# A Comparative Evaluation of Techniques for Single-Frame Discrimination of Nonstationary Sinusoids

Jeremy J. Wells and Damian T. Murphy, *Member, IEEE*

***Abstract*—Many spectral analysis and modification techniques require the separation of sinusoidal from nonsinusoidal signal components of a Fourier spectrum. Techniques exist for the estimation of the parameters of nonstationary sinusoids, and for discriminating these from other components, within a single Fourier frame. We present a comparative study of five methods for sinusoidal discrimination, considering their effectiveness and their computational cost.**

***Index Terms*—Audio coding, music, spectral analysis.**

## I. Introduction

SPECTRAL modeling, when the term is used in its broadest sense, is the description or simulation of signals and systems based on their frequency content. The more specific meaning of the term, adopted here, is the description of signals as the combination of processes which have a straightforward, or intuitive representation in the frequency domain. A well-known example of this for the modeling of audio is Spectral Modeling Synthesis (SMS), which represents signals as the combination of time varying sinusoids (for narrowband spectral components) and noise modified by time varying filters (for broadband components) [1].

Traditionally, systems such as SMS have performed their analysis "offline": generation of model parameters commences only after the whole sound has been captured. Subsequent modification of the model and resynthesis from it can be done online (in real-time) or offline depending on the application, the modification type and the processing power of the hardware. This means that the whole analysis-modification-synthesis cycle is offline, even if some individual parts of it are not. While this may be appropriate in some situations it may well not be ideally suited to all. Many audio processors (such as dynamics compressors, time-domain processors, etc.) have traditionally been used online with near-zero or minimal latency between audio input and output. One of the aspects of the analysis which benefits from the acquisition of the whole sound is the separation of sinusoidal from nonsinusoidal (residual) components. The work presented here is motivated by the desire to develop a sinusoidal-plus-residual analysis system for audio that can be used in (quasi) real-time. The qualifier "quasi" is used here since there must be some latency in a spectral analysis system which is causal. In fact, what is discussed here is a frame-by-frame system within which audio for a single analysis frame is analyzed, transformed and resynthesized without the need for subsequent frames to be acquired. Although previous frames are available to the discriminator such frame-by-frame modeling will at least occasionally require single frame discrimination between sinusoidal and residual components, after transients for example, where prior frames will not be relevant to the current frame. If the quality of single-frame processing is such that it obviates the need for the consideration of multiple previous frames then this would make it highly desirable in many applications.

Presented here is a comparison between different methods for such single-frame discrimination. Such methods are different from those presented in, for example [1] and [20], in that they do not rely on the inter-frame behavior of parameters (e.g., phase) but only make use of parameters estimated within a single frame. Such methods for single-frame discrimination include the Thomson F-test [16], the non-stationary spectral cross-correlation [14] (which is a modification of the discriminator in [21]) and the reassignment-derived mixed partial derivative (MPD) of the phase [15]. These methods are described in detail in Section III of this paper. The comparison between methods in this paper is done in the context of a high-accuracy frame-by-frame sinusoidal-plus-residual spectral modeling system designed for creative sound transformation. Following Fourier analysis of a frame and identification of local maxima in the magnitude spectrum, parameters for amplitude, phase, frequency, amplitude change, and frequency change are determined using a system that was described in [2] and [3]. This system has been chosen for this study since it has been tested with highly nonstationary sinusoids; however, it should be noted that there are alternative methods, such as those described in [4]–[6] and which could be used and such ranges of nonstationarity are not common in typical nonsynthetic audio signals. The term "stationary" here refers to nonconstant amplitude and/or frequency during a Fourier analysis frame. An overview of this high-accuracy analysis system is given in Section II of this paper. Once these parameters have been estimated by the system, the behavior of the components represented by local maxima is investigated to determine how "sinusoidal" the underlying component is. Five different methods for doing this are

Manuscript received December 15, 2008; revised December 03, 2009. Current version published February 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Bertrand David.

The authors are with the Audio Lab, Intelligent Systems Group, Department of Electronics, University of York, Heslington, York, YO10 5DD, U.K. (e-mail: jjw100@ohm.york.ac.uk; dtm3@ohm.york.ac.uk).

Digital Object Identifier 10.1109/TASL.2009.2039088

described in Section III. Section IV gives details of the methods used to compare the efficacy of each of these methods and their computational cost. Section V presents results obtained from using these methods for each of the sinusoidal discriminators described in Section III. Finally, conclusions are drawn and recommendations made in Section VI.

The main aim of this paper is to present a comparison between new and established methods for differentiating between sinusoidal (both stationary and nonstationary) and nonsinusoidal components in a mixed spectrum, in particular to compare two novel techniques recently developed by the authors (discriminators A and B in Section III), with established methods (some with novel adaptations for the purposes of this work). The work presented here is significantly expanded and enhanced compared to the survey presented in [7] and is the first large-scale study of this kind of which the authors are aware. Although the discriminators described here were developed, or adapted, for a specific audio modeling system the findings have relevance to any single-frame sinusoidal discrimination application.

## II. Frame-by-Frame Spectral Modeling of Audio

The SMS system described in [1] performs its analysis of sounds offline. The acquisition of the entire signal prior to estimation of the model parameters is advantageous when discriminating between the two different types of model component, which are referred to as sinusoidal (or deterministic) and nonsinusoidal (or stochastic or residual). Since sinusoidal is taken to mean long-term components (i.e., those which exist over a number of frames) which can be modeled as slowly time-varying sinusoids, their identification benefits from analysis which can take place across multiple frames, both in the past and the future. Furthermore, sounds generated by physical, vibrating systems tend to be characterized by an onset, during which the excitation energy changes rapidly, followed by a steady-state period where the excitation energy is nonzero but near-constant, and finally an exponential decay period once the excitation energy falls to zero and the vibrational energy is damped to zero. Identification of long-term, slowly varying sinusoidal components is best performed during the last two of these stages rather than at the sound onset [8].

Such a system offers a robust analysis and classification of sinusoidal components into "active" (requiring an oscillator for resynthesis in the current frame) and "sleeping" (not active in this frame, but active in a previous frame and in a subsequent frame) whose parameters are interpolated between active frames. However, since this system cannot produce its model of sound causally it can only be used for synthesis of existing models in real-time. A system which can model, as well as transform and synthesize from that model, with as close to real-time performance as possible could be used for live sound transformation as well as synthesis. For example such a low-latency implementation of this modeling paradigm might be successfully employed in the traditional real-time studio effects unit. The closer to real-time the whole analysis-modeling-synthesis process is, the more amenable it will be to live sound transformation. The closest to real-time that is possible is a latency of one analysis frame and it is for this reason that single-frame modeling techniques are investigated in this paper.

In this section, a brief overview is given of a nonstationary frame by frame spectral modeling system which is designed to address this. For a more detailed discussion the reader is referred to [2]. The output of the parameter estimation part of this system is used by each of the sinusoidal discriminators that are compared in this paper. This analysis method is used here since it offers high accuracy even where there is a high degree of nonstationarity [2]. Sinusoids are assumed to be nonstationary and are synthesized using oscillators that do not overlap in time (i.e., the output is not overlap-added). Since they are synthesized on a frame-by-frame basis, this requires high-accuracy estimates of amplitude and frequency and how they change in a single frame. Without this accuracy, there is the danger of severe discontinuities in amplitude and frequency trajectories at synthesis frame boundaries.

The nonstationary sinusoids discussed in this paper are globally assumed to be of the form

$$s(t) = A(t)\sin\left(\int_{\tau=0}^{\tau=t} 2\pi f(\tau)d\tau + \varphi\right). \tag{1}$$

The model adopted here approximates this global function by a succession of local functions. A local nonstationary sinusoidal function, for a single frame centered on the global time $\mathrm{t}_0$ is given by

$$s(\tau) = \overline{A}10^{\alpha\tau}\sin\left(\phi + 2\pi\left(\overline{f}\tau + \frac{\Delta f\tau^2}{2T}\right)\right) \tag{2}$$

where $t = t - t_0$ (in seconds), $\overline{A}$ and $\overline{f}$ are the amplitude and frequency at $\mathrm{t}_0$, $\varphi$ is the starting phase of the sinusoid, $\Delta f$ is the frequency change within the analysis window, and $\alpha$ is given by

$$\alpha = \frac{\Delta A}{20T} \tag{3}$$

where $T$ is the length of the window (in seconds) and $\Delta A$ is the amplitude change within the window. Since synthesis frames do not overlap (although analysis frames do overlap) for this model the amplitude is piecewise exponential, the frequency is piecewise linear and the phase is piecewise quadratic. The piecewise nature of these trajectories is inherent in the frame-by-frame approach. In these equations, $\Delta A$ and $\Delta f$ are expressed as *rates* of change, subsequently in this paper we refer to these quantities as *amounts* of change which occur within a single analysis frame.

The analysis begins with the discrete Fourier transform (DFT) for which we use the following notation.

| | |
|---|---|
| $X_{h,k,p}$ | The $k$th bin of the Fourier transform of the $p$th frame of the input sequence weighted by the window function $h$. |
| $F_s$ | Sampling rate of the input sequence. |
| $\mu$ | Hop size (the distance in samples between adjacent DFT frames). |
| $N_{\mathrm{frame}}$ | Length of the input frame in samples. |
| $N$ | The (zero-padded) size of the DFT, which is performed using the FFT. |

Many methods exist for the estimation of the mean instantaneous frequency of components in the Fourier domain. These include measuring the phase difference between successive frames, interpolation of the magnitude spectrum, and time–frequency reassignment [9]. Reassignment is used in this system since estimates are obtained from a single analysis frame and it provides better estimates than other single-frame methods [9]. Time–frequency reassignment estimates the exact centroid of a component energy in terms of radial frequency $(\omega)$ and time $(t)$, in seconds, by taking two additional DFTs per frame

$$\omega_{k,p} = B\left(k - \Im\left\{\frac{X_{dh,k,p}}{X_{h,k,p}}\right\}\right) \tag{4}$$

$$t_{k,p} = \left(\left\lceil\frac{N}{2}\right\rceil + p\mu + \Re\left\{\frac{X_{th,k,p}}{X_{h,k,p}}\right\}\right) \Big/ F_s \tag{5}$$

where $B$ is the width of an analysis bin (as radial frequency), $dh$ is the first derivative (or first order difference if no analytical solution for the derivative exists) of $h$ and $th$ is the result to applying a linear ramp to $h$ ($= 0$ in the center of the frame) [10]. $\omega_{k,p}$ is a measure of the mean instantaneous radial frequency during the frame.

The parameters of nonstationarity, $\Delta A$ and $\Delta f$, are calculated using a form of phase distortion analysis (PDA) called reassignment distortion analysis (RDA). PDA was first described in [11] and RDA in [2]. Particular attention was paid in the latter work to reducing the interdependency of the $\Delta A$ and $\Delta f$ estimates, producing much greater accuracy over a wider range of values. These two parameters are estimated by least squares fitting of a cubic polynomial to the time offset data [the second term in (5)] around a magnitude peak. The nonstationary estimates can also be used to improve the accuracy of those for $A$ and $f$ as described in [3]. The phase bias caused by nonlinearity is corrected using 2-D table look-up. A reasonable compromise between input–output latency and time–frequency resolution for this system, when operating on signals that have been sampled at 44.1 kHz, is an analysis frame size of 1025 samples, zero-padded to 8192. The Hann window is used to taper analysis frames since this offers a good trade-off between main-lobe width and spectral leakage, moreover its side lobes decrease in level monotonically as the distance from the main lobe increases—a useful property when attempting to identify spectral peaks [12]. Although RDA is described in detail in [2] and [3] a brief overview of the algorithm is given here. For each input frame a reassigned DFT analysis is performed. Estimation of the parameters for a single magnitude peak is performed in the following sequence:

1) Use magnitude of peak bin as first estimate of mean instantaneous amplitude, $\overline{A}$.
2) Obtain first estimate of mean instantaneous frequency, $\overline{f}$, using (3).
3) Fit a second-order polynomial to the time-reassignment offset, $\Re\{X_{th,k,p}/X_{h,k,p}\}$, for the peak bin and four of its neighbors $p = P, P\pm2, P\pm4$, where $P$ is the index of the peak bin. $P\pm1, P\pm3$ are not used for fitting the polynomial since they are subsequently used for variance testing for one of the discriminators (A, described in the following section). The polynomial fitting is achieved by minimization of the least squares error between a polynomial

$$y = a_2x^2 + a_1x + a_0 \tag{6}$$

where $y = \Re\{X_{th,k,p}/X_{h,k,p}\}$ for $p = P, P\pm2, P\pm4$, and $x = pB$. The number of points taken for fitting (five) is different to that taken in [2] and [3] (three), but apart from this difference the estimation method is the same.
4) Use the parameters $a_0$ and $a_1$ of this polynomial to obtain estimates of $\Delta A$ and $\Delta f$ using iterative 2-D table look-up [2].
5) Use these parameters to improve estimates of $\overline{A}$ and $\overline{f}$ [2].
6) Use the estimates of $\Delta A$ and $\Delta f$ to estimate the phase bias using 2-D table look-up and subtract this from the measured value to give the corrected phase.

All of the 2-D table look-ups in this sequence use data from the peak bin. Deviation of the energy peak from the center of the bin does introduce bias in the values obtained via look up (in the same way that magnitude estimates are biased by deviation of a component from the center of a bin in a standard DFT [12]); however, these biases are small except where there is a very high degree of nonstationarity and are reduced by the use of a high zero-padding factor. A detailed study of these biases will the subject of future work.

## III. Single-Frame Sinusoidal Discrimination Methods

Once the parameters for each local maximum in the magnitude spectrum have been estimated the next step is to decide which component type (sinusoidal or nonsinusoidal) each maximum represents. This is the purpose of the discriminators which are evaluated in this paper. All of them have the following initial steps in common.

1) Perform a zero-phase, 8× zero-padded FFT on a 1025 point Hann-windowed signal.
2) Search the magnitude spectrum for peaks. This is defined here as bins where the magnitude is higher than that of the four neighbors on either side. A range of four bins either side of a peak in an 8× zero-padded spectrum approximately corresponds to one nonzero-padded analysis bin and, therefore, the underlying resolution of the DFT analysis (over which the zero-padded magnitude behavior would be expected to match that of an underlying sinusoidal peak).
3) Discard peaks where the reassigned frequency falls a fixed distance beyond the edges of the bin. This distance depends on the range of $\Delta f$ considered valid for a nonstationary sinusoid. For a stationary sinusoid, if the estimated instantaneous frequency lies outside of the bin it is not due to a sinusoid within it. Here the biasing effect of the range of nonstationarities considered can produce frequency estimation errors of nearly 14 bins [2]. Therefore, at this stage of the analysis a bin is only rejected due to contamination if the reassigned frequency is a distance of 14 or more bins away from the center frequency of the peak bin. Once biasing effects have been accounted for this distance can be reduced [3].

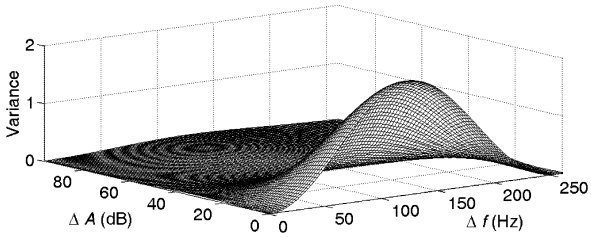


Fig. 1. Variance of second-order polynomial fit across a sinusoidal peak as a function of the modulus of the intra-frame amplitude and frequency change. The frame size is 1025, zero-padded to 8192. The sample rate is 44.1 kHz.

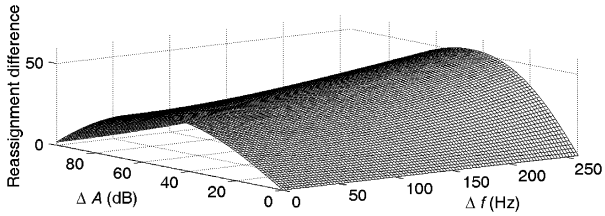


Fig. 2. Difference between reassignment estimates for group delay at a sinusoidal peak as a function of the modulus of the intra-frame amplitude and frequency change. The frame size is 1025, zero-padded to 8192. The sample rate is 44.1 kHz.

4) For each retained peak find the parameters $\Delta A$ and $\Delta f$ followed by $A$, $\varphi$, and $f$ using the high-accuracy technique described in [2] (although these latter two parameters are not needed by all of the discriminators).
5) Classify the retained peaks using the discriminator under test (DUT). Each of these discriminators compares a measured value of "sinusoidality" with a threshold. In testing, each of these thresholds is varied in order to produce the receiver operating characteristics (ROC) curves shown in Section V.

The rest of this section describes each of the sinusoidal discriminators considered in this survey (there are five in total). Hereafter, the discriminators are referred to by the letters in the following subsection headings (e.g., the time reassignment "goodness of fit" discriminator is referred to as discriminator A).

### A. Time Reassignment "Goodness of Fit" Measure

This method builds directly on the parameter estimation system described in [2], where a second-order polynomial is fitted to time-reassignment points around a peak. By taking different points around the peak there will be some that do not lie on the curve described by the polynomial if they do not all fit the same parabola. The points taken are $p = P \pm 1, P \pm 3$ so that they lie symmetrically around the peak but are not the same as those used to compute the coefficients of (6). The discriminator compares the variance of the measured data for a peak with that of a sinusoid which has the same parameters as those estimated from the peak. If there is little difference in variance then the indication is that the underlying component is a sinusoid. If the variance is large then this indicates that the component is not behaving as a sinusoid would be expected to do.

Fig. 1 shows the variance of the polynomial fit, as measured for a sinusoid with fixed values of $\varphi$, $A$ and mean instantaneous frequency (the values computed for the lookup table are all computed for a sinusoid whose mean instantaneous frequency is exactly at the center of an analysis bin), as a function of $\Delta A$ and $\Delta f$. The variance does not change with amplitude but it does vary to a small extent with distance between the mean instantaneous frequency and the center of the peak bin. This variation is significantly reduced by the use of a high zero-padding factor ($8\times$) and is ignored here so that just one look-up table is used requiring only a modest computational effort. Note that the variance does not increase monotonically with $\Delta f$. In fact, it returns to zero when $\Delta f$ reaches the maximum value for which its own estimator is monotonically increasing [3].

A single $100 \times 100$ element array is used to store the expected variances and values are obtained on a finer grid by linear 2-D interpolation. The discriminator calculates the modulus of the difference between the expected and measured variance and compares this with a threshold value. If the variance difference is lower than the threshold the component is classified as a sinusoid, if it is higher than the threshold it is classified as a nonsinusoid.

### B. Amplitude and Phase Reassignment Estimate Difference

The reassignment method described in [10] specifically relates to reassignment performed by differentiating with respect to phase. It is possible to obtain alternative reassignment measures by differentiating with respect to amplitude. "Phase" reassignment and "amplitude" reassignment measures will be close where the underlying component is a stationary sinusoid but will exhibit differences where it is not, except where a Gaussian window is used [13]. For a Gaussian window the measures will be identical since the Fourier transform of a Gaussian is also a Gaussian. A detailed study of amplitude reassignment can be found in [13].

For this method the discriminator compares the expected difference between amplitude and phase time-reassignment measures for the estimated values of $\Delta A$ and $\Delta f$ in the peak bin. As for the previous discriminator, the estimates of $\Delta A$ and $\Delta f$ are obtained using the high-accuracy RDA method. Fig. 2 shows how the difference in amplitude and phase estimates of the time-reassignment offset varies with $\Delta A$ and $\Delta f$. This "expected" difference in estimates, for a single bin $k$, is obtained from

$$D_{k,p} = \Re\left(\frac{X_{th,k,p}}{X_{h,k,p}}\right) - \left(-\frac{N}{\pi}\Re\left(\frac{X_{dh,k,p}}{X_{h,k,p}}\right)\right) \qquad (7)$$

In this case, the standard window is the Hann window, and time and frequency ramps are applied to this to produce the other two window types. The second term on the right-hand side of (7) is the standard phase reassignment estimate of the deviation (in samples) of the group delay from the center of the analysis frame, the second term is the estimation of group delay obtained by amplitude reassignment as given in [13]. Time-reassignment (group delay) measures were chosen over frequency-reassignment because they exhibit the least variation with frequency deviation from the center of an analysis bin [3]. Again, the values

measured for the look-up table are for a component that is exactly in the center of an analysis bin and variations due to a component not being at the center of an analysis bin are ignored.

As for the previous discriminator, a $100 \times 100$ array and 2-D linear interpolation is used to determine expected values of the reassignment difference given the estimates obtained for $\Delta A$ and $\Delta f$. The modulus of the difference between the actual and expected reassignment differences is calculated by the discriminator and compared with a threshold. A component that produces a value above the threshold is classified as a nonsinusoid, a component that produces a value lower than the threshold is classified as a sinusoid.

### C. Nonstationary Spectral Cross-Correlation

A method for testing the sinusoidality of a peak in the magnitude spectrum by cross-correlation in the Fourier domain is proposed and examined in [14]. The correlation is performed between a range of bins around the peak and the corresponding bins of a peak which results when a sinusoid is synthesized from the estimated parameters (including $\Delta A$ and $\Delta f$). In that work, the correlation method is evaluated for frequency and amplitude changes on a "highest score" basis: for each frame a certain proportion (10%) of the spectral peaks with the highest correlation measure are considered to be sinusoidal. In the current study, as for the previous discriminators, a threshold is used to determine (non-) sinusoidality.

$\Delta A$ and $\Delta f$ then $A$ and $\overline{f}$ are estimated, and a sinusoid with these estimated parameters is synthesized. The spectrum of this sinusoid is then compared with that of the actual spectral peak using (8), where $m$ is the number of zero-padded bins taken either side of the peak, $H(f)$ is the actual spectrum, and $W(f)$ is the spectrum of the sinusoid synthesized from the estimates. The value of $m$ used in this study is 8 and the frame is 8 times zero-padded. Equation (8) differs from that given in [10] in that here components are normalized according to the total energy measured across all $2m+1$ bins rather than the energy of the peak bin. This gives a range of correlation values of 0.0 to 1.0

$$\Gamma_k = \frac{\left| \sum_{\kappa=k-m}^{\kappa=k+m} H(\kappa) W^*(\kappa) \right|}{\sqrt{\sum_{\kappa=k-m}^{\kappa=k+m} |H(\kappa)|^2 \sum_{\kappa=m-B}^{\kappa=m+B} |W(\kappa)|^2}}. \tag{8}$$

As for the previous discriminators, a threshold is applied to this value: a value above the threshold is classified as a sinusoid, a value below is classified as a nonsinusoid.

### D. Reassigned Mixed Partial Derivative

In [15], the mixed partial derivative (MPD) of the phase of the Fourier transform is defined as

$$\frac{\partial^2 \phi}{\partial t \partial \omega}(X_k) \equiv \frac{\partial^2 \phi}{\partial \omega \partial t}(X_k). \tag{9}$$

This can be interpreted, considering the left side of (9), as the rate of change of measured mean instantaneous frequency (first derivative of phase) as the frequency axis of the transform is traversed. Where there is little change in mean instantaneous frequency across a wide range of the transform frequency spectrum this indicates the presence of a dominant component which is highly localized in frequency. There is a dual interpretation, considering the right-hand side of (6) which is the rate of change of the deviation of the group delay from the center of an analysis frame. By performing a fourth DFT (in addition to the three time–frequency reassignment DFTs) using a frequency-, then time-ramped version of the standard window the MPD can be estimated from

$$\Re\left\{\frac{X_{tdh,k,p}}{X_{h,k,p}}\right\} - \Re\left\{\frac{X_{th,k,p} X_{dh,k,p}}{(X_{h,k,p})^2}\right\} \tag{10}$$

where $tdh$ is the result of applying a zero-centered time-ramp to [15].

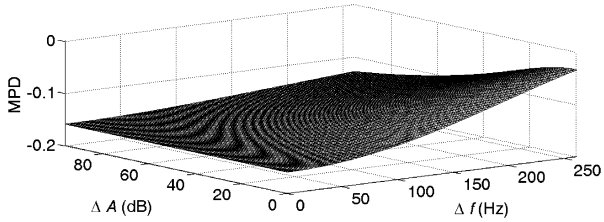


Fig. 3. Mixed partial derivative (MPD) as a function of intra-frame amplitude and frequency change. The frame size is 1025, zero-padded to 8192. The sample rate is 44.1 kHz.

The plot in Fig. 3 shows how the MPD varies with $\Delta A$ and $\Delta f$. As for the "goodness of fit" and "reassignment difference" measures, the distance of a component from the center of analysis bin has a small effect on the measured value of the MPD. Again, these differences are ignored so that a single look-up table can be used. As for those two previous discriminators a $100 \times 100$ array and 2-D linear interpolation is used to determine expected values of the MPD given the estimates obtained for $\Delta A$ and $\Delta f$. The difference between the actual MPD for the peak being tested and the MPD expected for a sinusoid with these parameters is thresholded. Values below the threshold are classified as sinusoids, values above are classified as nonsinusoids.

### E. Thomson's Multi-Taper Method

The final discriminator considered in this study is a modified version of Thomson's F-Test [16], [17]. Thomson's method employs a set of orthogonal windows which are optimally concentrated in frequency, prolate spheroidal functions (also known as Slepian sequences), to obtain a number of Fourier spectrums—one for each window applied to the signal under analysis. Combining estimates from each Fourier spectrum reduces bias and variance at the expense of greater spectral leakage.

Since there is more than one estimate for each component the values of each of these estimates can be compared for a single component: to what extent are each of the values consistent with the others as the estimate of a single sinusoidal process? The F-test is a Fisher–Snedecor test which compares the mean of the variance which can be explained by the presence of a single stationary sinusoid, with the mean of the variance which is not due to that sinusoid (i.e., the energy in that bin which is not due

to a sinusoid of the estimated amplitude) [17]. The test statistic, for an analysis bin $k$, is given by

$$F_{k,p} = W \frac{\hat{A}_{k,p}^2 \sum_{w=1}^{W} V_{0,w}^2}{\sum_{w=1}^{W} |X_{-w,k,p} - \hat{A}_{k,p} V_{0,w}|^2} \tag{11}$$

where $w$ is the window function (the $w$th Slepian sequence), $W$ is the number of different windows used,

$$\hat{A}_{k,p} = \frac{\sum_{w=1}^{W} X_{w,k,p} V_{0,w}}{\sum_{w=1}^{W} V_{0,w}^2} \tag{12}$$

is the estimated amplitude of the sinusoid obtained from the multi-taper spectral estimates and $V_{0,w}$ is the magnitude of the window spectrum function at DC (0 Hz). The numerator of the fraction in (8) is the variance due to a sinusoid with the estimated amplitude, the denominator is the residual variance (i.e., that not due to a sinusoid with that estimated amplitude). For this study, a bandwidth of two analysis bins is chosen and three Slepian windows are applied ($W = 3$). Although the input frame is tapered using a Hann window for the nonstationary parameter estimation, it is the original untapered frame to which the Slepian windows are applied.

One approach to adapting this discriminator to nonstationary sinusoids would be to sample $F_{k,p}$ for sinusoids with a range of $\Delta A$ and $\Delta f$ values and use these with a single interpolated 2-D lookup table, as for some of the discriminators already described (A, B, and D). However, the value of $F$, even for a stationary sinusoid, varies with the deviation of $f$ from the center frequency of the analysis bin. As $f$ approaches the center of the bin $F$ increases, since the value of $X_{w,k,p}$ approaches that of $\hat{A}V_0$, tending to infinity as the frequency deviation tends to zero. This means that there will be a wide range of expected $F$ values, meaning that the effect of the difference in $F$ due to deviation from center of the bin could be much greater than that due to the nature of the underlying component. Therefore, this approach cannot be used. Instead $W$ (= 3 in this case) 2-D lookup tables are used to store the values of $V_{0,w}$ for varying $\Delta A$ and $\Delta f$. From these three tables, the magnitude of each of the three windows for the estimated $\Delta A$ and $\Delta f$ can be used in (11) to generate a single value $F$. The tables for the three Slepian windows used in this study are shown in Fig. 4. A single ridge-like discontinuity can be seen in the plot for $V_2$. This is due to the position of the component shifting relative to the center of the analysis bin for this window (the position within an analysis bin for a given component is not the same for each window). As the position of a component in a second bin moves from the edge of one bin to the opposite edge of the adjacent bin this discontinuity in the measured value is observed. Again, variations in the measured values due to distance from the center of the analysis bin are ignored. Once $F$ has been estimated for a spectral component it is thresholded: higher than the threshold the component is classified as a sinusoid, a nonsinusoid if it is lower.

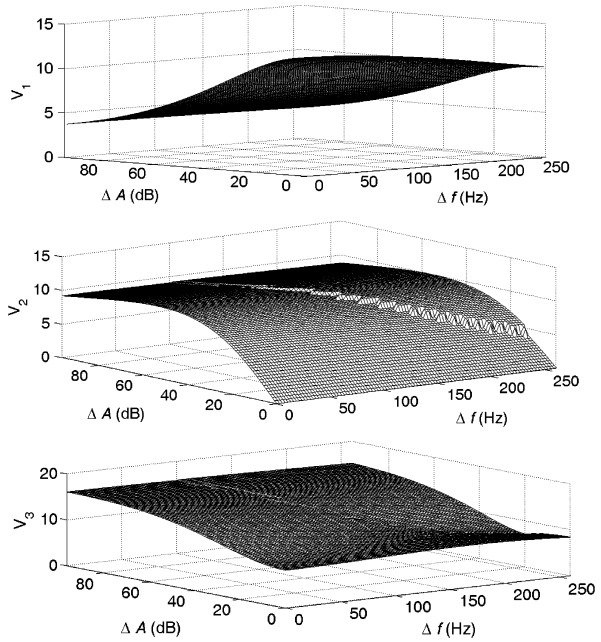


Fig. 4. Window magnitude for three Sepian sequences at center of peak for a sinusoid with unity amplitude and intra-frame amplitude and frequency change $\Delta A$ and $\Delta f$.

## IV. Evaluation of Discriminators

The performance of the discriminators, for sinusoids with a range of $\Delta A$ and $\Delta f$ values combined with noise, are compared in this study by the use of ROC graphs. These graphs offer a straightforward way of evaluating and comparing the performance of binary classifiers [25]. They are 2-D plots of the "false positive rate" (FPR) against the "true positive rate" (TPR) where

$$\mathrm{FPR} = \frac{\text{false postitives}}{\text{false postitives} + \text{true negatives}} \tag{13}$$

$$\mathrm{TPR} = \frac{\text{true postitives}}{\text{true postitives} + \text{false negatives}}. \tag{14}$$

A point at the top left-hand corner of an ROC curve ($\mathrm{TPR} = 1.0$, $\mathrm{FPR} = 0.0$) indicates a discriminator that is performing perfectly whereas a point in the diagonally opposite corner indicates the poorest performance possible (no correct classifications). A point that lies on the line $\mathrm{TPR} = \mathrm{FPR}$ indicates a classifier whose decision making process is entirely random. A curve for a thresholding discriminator can be produced by varying the threshold across a range of values that will produce $\mathrm{FPR} = 1.0$ and $\mathrm{TPR} = 0.0$ at one extreme of the range and $\mathrm{TPR} = 1.0$ and $\mathrm{FPR} = 0.0$ at the other. A perfect classifier will produce a line which moves from (0.0, 0.0) to (0.0, 1.0) and from there to (1.0, 1.0). The closer a curve is to this perfect trajectory in ROC space the better its performance is for the signal and conditions for which it is being tested, as illustrated in Fig. 5. A useful feature of ROC curves is that they do not change if the distribution of instances (proportion of sinusoids to nonsinusoids in this case) changes [18]. Individual ROC curves which provide a limited comparison of some of the discriminators described here were presented in [7]. Further examples of ROC curves are given in the Results section of this paper.

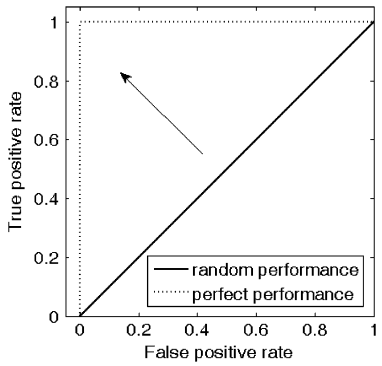


Fig. 5. Representations of binary discrimination performance with ROC curves.

The results of this study are presented in the next section, along with a discussion of their computational costs.

Individual ROC curves are only able to give an indication of the performance of a discriminator for a single set of conditions. A single value which can be derived from an individual curve is the area under the ROC curve (AUC). The AUC provides an estimate of the probability that a discriminator will rate a randomly chosen positive instance (a sinusoid in this case) higher than a randomly chosen negative instance (a nonsinusoid) [18]. A perfectly performing discriminator will have an AUC of 1.0, the AUC will be 0.5 for random performance. An AUC of 0.0 indicates a discriminator which consistently makes the wrong classification. An advantage of ROC analysis is that it illustrates the behavior of the discriminator for all possible thresholding values. There is no one "correct" threshold value for any of the discriminators—its setting will depend upon the required balance between TPR and FPR and could possibly be made to adapt to the input signal.

For this paper, the AUC for each discriminator is measured against the sinusoid-to-noise ratio (SNR) for different sets of $\Delta A$ and $\Delta f$ values. These sets of values represent: stationarity ($\Delta A = 0$ dB and $\Delta f = 0$ Hz), ranges of nonstationary amplitude and frequency commonly observed in speech ($\Delta A \in \mathrm{U}[0, 20]$ dB and $\Delta f \in \mathrm{U}[0, 30]$ Hz), and extreme nonstationarity ($\Delta A \in \mathrm{U}[0, 96]$ dB and $\Delta f \in \mathrm{U}[0, 260]$ Hz). For the latter set the range of values for $\Delta A$ represents the dynamic range of a 16 bit system and, for $\Delta f$, the range of values for which the $\Delta f$ estimator is monotonically increasing [2]. In order to evaluate the effect of amplitude and frequency change in isolation the AUC is also determined for ($\Delta A \in \mathrm{U}[0, 96]$ dB and $\Delta f = 0$ Hz) and ($\Delta A = 0$ dB and $\Delta f \in \mathrm{U}[0, 260]$ Hz). The SNR is measured at 5-dB increments over the range $-30$ to 30 dB. For each frame a sinusoid with the desired parameters is first synthesized. Then a vector of the same length containing white Gaussian noise is synthesized. This vector's energy is normalized to that of the sinusoid, the required gain (or attenuation) is then applied to the noise and it is added to the sinusoid to produce the frame. The values of $\varphi$ and $f$ are randomly selected from uniform distributions with the ranges [0, $2\pi$] and [20, 20 k], respectively. The value of $f$ is additionally constrained so that the instantaneous frequency is always between 20 Hz and 20 kHz for the duration of the frame with account taken of the value of $\Delta f$ so that no chirps can ever have a frequency outside of this range. The value of $A$ is not randomized since it has no effect on the performance of the estimators. When using ROC analysis in this way we are able to compare the output of the classifier (sinusoid/nonsinusoid) with the identity of the component that was synthesized (sinusoid or noise). However, it is possible that occasionally the noise signal may, by chance, be sinusoidal in parts of the spectrum. In this case, if the discriminator does correctly classify the peak as a sinusoid, it will be marked as a false positive. This will cause the ROC curve to be biased downwards and to the right. In such cases our knowledge of the spectrum is limited to that of the discriminators, since we have no *a priori* knowledge of how "sinusoidal" a noise component may be. One approach to this problem might be to subtract from the time-domain signal a sinusoid synthesized from the estimated parameters for the peak, to see whether this leads to a reduction in the overall amplitude of the signal (an approach similar to that employed in Matching Pursuits [22]). The difficulty here is in establishing what constitutes a significant reduction in amplitude (how close it should be to the estimated amplitude of the sinusoid) and the problem faced is in designing and evaluating a discriminator to test for this, which would require ROC or similar analysis at which point the investigation would become circular. In the results presented here the only components which would be "true positive" for a classification as a sinusoid are those that are synthesized as such.

One ROC curve and its associated AUC for each discriminator are produced from 1000 such frames. This process is repeated 20 times for each set of $\Delta A$ and $\Delta f$ and values at each SNR point. The values of $\Delta A$ and $\Delta f$ and values that are used by the discriminators are the estimates generated by the first part of the analysis, as if there were no *a priori* knowledge of them. This means that the discriminators are tested in context and the results give an indication of how they would perform in real-world situations (i.e., ones in which the parameters of the sinusoidal components are not known and must be estimated). For comparison, a final set of results is presented for ($\Delta A \in \mathrm{U}[0, 96]$ dB and $\Delta \mathrm{f} \in \mathrm{U}[0, 260]$ Hz), where the parameters ($A$, $f$, $\varphi$, $\Delta A$, and $\Delta f$) of the sinusoid *are* known *a priori*.

A summary of the test procedure is as follows.

1) Generate 1025 sample frame at 44. 1 kHz containing noise and a single sinusoid with the required parameters.
2) Perform steps 1–3 listed at the beginning of Section III.
3) Obtain a sinusoidality measure for each magnitude peak identified for each discriminator under test (DUT).
4) Repeat the above three steps 1000 times. For each DUT calculate its ROC curve and the AUC.
5) Repeat steps 1–4 20 times and take the mean.
6) Repeat steps 1–5 once for each SNR point.
7) Repeat steps 1–6 once for each set of values for $\Delta A$ and $\Delta f$).

The results of this study are presented in the next section, along with a discussion of their computational costs. This study only considers discrimination ability for single sinusoids in a spectrum—such a signal type is uncommon in music and speech. However, the purpose of this work is to evaluate the "raw" discrimination ability of these algorithms. Consideration

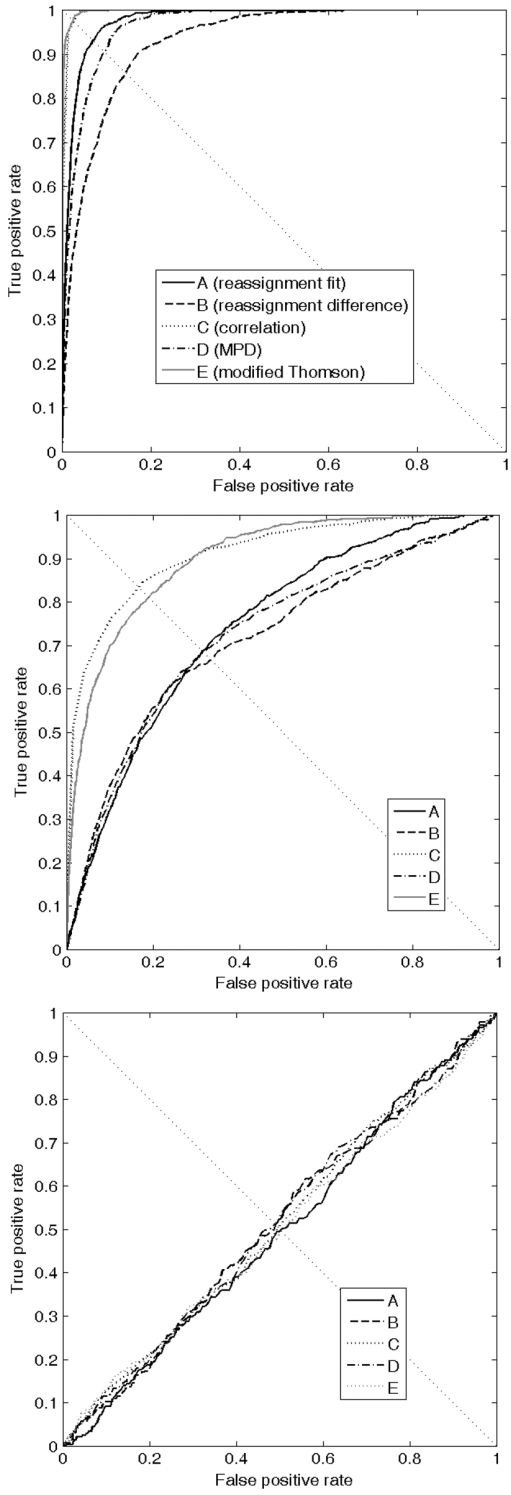


Fig. 6. ROC curves for SNR $=$ 30 dB (top), 0 dB (middle), and $-30$ dB (bottom). Intra-frame $\Delta A \in \mathrm{U}[0, 96]$ dB and $\Delta f \in \mathrm{U}[0, 260]$ Hz.

of how performance is affected by multi-component signals will be the subject of future work.

## V. Discriminator Performance

### A. Example ROC Curves

To further illustrate how ROC curves represent discriminator performance and to give examples of thresholding values and range for each of the discriminators, three curves are presented in Fig. 6. These are curves generated from one run of 1000 tests for $\Delta A \in \mathrm{U}[0, 96]$ and $\Delta f \in \mathrm{U}[0, 260]$. The top set of curves is for SNR $=$ 30 dB and it can be seen that the performance of all discriminators is much better than random, with discriminators C and E performing best and discriminator B performing the worst. The middle set of curves is for SNR $=$ 0 dB, here the performance of all discriminators is worse, with discriminators C and E again performing best and the other three having similar performance to each other. The bottom set of curves is for SNR $=$ $-30$ dB, here all of the discriminators exhibit close to random performance. To give some idea of the range and distribution of values returned by each of the discriminators, Table I gives the minimum and maximum values encountered for each discriminator across all SNR values and the value of the discriminator at the point at which it intersects the diagonal dashed line in each of the plots.

TABLE I
Example Discriminator Values and Ranges

| | Discriminator | | | | |
|---|---|---|---|---|---|
| | A | B | C | D | E |
| minimum | $3.3\text{x}10^{-6}$ | $3.0\text{x}10^{-5}$ | $2.7\text{x}10^{-3}$ | $4.3\text{x}10^{-7}$ | $1.2\text{x}10^{-1}$ |
| maximum | $3.3\text{x}10^{9}$ | $9.2\text{x}10^{2}$ | $1.0\text{x}10^{0}$ | $1.8\text{x}10^{1}$ | $4.3\text{x}10^{8}$ |
| 30 dB SNR intersect | $2.3\text{x}10^{-1}$ | $3.8\text{x}10^{0}$ | $1.0\text{x}10^{0}$ | $8.1\text{x}10^{-2}$ | $2.8\text{x}10^{2}$ |
| 0 dB SNR intersect | $2.4\text{x}10^{0}$ | $8.7\text{x}10^{0}$ | $9.9\text{x}10^{-1}$ | $2.6\text{x}10^{-1}$ | $8.8\text{x}10^{1}$ |
| -30 dB SNR intersect | $7.7\text{x}10^{0}$ | $1.4\text{x}10^{1}$ | $9.8\text{x}10^{-1}$ | $4.0\text{x}10^{-1}$ | $1.9\text{x}10^{1}$ |

### B. Discrimination Ability

Overall plots of AUC against SNR for each of the parameter sets are shown in Figs. 7 and 8. Tables II and III give the mean of the mean AUC values for low and high SNR ranges. This gives a crude measure of the effectiveness of each discriminator.

It is clear that the "non-stationary correlation" (discriminator C) method is the most effective, having higher mean values of AUC than the others overall. Its performance is near perfect, and significantly better than any other discriminator, where parameters are known *a priori* (bottom of Fig. 8). The modified Thomson method is the second best discriminator overall although it does not perform as well when the sinusoids are stationary (top of Fig. 7). An explanation for this is that the Fourier transform of the second Slepian sequence is 0 $(V_{0,2} = 0)$ at the peak due to a stationary sinusoid. This reduces the information available to the discriminator when a peak due to a sinusoid is being evaluated by (11).

It is more difficult to separate the remaining methods in terms of their discrimination ability. The "time reassignment goodness of fit" (A) discriminator performs marginally better then both the "phase and amplitude reassignment difference" (B) and MPD (D), although B performs relatively well for low SNR and where parameters are known *a priori*. It is interesting to note that for stationary sinusoids discriminator E has performance comparable to A, B, and D. However, it should be remembered that the discriminators are using estimates of $\Delta A$ and $\Delta f$, not assuming that these two parameters are 0 (as is the case for original Thomson F-test method). It is reasonable to assume that, in

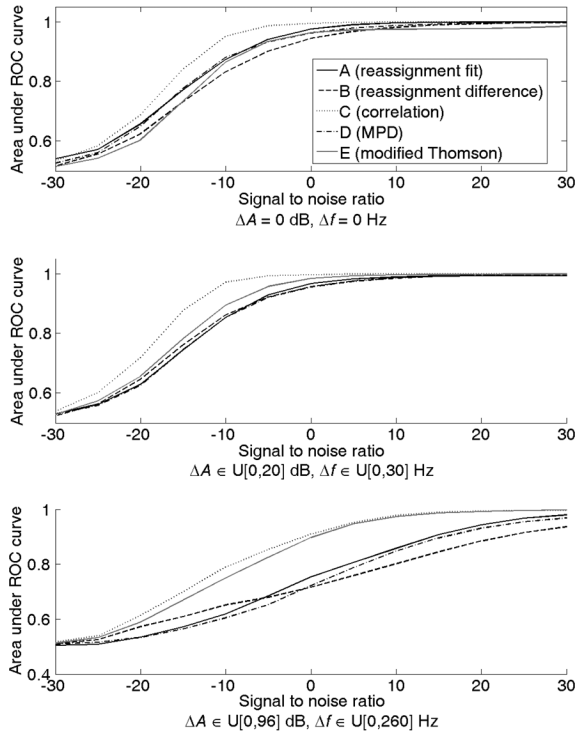


Fig. 7. Mean area under receiver–operator characteristic curve against signal-to-noise ratio for stationary sinusoids and combinations of intra-frame $\Delta A$ and $\Delta f$ (parameters of sinusoids not known *a priori*).

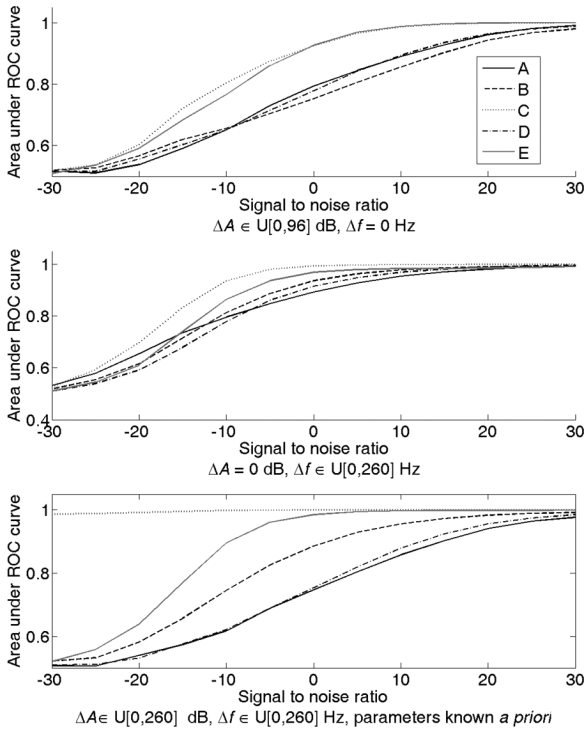


Fig. 8. Mean area under receiver–operator characteristic curve against signal-to-noise ratio for stationary for isolated intra-frame $\Delta A$ and $\Delta f$ (top and middle, parameters of sinusoids not known *a priori*) and for combined intra-frame $\Delta A$ and $\Delta f$ (bottom, sinusoidal parameters are known *a priori*).

this case, the Thomson discriminator, and the others, would perform better if $\Delta A$ and $\Delta f$ were known, *a priori*, to be 0.

TABLE II
MEAN OF AREA UNDER CURVE FOR SNR $= -30, -25, -20, -15, -10, -5$, 0 dB

| | Discriminator | | | | |
|---|---|---|---|---|---|
| $\Delta A$ , $\Delta f$ | A | B | C | D | E |
| 0 dB, 0 Hz | 0.76 | 0.73 | 0.80 | 0.76 | 0.74 |
| [0,20] dB, [0,30] Hz | 0.75 | 0.75 | 0.81 | 0.74 | 0.77 |
| [0,96] dB, [0,260] Hz | 0.60 | 0.61 | 0.70 | 0.59 | 0.68 |
| mean | 0.70 | 0.70 | 0.77 | 0.70 | 0.73 |

TABLE III
MEAN OF AREA UNDER CURVE FOR SNR $=$0, 5, 10, 15, 20, 25, 30 dB

| | Discriminator | | | | |
|---|---|---|---|---|---|
| $\Delta A$ , $\Delta f$ | A | B | C | D | E |
| 0 dB, 0 Hz | 0.99 | 0.98 | 1.00 | 0.99 | 0.98 |
| [0,20] dB, [0,30] Hz | 0.99 | 0.99 | 1.00 | 0.99 | 1.00 |
| [0,96] dB, [0,260] Hz | 0.89 | 0.84 | 0.97 | 0.87 | 0.97 |
| mean | 0.96 | 0.94 | 0.99 | 0.95 | 0.98 |

A question that arises is how discriminator performance is affected if the shape of the amplitude variation is not monotonically exponential throughout the frame. This will be subject of a detailed study in the future but initial experiments have shown that, taking the extreme case of a sinusoid that instantaneously onsets in the middle of the frame, the reassignment difference (B) and spectral correlation (C) methods exhibit a significant decline in discrimination performance compared to the others. The modified Thomson discriminator is the best overall in this case. Such a sinusoid would have a greater probability of correct detection in the next frame since it would span the whole of that frame, and many would consider an instantaneous onset a limiting case of what constitutes a sinusoid.

### C. Computational Cost

An important aspect of a discriminator's performance is the computational cost of its discrimination ability, although its importance will depend on the application. For example, it is likely the computational cost will be of high importance in a real-time analysis-for-modeling context but will be less critical where the discriminator is being used offline (e.g., to improve the robustness of a partial tracking algorithm). Ideally, a numerical measure of "unit AUC per unit time" would exist that would fully describe the ability of a discriminator and cost of that ability. Such a universal measure cannot be realized due to differences in cost between types of operation (e.g., addition, multiplication, memory read/write access) across hardware and software, as well the signal that the discriminator is operating on. The dependency of cost on signal type can vary between discriminators. For example, for discriminator C the number of complete FFTs that need to be performed per frame is dependent upon the number of magnitude peaks identified, whereas for the others a fixed number of FFTs per frame are performed. In addition, it may be that calculations already performed in previous analysis steps can be reused so there is a context-dependency to the computational cost. Table IV provides a summary of the complexity of each of the discriminators and ranks them in terms of computational cost (the highest cost is denoted by the lowest rank). $N$ is the FFT size, $L$ represents the FFT complexity, which is

TABLE IV
DISCRIMINATOR COMPLEXITY AND COST

| | Discriminator | | | | |
|---|---|---|---|---|---|
| | A | B | C | D | E |
| Complexity | $O(R+L)$ | $O(R+L)$ | $O(R(N+L))$ | $O(R+L)$ | $O(R+L)$ |
| FFTs per frame | 3 | 3 | $3+R$ | 4 | 6 |
| Rank | 4 | 4 | 1 | 3 | 2 |

$O(N\ln(N))$, and $R$ is the number of peaks detected in one frame. This table does not include the cost of synthesizing sinusoids, required by discriminator C which can be an intensive process (especially if the sinusoids are nonstationary); however, if the range $m$ over which the correlation is measured is small, a cheaper alternative to performing a full FFT (i.e., a DFT just over the correlation range) could be used.

## VI. CONCLUSION

Five methods for separating nonstationary sinusoids from Gaussian white noise have been presented and evaluated in the context of a high-accuracy analysis system for frame-by-frame spectral modeling of audio. Of the five the "nonstationary spectral correlation" discriminator (C) is the most effective but is also the most expensive, except in situations where only three or less magnitude peaks are detected (at which point the Thomson discriminator (E) becomes the most expensive since it requires three FFTs regardless of the number of magnitude peaks). A general trend observed is that the more expensive the discriminator, the better its performance is. The exception to this is the MPD (discriminator D) which requires an additional FFT per frame but performs relatively poorly. Although it does not demonstrate much promise as a discriminator of stationary or nonstationary sinusoids, the MPD may well still be of use in spectral modeling for the identification of time domain impulses, as described in [15]. Where computational resources are limited discriminators A and B offer reasonable performance at low cost.

## REFERENCES


[1] X. Serra, "A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition," Ph.D. dissertation, Center for Comput. Res. in Music and Acoust., Stanford Univ., Stanford, CA, 1989.

[2] J. Wells and D. Murphy, "High-accuracy frame-by-frame non-stationary sinusoidal modeling," in *Proc. 2006 Int. Conf. Digital Audio Effects (DAFx-06)*, pp. 253–258.

[3] J. Wells, "Real-time spectral modeling of audio for creative sound transformation" Ph.D. dissertation, Dept. of Electronics, Univ. of York, New York, 2006 [Online]. Available: http://www.jezwells.org/Jez_Wells_PhD.pdf

[4] M. Abe and J. Smith, III, "AM/FM rate estimation for time-varying sinusoidal modeling," in *Proc. IEEE ICASSP*, 2005, pp. 201–204.

[5] M. Betser, P. Collen, G. Richard, and B. David, "Estimation of frequency for AM/FM models using the phase vocoder framework," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 505–517, Feb. 2008.

[6] S. Marchand and P. Depalle, "Generalization of the derivative method to non-stationary sinusoidal modeling," in *Proc. 2008 Int. Conf. Digital Audio Effects (DAFx-08)*, pp. 281–288.

[7] J. Wells and D. Murphy, "Single-frame discrimination of non-stationary sinusoids," in *Proc. IEEE WASPAA*, 2007, pp. 94–97.

[8] X. Serra, "Musical sound modeling with sinusoids plus noise," in *Musical Signal Processing*, C. Roads, S. Pope, A. Picialli, and G. De Poli, Eds., 1st ed. Lisse, The Netherlands: Swets and Zeilinger, 1997, pp. 91–122.

[9] F. Keiler and S. Marchand, "Survey on extraction of sinusoids in stationary sounds," in *Proc. Int. Conf. Digital Audio Effects (DAFx-02)*, 2002, pp. 51–58.

[10] F. Auger and P. Flandrin, "Improving the readability of time–frequency and time-scale representations by the reassignment method," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1068–1089, May 1995.

[11] P. Masri and A. Bateman, "Identification of nonstationary audio signals using the FFT, with application to analysis-based synthesis of sound," in *Proc. IEE Colloq. Audio Eng.*, May 1995, pp. 1–6.

[12] M. Desainte-Catherine and S. Marchand, "High-precision Fourier analysis of sounds using signal derivatives," *J. Audio Eng. Soc.*, vol. 48, pp. 654–667, July/Aug. 2000.

[13] S. Hainsworth and M. Macleod, Time Frequency Reassignment: A Review and Analysis Cambridge Univ. Eng. Dept., CUED/F-INFENG/TR.459, 2003 [Online]. Available: http://citeseer.ist.psu.edu/hainsworth03time.html, Tech. Rep.

[14] M. Lagrange, S. Marchand, and J. Rault, "Sinusoidal parameter extraction and component selection in a non-stationary model," in *Proc. Int. Conf. Digital Audio Effects (DAFx-02)*, 2002, pp. 59–64.

[15] S. Fulop and K. Fitz, "Separation of components from impulses in reassigned spectrograms," *J. Acoust. Soc. Amer.*, vol. 121, pp. 1510–1518, Mar. 2007.

[16] D. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, Sep. 1982.

[17] D. Percival and A. Walden, *Spectral Analysis for Physical Applications*. Cambridge, U.K.: Cambridge Univ. Press, 1993.

[18] T. Fawcett, *ROC Graphs: Notes and Practical Considerations for Researchers*. Norwell, MA: Kluwer, March 16, 2004 [Online]. Available: http://home.comcast.net/~tom.fawcett/public_html/papers/ROC101.pdf

[19] W. J. Conover, *Practical Nonparametric Statistics*, 3rd ed. New York: Wiley, 1999.

[20] G. Peeters and X. Rodet, "SINOLA: A new analysis/synthesis method using spectrum peak shape distortion, phase and reassigned spectrum," in *Proc. Int. Comput. Music Conf. (ICMC 99)*, 1999.

[21] X. Rodet, "Musical sound signal analysis/synthesis: Sinusoidal+residual and elementary waveform models," in *Proc. IEEE Symp. Time-Freq. Time-Scale Anal. (TFTS 97)*, 1997.

[22] S. Mallat and Z. Zhang, "Matching pursuits with time–frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[23] Oppenheim, R. Schafer, and J. Buck, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[24] M. Abe and J. Smith, III, Design criteria for the quadratically interpolated FFT method (III): Bias due to amplitude and frequency modulation Dept. of Music, Stanford Univ., Stanford, CA, Tech. Rep. STAN-M-116, Oct. 2004.

[25] J. Egan, *Signal Detection Theory and ROC Analysis*, 1st ed. New York: Academic, 1975.



**Jeremy J. Wells** is a graduate of the Tonmeister course at the University of Surrey, Guildford, U.K., where he received the B.Mus. (Hons) degree in 1994, and he received the M.Sc. degree in music technology and the Ph.D. degree from the University of York, York, U.K., in 2001 and 2007, respectively.

He is currently a Lecturer for Music Technology in the Department of Electronics, University of York, where he has also worked as a Teaching Fellow and Research Associate.

He previously held posts with Digital Audio Research, Ltd., and Fairlight as well as working as a freelance Recording Engineer, Lecturer, and Musician. His research interests are spectral modeling of audio, especially high-accuracy single-frame analysis methods, room acoustics modeling, and studio effects processing. He is a trustee of Accessible Arts and Media, a charity which aims to widen participation in arts and performance as well as improving access to their related technologies. As a DJ, he has appeared at various clubs and festivals.

Dr. Wells is a member of the Audio Engineering Society.

**Damian T. Murphy** (M'09) received the B.Sc. (Hons.) degree in mathematics, the M.Sc. degree in music technology, and the D.Phil. degree in music technology, all from the University of York, York, UK, in 1993, 1995, and 2000, respectively.

In 1999, he was Lecturer in Music Technology in the School of Engineering, Leeds Metropolitan University, Leeds, U.K., and in 2000 was appointed as Lecturer in the Department of Electronics, University of York, where he is now Senior Lecturer in Music Technology. He has worked as an Independent Audio Consultant and since 2002 has been a Visiting Lecturer in the Department of Speech, Music, and Hearing, KTH, Stockholm, Sweden. His research is in the areas of physical modeling, virtual acoustics, and spatial sound, with particular interests in applications of the digital waveguide mesh. He is an active composer in the fields of electroacoustic and electronic music and in 2004 was appointed as one of the U.K.'s first AHRC/ACE Arts and Science Research Fellows, investigating the compositional and aesthetic aspects of sound spatialization and acoustic modeling techniques. He has recently held visiting researcher status during a series of collaborative exchanges with Helsinki University of Technology, The Sonic Arts Research Centre, Queen's University Belfast, and the Centre for Interdisciplinary Research in Music Media and Technology at McGill University, Canada.

Dr. Murphy is a member of the Audio Engineering Society.