THE CHARACTERISATION OF BIO-ACOUSTIC SIGNALS THROUGH TIME-FREQUENCY METHODS.

Paul White[1]

[1] Institute of Sound and Vibration Research, University of Southampton, Highfield, Hants, U.K., S017 1BJ *prw@isvr.soton.ac.uk*

1. ABSTRACT

The application of time-frequency methods for the analysis of marine mammal calls is considered. Specifically the use of data adaptive time-frequency methods is motivated and principle behind a variety of such algorithms is discussed. These algorithms, and more classical methods, are applied to both synthetic and measured data sets.

KEY WORDS: Time-Frequency; Wavelets.

2. INTRODUCTION

The general structure of a signal classification algorithm contains several basic building blocks. First the signals of interest are detected, they are then characterised/analysed, salient descriptive features (parameters) are then computed and these features are finally supplied to the classification algorithm. Our interest is in the construction of such a scheme to classify the calls of marine mammals. This paper concentrates on the problems associated with the algorithms used for the characterisation of the calls, from which features will be extracted.

3. CLASSICAL TIME-FREQUENCY ANALYSIS

The fact that marine mammal calls are highly non-stationary means that the natural tool for their analysis is a form of time-frequency representation, such representations have a long history [1]. Time-frequency methods seek to capture the time-varying spectral characteristics associated with a non-stationary signal. One needs to resort to some form of time-frequency distribution because a purely temporal representation fails to adequately characterise the time-varying spectral characteristics of a signal, whilst the spectrum fails to successfully represent the temporal variations in a signal. The spectrogram is justified in a heuristic manner; specifically one can characterise a signal's spectrum at time, t_0, by windowing the signal about that time and computing the spectrum from this truncated data segment. This leads to the definition of the spectrogram, $S(t,f)$, as:

$$S(t,f) = \left| \int w(\tau - t_0)\, x(\tau) e^{-2\pi i f \tau} d\tau \right|^2 \qquad (1)$$

where $w(t)$ is a fixed windowing function, t and f represent time and frequency respectively. The spectrogram is a positive, real, valued function. One loosely considers the value of $S(t,f)$ as reflecting the signal's energy density at time t and f. The choice of the windowing function $w(t)$ critically affects the appearance of the distribution $S(t,f)$. The shape of the windowing function has a limited impact on the form of the spectrogram, however the choice of the window's duration can have profound affects on the resulting distribution. This point is illustrated by the example shown in Figure 1, which depicts two spectrograms for a section of a call from a Beluga or white whale (*Delphinapetrus leucas*). Both spectrograms are computed using Gaussian windows with different durations (129 samples for a) and 513 for b)). Whilst these two spectrograms represent exactly the same signal their appearance is very different. The 513 point window (Figure 1b)) is too long to capture the impulsive nature of the signal and this feature is obscurred.

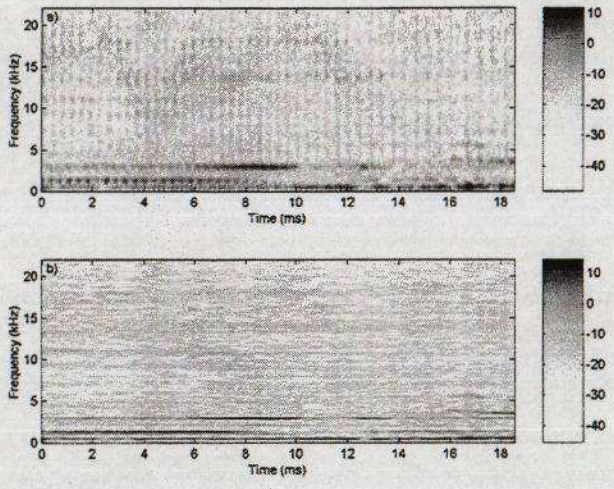

Figure 1: Spectrograms of a Beluga Whale Call a) Using a Gaussian Window of 129 samples, b) Using Gaussian Window of 513 samples

The choice of window duration involves a trade-off between resolution in time and resolution in frequency. If a short duration window is used then the time at which an event occurred can be well discerned from the spectrogram, *i.e.* good time resolution is achieved. Since the spectrogram is based on the Fourier transform then the frequency resolution is inversely proportional to the window duration. Meaning that spectrograms constructed using short duration windows necessarily have poor frequency resolution. If a long duration window is applied then spectrogram suffers from poor temporal resolution but has good frequency resolution. This limitation derives from Heisenburg's uncertainty principle [1,2], which states that there is a lower limit to the product of a signal's bandwidth and its duration. One often chooses to using a Gaussian windowing function *w(t)* in the spectrogram since that offers maximal frequency resolution for a given temporal resolution, it is the function that achieves the lower bound predicted in the uncertainty principle. This basic shortcoming of the spectrogram has long been recognised [1], but has not stopped the spectrogram becoming the most widely used tool for time-frequency analysis.

Marine mammals calls exhibit a wide variety of forms, including short duration clicks, moans and whistles. Such signals have very different durations and bandwidths making it difficult to select a single windowing function that well represents all signal types. This makes the spectrogram an imperfect tool for characterising marine mammal calls. It is untenable to suggest that the spectrogram is of no use for characterising the calls of marine mammals, since its extensive use clearly highlights that useful information can be gleaned from it. The contention here is that the use of better signal representation may lead to improved classification performance.

A second type of time-frequency representation is based on the wavelet transform. These distributions are in fact based on decomposing the signal into a function of time and scale, but in practice one can relate scale to frequency in almost all cases. The squared magnitude of the wavelet transform is referred to as the scalogram [3].

The scalogram can be viewed as a spectrogram constructed using windows whose duration varies with frequency. Specifically the duration of the window reduces in proportion to $1/f$. Consequently the scalogram exhibits good temporal resolution at high frequencies, with correspondingly poor frequency resolution, whilst at low frequency it has good frequency resolution and poor time resolution. This means that the scalogram has variable resolution, but

these variations are predetermined. The user has to select the shape of the basic decomposing function (the mother wavelet), this can be thought of as selecting the duration of the window at a given scale. Like the spectrogram the choice of this parameter has a great effect on the form of the computed scalogram.

A more general class of time-frequency representation (of which the spectrogram is a specific example) is variously referred to as Cohen's class, quadratic or bilinear representations. These representations are all based on a signal distribution, specifically the Wigner-Ville Distribution (WVD), *W(t,f)*, which is defined as:

$$W(t,f) = \int x(t-\tau/2)^{*} x(t+\tau/2) e^{-2\pi i f \tau} d\tau \qquad (2)$$

where * denotes complex conjugation. The WVD is necessarily real, but is not strictly positive (except for specific signals). The computation of this distribution requires no windowing of the signal. The WVD is the Fourier transform of the time-varying correlation function:

$$r(t,\tau) = x(t-\tau/2)^{*} x(t+\tau/2) \qquad (3)$$

The WVD possesses many properties [2,3] that one associates with a true energy density function. Two of the most important of which are the so-called marginal conditions:

$$\int W(t,f)df = |x(t)|^2 \quad \text{and} \quad \int W(t,f)dt = |X(f)|^2 \qquad (4)$$

So the integral of the WVD, across all frequencies, for a fixed time, yields the instantaneous energy of the signal at that time. Similarly when it is integrated across all times for a particular frequency, one obtains the spectral level at that frequency. Together these properties also imply that the integral of the WVD over all time and frequency yields the total energy of a signal. None of these conditions hold for the spectrogram.

Further to this the WVD can offer a resolution that is superior to that which can be achieved using the spectrogram. The reason being related to the manner in which the WVD is constrained by the uncertainty principle [2]. Specifically for the spectrogram the uncertainty principle applies to the windowed segment of the signal, whereas for the WVD it applies to the whole signal, such that the frequency resolution is limited by the total signal duration.

However there are two predominant drawbacks with the WVD. The first is that, for nearly all signals, it is negative for some values of *t* and *f*. This is primarily a theoretical issue, since it excludes the possibility of regarding the WVD as a true energy density. The second is a far more practical problem, specifically if a signal contains two distinct components, occupying different regions of the time-frequency plane, then the WVD will consist of three elements, one due to each of the components and a third interference term. Such interference terms are also referred as cross-terms, interaction terms or ghost terms. These interference terms ensure that the WVD will take on large values in areas of the time-frequency plane that are not associated with any components in the input signal. Interference terms differ from true signal terms in that they tend to be oscillatory.

A secondary effect of the interference terms, that is important in our application, is their affect on noise. In most circumstances the effect of noise on the WVD is greater than it effect on the spectrogram. This is can be seen by realising that noise contains a great many components, each pair of components gives rise to a noise like interference terms that raise the overall noise floor. The result is that the SNR observed in the WVD may be lower than that observed in a spectrogram.

The general class of time frequency representations are derived from the WVD are based on applying a two-dimensional low-pass filter to the WVD in order to suppress the interference terms. The reduction of these interference terms also serves to increase the SNR in the representation.

The application of such low-pass filters inevitably results in a loss of resolution in the distribution. A general Cohen's class time-frequency representation, *C(t,f)*, is given by:

$$C(t,f) = \iint W(\tau,\nu)\Pi(t-\tau,f-\nu)d\tau d\nu \qquad (5)$$

where the function $\Pi(t,f)$ is referred to as the kernel function and each member of Cohen's class is specified by defining its associated kernel function. The spectrogram can obtained in this manner by a specific choice of kernel function. The scalogram can also be obtained by smoothing the WVD [3] albeit using an affine smoothing operator.

There have been a great many time-frequency distributions proposed [2,3], the kernel functions for which have usually been selected to ensure that the resulting distribution has certain desirable properties. In each case the kernel function is defined with, at least one, user selected parameter that allows one to control the degree of smoothing applied to the distribution. The selection of these smoothing parameters entails a trade-off, not dissimilar to that involved when choosing the window length in the spectrogram. In the case of a Cohen's class distribution the trade-off is between resolution in the distribution and degree of suppression of the interference terms. Which, in the presence of noise, also involves the SNR of the representation.

Hence the use of Cohen's class distributions, like the use of the spectrogram requires one to partake of a trade-off. Further the choice of a good value for the smoothing parameter usually depends on the character of the signal under analysis, so that in situations where signals of differing character may arise then choosing a suitable set of smoothing parameters may be a severe obstacle. These observations have led researchers to develop alternative methods of time-frequency analysis that are more robust to the exact nature of the signal be analysed. Herein such algorithms will be collectively referred to as "data adaptive" although in some cases it is accepted that this title is somewhat inappropriate.

4. DATA ADAPTIVE TIME-FREQUENCY METHODS

It is not possible to provide details on all the various alternative methods for constructing time-frequency representations here. To provide a feel for the some of the available approaches 5 techniques will be briefly outlined, details for each method are to be found in the references given. Their advantages and shortcomings will be highlighted through results showing their effectiveness on both synthetic and real signals.

4.1 Re-Assignment Methods

One approach to improving the performance of time-frequency methods is through the use of re-assignment [4-6]. This methodology takes a classical time-frequency distribution and modifies it to effectively sharpen the peaks in the distribution. Re-assignment can be applied to any of Cohen's class [5,6], but here results using only the spectrogram will be shown. The reason for this is that the results from applying re-assignment are similar regardless of the original distribution[6].

The principle behind re-assignment is to consider each point in the time-frequency representation in turn and for it, compute a local estimate of the instantaneous frequency and group delay. These two quantities can be regarded as estimates of the centre frequency and temporal location of the energy close to the point under consideration. The re-assignment proceeds by then relocating the energy on the original distribution to the point defined by the estimated instantaneous frequency and group delay. This has the effect of clustering points on lines corresponding to true time-frequency laws of the components in the signal.

There are various techniques for computing the local estimates of instantaneous frequency and group delay, but good results can be obtained using an algorithm that requires a doubling of the

computational load. There are also other algorithms closely allied to this technique, which seek to describe the loci of time-frequency laws [7]

4.2 Local Matching Algorithms

There have been a variety of methods proposed that fall into this category, amongst the most widely known are the adaptive spectrogram [8] and the matching pursuit algorithm [9]. Here a generic algorithm is outlined.

The crux of these approaches is a decomposition of a signal using a set of basis functions that are over-complete. The algorithm described here employs a parameterised set of basis functions, for simplicity assume the set to be used is:

$$\phi(t;\sigma,\tau,f) = e^{-(t-\tau)^2/\sigma} e^{2\pi i f t} \qquad (6)$$

This is the set of complex sinusoids with a Gaussian envelope. The parameter τ controls the location of the function on the time axis, σ controls the width (duration) of the function and f dictates its centre frequency. The decomposition of a signal, $x(t)$, can then be written:

$$x(t) = \sum_{n=1}^{N} A_n \phi(t;\tau_n,\sigma_n,f_n) \qquad (7)$$

where A_n are the complex amplitudes. Note that each component in the decomposition employs a different parameter set, most importantly each basis function potentially has a different width. The immediate problem relates to how one constructs such a decomposition.

The simplest approach is based on the observation that the spectrogram computed using a Gaussian window, with width parameter σ, can be regarded as the square of the cross correlation of the signal with the function $\phi(t;\tau,\sigma,f)$ [8]. Hence a peak in the spectrogram will at (t,f) when the signal matches the basis function. One proceeds to construct the decomposition by selecting a prototype width parameter, σ , and computes the associated spectrogram. The time and frequency of the largest peak in the spectrogram is noted. Then the following cost function is minimised with respect to the four parameters A,τ,σ,f:

$$\int |e(t)|^2 dt \qquad e(t) = x(t) - A\phi(t;\tau,\sigma,f) \qquad (8)$$

using the location of the spectrogram peak and the prototype width as initialisation values for τ,f,σ, respectively. The amplitude A can be eliminated from the optimisation using standard techniques and is only computed after the other three parameters have been evaluated.

The above process finds the single basis function $\phi(t;\tau,\sigma,f)$ that best matches the data. The process proceeds iteratively taking the error signal $e(t)$ as the signal $x(t)$ in the next iteration. This process proceeds either until a predefined number of terms have been computed or until the error signal has become sufficiently small.

The above provides one scheme for constructing the model (7). Once this model has been formed one can construct the associated time-frequency representation. For example for the Gaussian functions (6) the WVD can be expressed analytically and exhibits no cross terms and is positive. So a natural candidate for the time-frequency representation of $x(t)$ based on the decomposition (7) is:

$$A(t,f) = \sum_{n=1}^{N} A_n^2 W_\phi(t,f;\tau_n,\sigma_n,f_n) \qquad (9)$$

where $W_\phi(t,f;\tau_n,\sigma_n,f_n)$ is the WVD of $\phi(t;\tau_n,\sigma_n,f_n)$. The distribution in (9) will be positive and will contain no interference terms, making it a promising proposition.

The above principles can be extended in many ways including using a more complex set of basis functions, usually involving a greater number of parameters and so complicating the fitting process. Alternatively one can constrain the basis functions to only be permitted to take on a small subset of parameter values, which simplifies the fitting process but at the expense of some loss of generality.

4.3 Optimised Kernels

The discussion in section 3 highlighted the fact that classical time-frequency methods require one to select parameter values that control the trade-off involved in computing time-frequency representations. The user, usual through a process involving a degree of "trial and error" selects these parameter values. The optimised kernel methods [10] can be viewed as automating this process or more generally allowing the kernel function, $\Pi(t,f)$, to select its own shape to match the signal.

This process evidently requires an optimisation procedure. To construct such a procedure one needs to define the parameters to be selected and cost function against which performance is to be measured. The kernel function can be parameterised using one of the established Cohen's class distributions and treating the user defined smoothing parameters as the objects of the optimisation. Alternatively one can define a more general kernel function, as described in [10], with a great many more degrees of freedom the parameters which are then subjected to the optimisation scheme. Furthermore one can choose to have a single kernel function that is applied to the whole signal or the kernel function itself can be a function of position in the time-frequency plane. In the latter case the resulting optimisation will in general be more complex because of the associated increase in dimensionality. The cost function that the optimisation scheme seeks to maximise is usually a measure of "peakiness" of the output distribution, normally involving the fourth order statistics of that output.

4.4 Cohen-Posch Distributions

As discussed none of the classical time-frequency distributions can be regarded as true energy densities. The minimal requirements of such a density are that the distribution is non-negative and that the marginal properties (4) are satisfied. It can be shown that the only members of Cohen's class that are non-negative for all signals are spectrograms and further it can be shown that a spectrogram can not satisfy the marginal properties. The existence of distributions that do satisfy these properties was demonstrated in [11]. To overcome the apparent impasse these time-frequency representations are necessarily signal dependent and so are not members of Cohen's class.

Methods for constructing distributions that have these desirable properties have appeared much more recently [12] and here a brief discussion of the simplest approach is given here. These methods are iterative in nature. One starts with an initial distribution, usually selected to be positive, and the marginal conditions are imposed alternately. The distribution is forced to satisfy the marginal conditions, in the simplest case, through re-scaling lines within the distribution. This algorithm was originally proposed through consideration of cross-entropy [11] but can also be considered as an application of a general method called Projection onto Convex Sets [13].

4.5 Time-Frequency Distribution Series

This approach for computing the time-frequency representation of a signal is based on the observation that a WVD of any signal can be written in the following series expansion [8]:

$$W(t,f) = \sum_{m_1,n_1} \sum_{m_2,n_2} C_{m_1,n_1} C_{m_2,n_2} W_{\phi_1,\phi_2}(t,f) \qquad (10)$$

where $C_{m,n}$ are coefficients of the Gabor expansion of the signal [8] and $W_{\phi_1,\phi_2}(t,f)$ is the cross WVD between two Gaussian modulated complex sinusoids. The character of the cross WVD is such that when the pairs (m_1,n_1) and (m_2,n_2) are very different then $W_{\phi_1,\phi_2}(t,f)$ is highly oscillatory. So that in a sense one can associated the interference terms in $W(t,f)$ with the terms in (10) where (m_1,n_1) and (m_2,n_2) differ greatly.

This suggests a strategy for reducing the interference terms based on truncating the series in (10). Specifically if (10) is amended so that only terms such that $|m_1 - n_1| + |m_2 - n_2| \le D$ then the effect of the interference terms is reduced. Choosing D to be large results in a distribution close to the WVD and choosing D to be small results in a distribution that is a special case of the local matching algorithm described in section 4.2.

5. RESULTS

To compare the performance of the algorithms described here we shall firstly consider a synthetic time series, which consists of two components; one a Gaussian impulse centred on the 50th sample and a normalised frequency of 0.3 and the second a frequency modulated component, whose envelop is also Gaussian. The frequency modulation used is a sinusoidal modulation with a period of 500 samples. The time series can be seen in Figure 2a). Figures 2b)-d) show spectrograms computed using windows of lengths 33, 81 and 257 samples, whilst Figures 2e) and 2f) show the WVD and the smooth pseudo-WVD (a widely used member of Cohen's class). The plots use a grey scale representation on a decibel scale, 60 dB dynamic range, darker colours represent larger values. For short windows the spectrogram represents the impulsive component well, but the character of the frequency modulated component is obscured and the converse happens when a long window is used. The most obvious interference terms in the WVD (Figure 2e)) can be seen as a grey area roughly centred on 150 samples and normalised frequency of 0.18. The smoothing operation employed to construct Figure 2e) can be seen to have significantly reduced the interference terms, but some loss of resolution can also be seen.

Figure 3 depicts the time-frequency representations computed using the techniques discussed in Section 4. Inspecting these representations one can draw several conclusions. The results of the re-assigned spectrogram are very encouraging, the two components have been well resolved. The local fitting algorithm produces a good fit to the impulsive component, but requires several components to match the frequency modulated term. This results in a significant loss of resolution. The optimal kernel algorithm resolves the frequency modulated term, but fails to resolve the impulse. This is due to the fact that the version of the algorithm used here estimated a single kernel for the whole time series. For this signal the use of a local kernel, one that changes with time, should produce better results. The optimal kernel scheme has also introduced artefacts at relatively low amplitudes. In practice these artefacts may be masked by the background noise. The Cohen-Posch algorithm yields a good result for this synthetic signal. The impulse is well resolved whilst the frequency modulated component is less well resolved. Finally the time-frequency distribution series has yielded results in which the resolution of the two terms is similar, regardless of their orientation, and in which the interference terms have been largely eliminated.

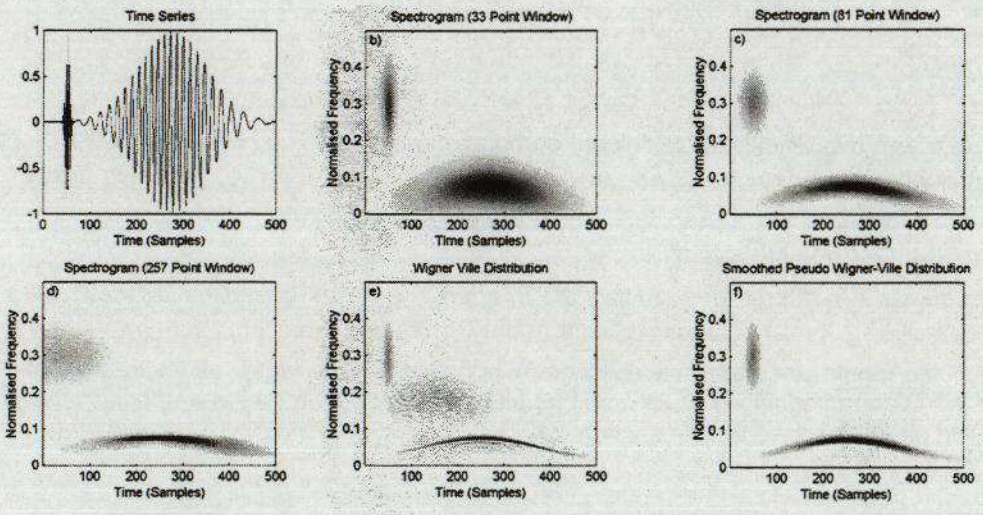

Figure 2: Synthetic Signal Analysed Using "Classical" Time-Frequency Methods

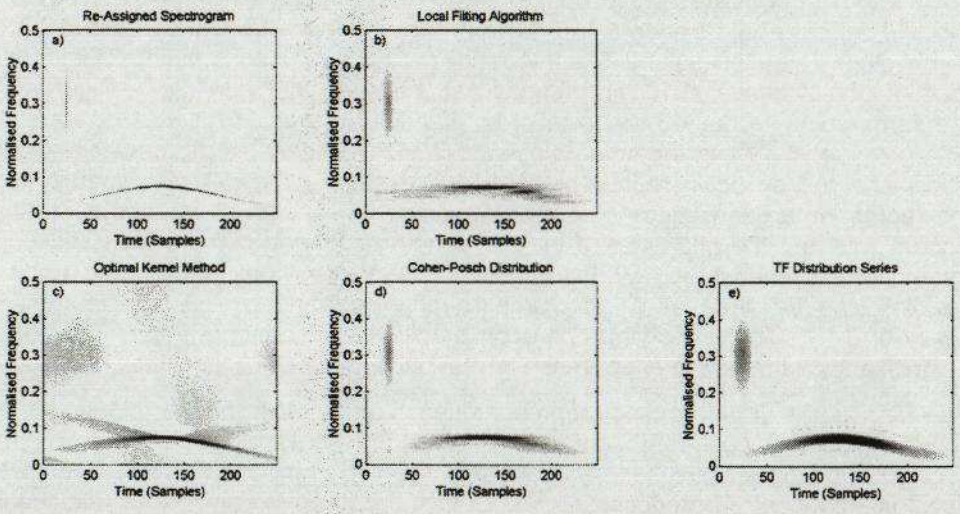

Figure 3: Results of Applying the Data Adaptive Time-Frequency Methods to the Synthetic Signal

Figure 4 shows the results of applying classical time-frequency methods to a time series containing two measured dolphin clicks (Atlantic white-sided dolphin, Lagenorhynchus dioptrica), this signal was captured using a sampling rate of 80kHz. The spectrogram using a 33 point window and the smoother pseudo-WVD, seem imply that the first click consists of two upward sweeping chirps and the second click consists of a single curved, nearly vertical component.

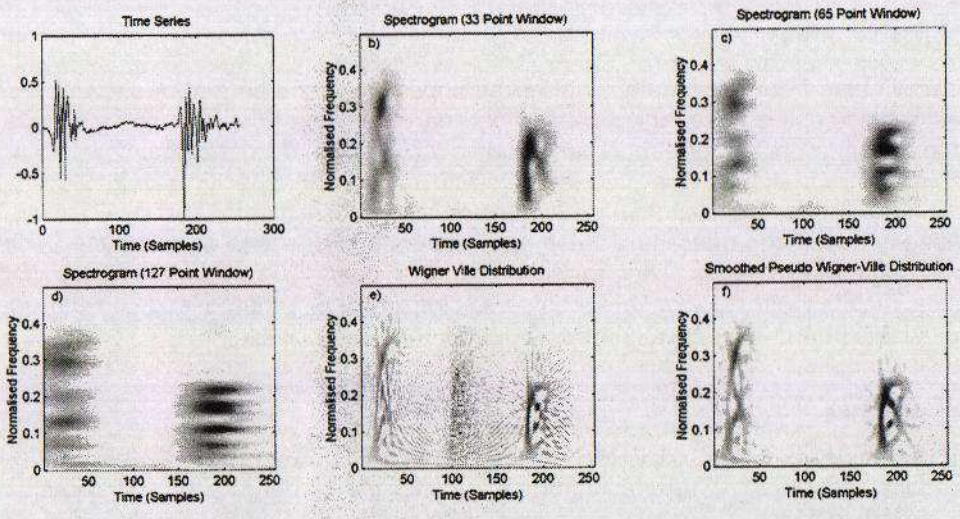

Figure 4: Dolphin Click Analysed Using "Classical" Time-Frequency Methods

Figure 5 depicts the data adaptive time-frequency representations for the signal shown in Figure 4a). The re-assigned spectrogram defines tight lines, although it is not clear that these are an appropriate characterisation of the click. The local fitting algorithm breaks the clicks into its constituent components, but the resulting representation fails to covey the signal's structure. The optimal kernel method produces a result similar to that obtained by the smoothed pseudo-WVD. The Cohen-Posch distribution has cheque board appearance this can be attributed to its imposition of the marginal constraints. Finally the TF distribution algorithm yields a representation similar to that obtained with the 65 point spectrogram, Figure 4c).

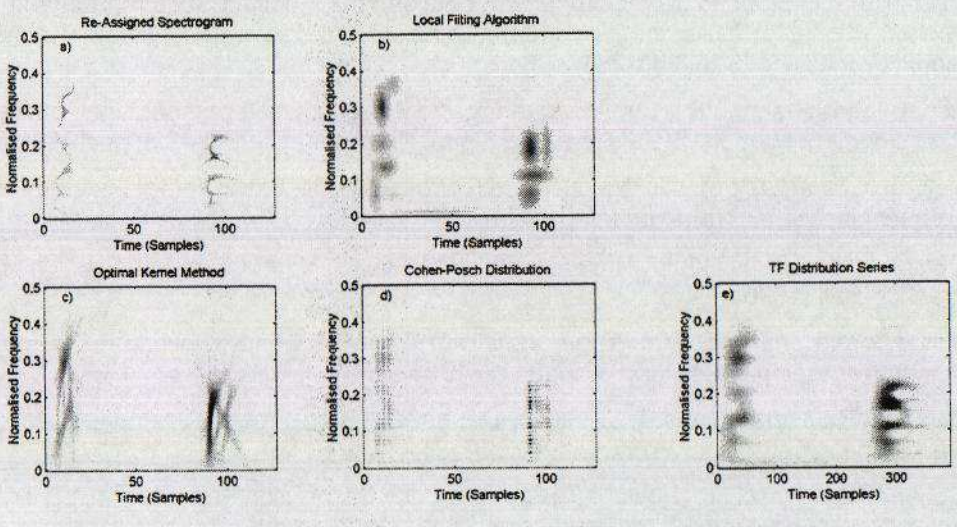

Figure 5: Dolphin Click Analysed Using Data Adaptive Time-Frequency Methods

6. CONCLUSIONS

Data adaptive time-frequency methods offer the opportunity for a time-frequency representation to adjust its characteristics to suit the signal being analysed. Classical methods require a degree of user intervention to obtain good representations over a range of input signal types. Herein results using only one real signal were presented in full and on this specific signal the data adaptive methods failed to offer significant improvements over classical methods. To demonstrate the true performance advantages offered by these approaches one needs to consider the performance of a fixed configuration over a wider range of signal types. The important comparison for this application requires one to design a complete classification scheme and to train and test that scheme using different time-frequency representations as inputs.

ACKNOWLEDGEMENTS

I would to thank Nick Felgate, DERA, Farnborough for supplying me with the data and Prof. S. Qian, National Instruments Corporation for software for the Time-Frequency Distribution series.

REFERENCES

[1] Gabor, D. "Theory of Communication" Journal of the IEE, 1946; 93: 3, 429-457.

[2] Cohen L. Time-Frequency Analysis pub. Prentice-Hall, 1995.

[3] Hlawatsch F. and Bourdreaux-Bartels G. "Linear and Quadratic Time-Frequency Signal Representations" IEEE Signal Processing Magazine, 1992, pp.21-67,.

[4] Kodera K., Villedary C. and Gendrin R. "A New Method for the Numerical Analysis of Non-Stationary Signals" Physics of the Earth and Planetary Interiors, 1976;12, pp.142-150.

[5] Moss J. and Hammond J. "A Comparison between the Modified Spectrogram and the Pseudo-Wigner-Ville Distribution with and without Modification" Mechanical Systems and Signal Processing, 8(3), 243-258, 1994.

[6] Auger F. and Flandrin P. "Time-Frequency and Time-Scale Representations by the Reassignment Method" IEEE Trans. on Signal Processing, 1996; 43:5; pp.1068-1089,.

[7] Delprat N. *et al* "Asymptotic Wavelet and Gabor Analysis: Extraction of Instantaneous Frequencies" IEEE Trans. on Information Theory 1992; 38:2), 644-664,.

[8] Quain S. and Chen D. Joint Time-Frequency Analysis: Methods and Applications pub. Prentice-Hall, 1996.

[9] Mallat S. and Zhang Z. "Matching Pursuit with Time-Frequency Dictionaries" IEEE Trans. on Signal Processing, 41(12), pp 3397-3415, 1993.

[10] Baraniuk R. G. and Jones D. L. "Signal-Dependent Time-Frequency Analysis Using a Radially Gaussian Kernel" Signal Processing, 32(3), 263-284, 1993.

[11] Cohen L. and Posch T. "Positive Time-Frequency Distribution Functions" IEEE Trans on Acoustics Speech and Signal Processing, 33(1), 31-38, 1985.

[12] Loughlin P., Pitton J., and Atlas L. "Construction of Positive Time-Frequency Distributions" IEEE Trans. on Signal Processing, 42(10), 2697-2705, 1994.

[13] Lopes, D. "Signal Reconstruction from Partiak or Modified Linear Time-Frequency Representations" Ph.D. Thesis, University of Southampton, 2000