4 = 64$ (so $P = 4$). Figure 2 shows the original envelope of the sinusoid, and the estimation of the amplitude parameters thereof using (1) conventional sinusoidal modeling parameter estimation ("SMS") (2) time reassignment and (3) the explicit onset model. For the explicit onset model, we show all of the time-reassigned points rather than their mean time value. It is clear from the figure that the explicit onset model better fits the original data than the other models. We note that this is true even if a pruning technique such as in [10] were applied to the time-frequency reassigned data shown.

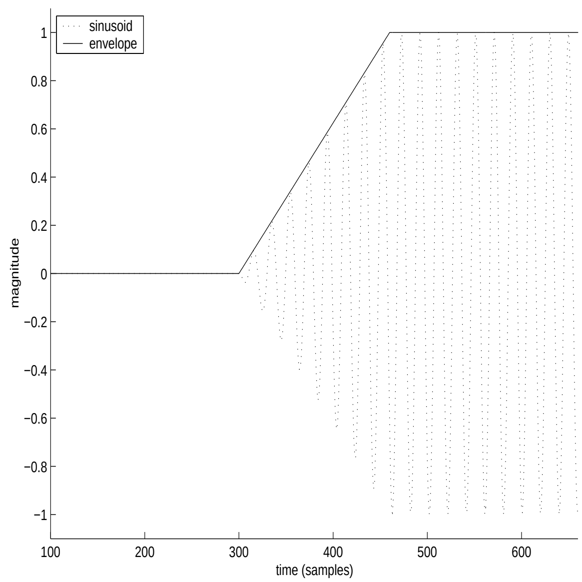

Fig. 3. Linear onset sinusoidal test signal and envelope.

We also applied our model and the others to signals with an abrupt linear onset. A signal whose onset occurs linearly over 160 samples is shown in figure 3. We tested our model on a set of such signals to investigate how far we can strain the step-function style onset assumption before our model becomes suboptimal. The table below shows our results. In the first column, we show the amount of time given for the linear onset to occur in the test signal. The second column shows the standard deviation in time onset estimates generated by our system. The remaining three columns show the mean squared error, in samples, of each of the three models. Because the amplitude data comes from the spectrogram and is consistent across all three models, we calculate these errors relative to where in time the detected amplitudes actually occur on the input signal envelope. (This also allows us to consider separately any interpolation used later in resynthesis techniques.) We see that our model is more accurate than the others even when the linear onset time exceeds the hop size of 64 samples. Importantly, we also see that standard deviations in the onset estimates greater than 41 samples (for this setup) tend to indicate that the explicit onset model is not appropriate. We thus select T_1 = 41 samples in our algorithm. To show what happens to our model at the breaking point (for the 160 sample linear onset case), we include the envelope parameter estimates in figure 4.

onset dur.	σ_{EOM}	E_{SMS}	E_{TFR}	E_{EOM}
1	0	8553	6250	**4293**
16	1.64	7248	4933	**3477**
32	4.97	6112	3803	**2722**
64	13.59	7355	5413	**3803**
128	26.28	6325	4673	**4463**
160	40.24	4421	2780	**2529**
192	45.24	5534	**4091**	4267
256	65.59	4919	**3635**	4213

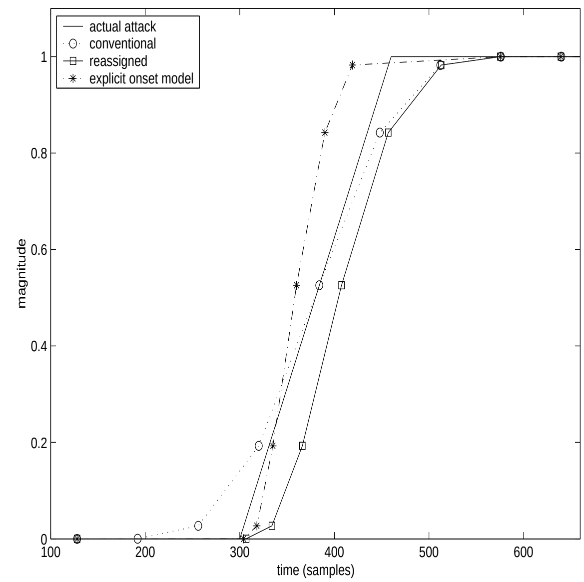

Fig. 4. Linear onset sinusoidal test signal: original envelope and detected envelopes.

We also apply our system to a real world glockenspiel signal. In figure 5, we show the signal with the various systems' envelopes for the largest amplitude sinusoid detected. (The negative amplitude signal values are cropped for plot clarity.) We see that the explicit onset model traces the envelope of the signal most closely. All three envelopes are below the maximum of the signal because the signal contains multiple sinusoidal components.

For output, any standard sinusoidal synthesis technique may be used given the parameters detected using our system, provided that the sharp envelope characteristics of the onsets are preserved as appropriate.

5. SUMMARY AND FUTURE DIRECTIONS

We have presented a system that uses time reassignment data to detect step-function style onsets and estimate their location in time. We specified a heuristic based on data consistency for use in the system, and provided data showing improved modeling of synthetic and real onsets. Our model incurs little additional computational expense relative to time-frequency reassignment, via its table lookup and consistency test. The technique discussed here also applies to the "turn-off" of sinusoidal components, though the post-masking phenomenon (see, e.g. [11]) makes such detection less psychoacoustically necessary in many applications. The

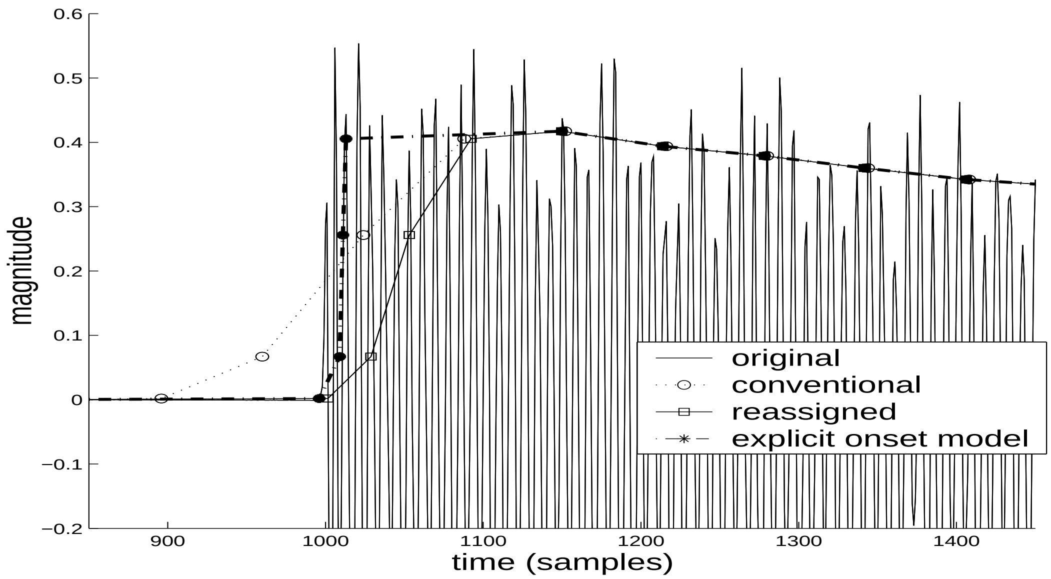

Fig. 5. Glockenspiel test signal with various model envelopes

theory and application remain identical to that presented above, with the time axis reversed.

In future work, we will consider expansions to the onset model, employing psychoacoustic criteria to determine the accuracy required.

6. REFERENCES

[1] Julius O. Smith and Xavier Serra, "PARSHL: A program for the analysis/synthesis of inharmonic sounds based on a sinusoidal representation," in *Int Computer Music Conf*. 1987, Computer Music Association, Also available as Stanford Music Department Technical Report STAN–M–43.

[2] R. J. McAulay and T. F. Quatieri, "Speech analysis-synthesis based on a sinusoidal representation," Tech. Rep. 693, Lincoln Laboratory, MIT, 1985.

[3] Xavier Serra and Julius O. Smith, "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, Winter 1990.

[4] Tony S. Verma and Theresa H. Y. Meng, "An analysis / synthesis tool for transient signals that allows a flexible sines + transients + noise model for audio," in *ICASSP*, 1998.

[5] Scott Nathan Levine, *Audio Representations for Data Compression and Compressed Domain Processing*, Ph.D. thesis, Electrical Engineering Dept., Stanford University (CCRMA), December 1998.

[6] Geoffrey Peeters and Xavier Rodet, "Sinola: A new analysis/synthesis method using spectrum peak shape distortion, phase and reassigned spectrum," in *International Computer Music Conference*, 1999, Beijing.

[7] Francois Auger and Patrick Flandarin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Transactions on Speech and Audio Processing*, vol. 43, no. 5, pp. 1068–1089, May 1995.

[8] Stephen W. Hainsworth and Malcolm D. Macleod, "Time frequency reassignment: A review and analysis," Tech. Rep. CUED/F-INFENG/TR.459, Cambridge University Engineering Department, 2002.

[9] K. Kodera, C. de Villedary, and R. Genderin, "Analysis of time-varying signals with small bt values," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 64–76, February 1978.

[10] Kelly Fitz and Lippold Haken, "On the use of time-frequency reassignment in additive sound modeling," *Journal of the Audio Engineering Society*, vol. 50, no. 11, pp. 879–893, November 2002.

[11] E. Zwicker and H. Fastl, *Psychoacoustics*, Springer-Verlag, Berlin, Germany, 1990.