MODAL DISTRIBUTION SYNTHESIS FROM SUB-SAMPLED AUTOCORRELATION FUNCTION

Thomas Lysaght
Department of Computer Science, NUI, Maynooth
Co. Kildare, Ireland
Tom.Lysaght@nuim.ie

Joseph Timoney
Department of Computer Science, NUI, Maynooth
Co. Kildare, Ireland
Joseph.Timoney@nuim.ie

Victor Lazzarini
Department of Music, NUI, Maynooth
Co. Kildare, Ireland
Victor.Lazzarini@nuim.ie

ABSTRACT

The problem of signal synthesis from bilinear time-frequency representations such as the Wigner distribution has been investigated [1,2,4] using methods which exploit an outer-product interpretation of these distributions. The Modal distribution is a time-frequency distribution specifically designed to model the quasi-harmonic, multi-sinusoidal, nature of music signals and belongs to the Cohen general class of time-frequency distributions. Existing methods of synthesis from the Modal distribution [3] are based on a sinusoidal-analysis-synthesis procedure using estimates of instantaneous frequency and amplitude values. In this paper we develop an innovative synthesis procedure for the Modal distribution based on the outer-product interpretation of bilinear time-frequency distributions. We also propose a streaming object-oriented implementation of the resynthesis in the SndObj library [6] based on previous work which implemented a streaming implementation of the Modal distribution [7]. The theoretical background to the Modal distribution and to signal synthesis of Wigner distributions is first outlined followed by an explanation of the design and implementation of the Modal distribution synthesis. Suggestions for future extensions to the synthesis procedure are given.

1. INTRODUCTION

The Modal distribution was introduced by Pielemeier and Wakefield [3] as a member of the Cohen general class of time-frequency distributions [5] for the analysis of music signals. It is primarily a Wigner distribution, or more specifically, a smoothed pseudo-Wigner distribution (SPWD), with a kernel that takes account of the *modes* present in quasi-harmonic, multi-sinusoidal, music signals. Being based on the Wigner distribution, it provides a more accurate measure of time-frequency localisation and does not suffer from the time-bandwidth trade-off inherent in the spectrogram (also a member of the Cohen class) implementations. One drawback of the Wigner distribution is the existence of cross-terms amounting to beats between partials not existing in the original signal. The Modal distribution kernel is designed to minimize the effect of these cross terms for music signals. Furthermore, implementation of the time-smoothing kernel for the Modal distribution greatly reduces the number of Digital Fourier Transforms (DFTs) that need to be performed on the smoothed autocorrelation function and results in applying the DFT at hop steps related to the size of the time-smoothing kernel. Ultimately this decreases the load in computing the distribution. In order to apply an outer-product based synthesis procedure to the Modal distribution, therefore, it is necessary to devise a method of signal recovery from sub-sampled autocorrelation functions.

2. THEORETICAL BACKGROUND

Leon Cohen [5] proposed a general class of time-frequency distributions which are related through linear transformations. The set of all linear transformations of the Wigner distribution has come to be known as the Cohen general class. A two-dimensional kernel determines the linear transformation involved. The Wigner distribution, equation (1), in terms of the signal $f(t)$ and the spectrum $F(\omega)$ is given by:

$$\begin{aligned} W(t,\omega) &= \frac{1}{2\pi}\int f^*\left(t-\frac{1}{2}\tau\right) f\left(t+\frac{1}{2}\tau\right) e^{-j\tau\omega}\, d\tau \\ &= \frac{1}{2\pi}\int F^*\left(\omega-\frac{1}{2}\theta\right) F\left(\omega+\frac{1}{2}\theta\right) e^{jt\theta}\, d\theta \end{aligned} \tag{1}$$

Here the kernel is 1. The autocorrelation with the lag variable, τ, produces the time-relative-time or temporal autocorrelation function given in equation (4). An important property of the Wigner distribution is that it is real with $W^*(t,\omega) = W(t,\omega)$. Also, the Wigner distribution gives a clear picture of the instantaneous frequency and group delay, which is not the case for the spectrogram. These are important for resynthesis [1,7].

2.1. The discrete pseudo-Wigner Distribution

The discrete implementation of the pseudo-Wigner distribution with a frequency smoothing window function $w(k)$, with length $M = 2L-1$, $w(k) = 0$ for $|k| \geq L$ is then defined by:

$$PWD\left(n, \frac{m\pi}{M}\right) = 2\sum_{k=-L+1}^{L-1} g(n,k)\,p(k)\,e^{-2jk\frac{mn}{N}}, \quad m = 0,\ldots,M \tag{2}$$

where

$$p(k) = w(k)\,w^*(-k) \tag{3}$$

and:

$$g(n,k) = f(n+k)\,f^*(n-k) \tag{4}$$

$g(n,k)$ is known as the temporal correlation function (TCF) or autocorrelation function. Equation (2) can be interpreted as the discrete Fourier transform of the autocorrelation function $g(n,k)$ with respect to *n* for each value of *m*.

2.1.1. Cross terms

Given a music signal model as follows:

$$f(t)=\sum_{k=1}^{M}A_k e^{j(w_k t+\phi_k)} \tag{5}$$

where k is the partial series index, t is time, and the k^{th} term in the summation represents a partial with constant amplitude A_k, frequency ω_k, and phase ϕ_k, the Wigner distribution is:

$$\begin{aligned} W_f(t,\omega) &= \sum_{k=1}^{M}A_k^2\delta(\omega-\omega_k) \\ &+\sum_{k=1}^{M}\sum_{l=k+1}^{M}A_k A_l\cos([\omega_k-\omega_l]t+\phi_k-\phi_l) \\ &\times\delta\left(\omega-\frac{(\omega_k+\omega_l)}{2}\right) \end{aligned} \tag{6}$$

The partials of $f(t)$ (*auto* terms) are given by the first term in equation (6). The second double summation indicates the cross terms, arising from products between partials, which lie between any pair of auto terms. The magnitude of the cross terms is the product $A_k A_l$ of the amplitudes of auto terms k and l and they oscillate at a frequency, $(\omega_k+\omega_l)/2$ equal to the difference between the frequencies of the two auto terms. For strictly harmonic signals, the cross terms form a partial series an octave below the fundamental, resulting in cross terms which fall at the same frequencies of and therefore corrupt the autoterms, and also cross terms at partial frequencies not in the original signal.

2.2. The Modal distribution

The modal distribution in equation (7) was designed to minimise these cross terms in equation (6) for music signals. The modal kernel consists of two different filter functions. The time-smoothing window, $h_{LP}(p)$, has the effect of smoothing the cross terms in the time direction, and the frequency-smoothing window, $g_{LP}(l)$, implements cross term suppression in cases of frequency modulation. $h_{LP}(p)$, is chosen to be a low pass filter with an upper cut-off just below the minimum frequency spacing in the distribution, this being the fundamental frequency for quasi-harmonic signals. The discrete form of the modal distribution is defined by:

$$M(n,k)=\sum_{n=-L+1}^{L-1}R_{f,l}(n,l)g_{LP}(l)e^{\frac{-j2\pi kl}{2L}} \tag{7}$$

where $R_{f,l}(n,l)=R_f(n-p,l)h_{LP}(p)$ is the time-smoothed temporal autocorrelation function (STCF). Computing the time-smoothing in the autocorrelation domain greatly reduces the number of DFTs that need to be performed. DFT's need to be computed only at *hop* steps that sample at a rate approximately equal to the period of the time smoothing window.

2.3. The Autocorrelation Function

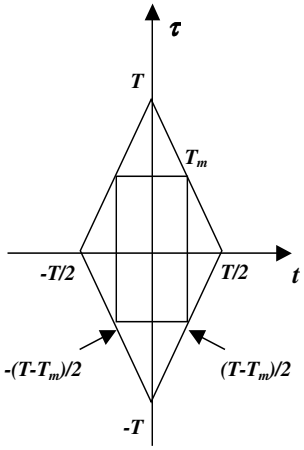

Figure 1: *Extent of the windowed autocorrelation function*

The autocorrelation function $g(n,k)$, represented by the diamond-shaped function in Figure 1, is sampled in time (t) at twice the Nyquist rate, or $2f_s$, and in relative-time (τ) at rate f_s. This function, then, has duration $2T$ in τ as shown in Figure 1. This requires that the discrete frequency index *k* in equation (2) be interpreted relative to this 2:1 sub-sampling rate [3]. With application of a 2-D kernel function, $2T_M$ represents the length of the frequency smoothing filter and the diamond-shaped region in the (t,τ) plane in Figure 1 is limited to the rectangular region [6]:

$$|t|>\frac{(T-T_m)}{2},|\tau|>T_m \tag{8}$$

3. MODAL DISTRIBUTION SYNTHESIS METHOD

From equation (2) the inverse discrete transform is given by:

$$y(n,2k)=g(n,k)p(k) \tag{9}$$

resulting in the autocorrelation function written as:

$$f(n+k)f^*(n-k)=\frac{y(n,2k)}{2g(k)g^*(-k)} \tag{10}$$

or with appropriate change of variable as:

$$\begin{aligned} &f(n)f^*(m)=c(n,m) \\ &=y(\frac{n+m}{2},n-m)\Big/2g\left(\frac{n-m}{2}\right).g^*\left(\frac{m-n}{2}\right) \end{aligned} \tag{11}$$

This outer-product formulation represents the product of two one-dimensional functions separable in *n* and *m* into the odd-or even-indexed sequences of signal samples. For the even-indexed samples, the outer-product formulation C_e can be written in matrix form as:

$$C_e = \begin{bmatrix} C_{0,0} \cdots C_{0,P} & C_{0,P+1} \cdots C_{0,L} \\ C_{P,0} \cdots C_{P,P} & C_{P,P+1} \cdots C_{P,L} \\ C_{P+1,0} \cdots C_{P+1,P} & C_{P+1,P+1} \cdots C_{P+1,L} \\ C_{L,0} \cdots C_{L,P} & C_{L,P+1} \cdots C_{L,L} \end{bmatrix} \quad (12)$$

where there are *P* known even samples and *L-P* samples to be recovered. This outer-product (OP) formulation is used in [1] for synthesis from overlapping blocks of C_e.

For signal synthesis from the modal distribution, however, only *hop* number of frames of the autocorrelation function are available and therefore the OP method cannot be directly applied. We derive an alternative method, which requires *hop* number of known samples to recover the sequence of odd- or even-indexed samples on a frame-by-frame basis.

3.1. Sub sampled autocorrelation function method

Synthesis of the signal samples from a sub sampled version of C_e is implemented in two stages: in the first stage, C_{e1}, processes all autocorrelation frames up to h_fft, half the DFT length (or $l = h\_fft/hop$ frames), where the size of each frame grows by hop number of samples and hop number of samples can be recovered from each frame. In the second stage, C_{e2} recovers $hop/2$ samples from all remaining frames. There are three cases only which must be processed separately:

i. the first frame of C_e contains the product $f_e(0)f_e^*(0)$ and so no processing is necessary

ii. the number of samples recovered from the second frame is *h/2*

iii. the number of samples recovered for frame *l+1* is:

$$h - a/2 \text{ where } a = (l+2) * h - (h\_fft - 1) \quad (13)$$

The matrix C_e can now be reformulated to take the hop step into account. For the even-indexed signal samples f_e we define the autocorrelation samples:

$$\begin{aligned} & f_e(n)f_e^*(m) = c_e(n,m) \\ & = y(\frac{n+m}{2}, n-m) \Big/ 2g\left(\frac{n-m}{2}\right).g\left(\frac{m-n}{2}\right) \end{aligned} \quad (14)$$

where $n = 0, hop, 2*hop, 3*hop, ..., t*hop$, and *t* is the total number of frames. Now we can write:

$$C_{e1} = \begin{bmatrix} C_{i_1,i_1}, & C_{i_1,i_1-2}, & C_{i_1,i_1-4} \cdots C_{i_1,i_1-2hop-2} \\ C_{i_2,i_2}, & C_{i_2,i_2-2}, & C_{i_2,i_2-4} \cdots C_{i_2,i_2-2hop-2} \\ C_{i_3,i_3}, & C_{i_3,i_3-2}, & C_{i_3,i_3-4} \cdots C_{i_3,i_3-2hop-2} \\ \cdots & & \\ C_{i_k,i_k}, & C_{i_k,i_k-2}, & C_{i_k,i_k-4} \cdots C_{i_k,i_k-2hop-2} \end{bmatrix} \quad (15)$$

where $\langle i_1, i_2, \cdots, i_k \rangle = \langle 2*hop, 3*hop, \cdots, l*hop \rangle$, gives the hop frame index. Given a matrix $A = \lfloor a_0 a_1 \cdots a_p \rfloor$ of *p=hop-1* known even samples, and $X_e = diag(A)$, a diagonal matrix generated from *A*, all even indexed samples from C_{e1} can be determined by:

$$F_{e1} = C_{e1} / X_e \quad (16)$$

Next we can write:

$$C_{e2} = \begin{bmatrix} C_{j_1,j_1+\alpha_1}, & C_{j_1,j_1+\alpha_1-2}, & \cdots C_{j_1,j_1+\alpha_1-hop+2} \\ C_{j_2,j_2+\alpha_2}, & C_{j_2,j_2+\alpha_2-2}, & \cdots C_{j_2,j_2+\alpha_2-hop+2} \\ C_{j_3,j_3+\alpha_1}, & C_{j_3,j_3+\alpha_1-2}, & \cdots C_{j_3,j_3+\alpha_1-hop+2} \\ \cdots & & \\ C_{j_s,j_s+\alpha_v}, & C_{j_{s1},j_s+\alpha_v-2}, & \cdots C_{j_s,j_s+\alpha_v-hop+2} \end{bmatrix} \quad (17)$$

where $\alpha_1 = hop/2 - 1$, $\alpha_1 = \alpha_2 = \alpha_v$ and $\langle j_1, j_2, \cdots, j_s \rangle = \langle k+2, k+3, \cdots, t \rangle$ for even *hop,* and $\alpha_1 = hop/2 - 2$, $\alpha_2 = hop/2 - 1$ and $\alpha_v = \alpha_1$ or $\alpha_v = \alpha_2$ depending on *s*, for odd *hop*. The sequence $(j_i + \alpha_{1or2}, \cdots, j_i + \alpha_{1or2} - hop + 2)$ where $i = 1,2,\cdots s$, represents the *hop/2* known samples used to recover *hop/2* even-indexed samples from each row of C_{e2}. An identical formulation applies to the odd-indexed samples f_o and so need not be outlined.

4. RESULTS

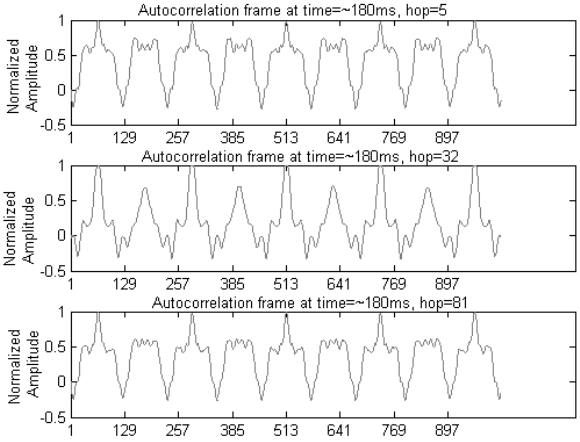

Figure 2: Autocorrelation function (for Bb Clarinet note G3) slices at time=180ms, for hop steps of 5, 32 and 81

Figure 2 shows TCF function relative-time slices at time 180ms, for a Bb clarinet G3 note of length 8000 samples with $f_s = 44100$ and $2T = 1024$ for hop sizes of 5, 32 and 81 respectively. In each case the peaks in the relative-time direction (horizontal axis) indicate the signal harmonics. For example, the

fundamental frequency can be seen from the 9 signal cycles in each plot, indicating a frequency of approximately $2f_s$ =9.1/1024 or ~196Hz (G3).

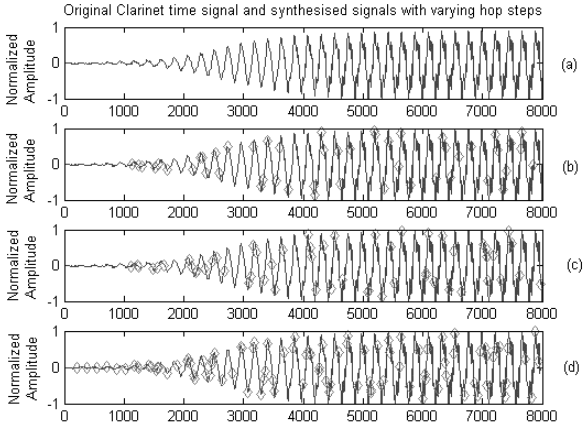

Figure 3: Comparison of (a) original Clarinet G3 note with synthesized Clarinet G3 note samples (b), (c) and (d) from respective autocorrelation functions in Figure 2. The diamonds in (b), (c), and (d) indicate gaps in the signal where samples could not be recovered.

Figure 3 shows the original signal (a), and the three signals recovered from the autocorrelation functions with hop steps of (b) 5, (c) 32, and (d) 81 respectively. The diamonds on plots (b)-(d) in Figure 3 indicate where zeros occur in the recovered signals. These missing samples are subsequently interpolated to avoid 'clicks' in the recovered signals. Hop sizes of arbitrary length were tested and in each case the synthesised samples recovered were identical to the original signal, apart from where zeros occurred, and the recovered signal was audibly indistinguishable from the sound of the original signal.

5. CONCLUSIONS AND FUTURE WORK

This frame-by-frame resynthesis method for Modal distributions exactly recovers the even- and –odd indexed signal samples for arbitrary hop steps. It provides an alternative signal recovery method for the Modal distribution based on the outer-product method in [1]. Current work focuses on a comparison of the outer-product approximation (OPA) method in [1] implemented for the Modal distribution using eigenvalue-eigenvector decomposition for signal recovery, with the method outlined in this paper. Immediate further work will implement signal filtering for the Modal distribution using these methods. Future work will also investigate the effect of the Modal distribution's smoothing kernel on this method and the possibility of signal modification in comparison with analysis-synthesis approaches. Finally, this frame-by-frame approach readily integrates into the SndObj library's Modal distribution routine [6,7], thus allowing a streaming implementation of Modal distribution synthesis in conjunction with many of the tools necessary for sound analysis and modification such as time stretching and vocoding.

6. REFERENCES

[1] K.-B., Yu,. and S. Cheng, " Signal synthesis from pseudo Wigner distribution and applications", *IEEE Trans. Acoust., Speech, Signal Processing,* Vol. ASSP-35, pp.1289-1302, Sept. 1987.

[2] W. Mecklenbraukner, and F. Hlawatsch (Editors), "*The Wigner Distribution – Theory and Applications in Signal Processing*", 1997, Elsevier Science B.V. pp. 135-209.

[3] W. J. Pielemeier, and G. Wakefield, "A high-resolution time-frequency representation for musical instrument signals", *Journal of Acoustical Society of America",* 99(4), Pt. 1, April 1996.

[4] G. F. Boudreaux-Bartels, "Time-Varying Filtering and Signal Estimation Using Wigner Distribution Synthesis Techniques", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-34, No. 3, June 1986.

[5] L. Cohen, *Time Frequency Analysis*. Prentice-Hall, New Jersey, 1995.

[6] V. Lazzarini, "The sound object library". *Organised Sound 5 (1).* Pp. 35-49, 2000.

[7] T. Lysaght, V. Lazzarini, and J. Timoney, 'A Streaming Object Oriented Implementation of the Modal Distribution' *In Proc. Of International Computer Music Conference (ICMC-05)*, Barcelona, Spain, September 5-9, 2005.