UNIVERSITY OF Southampton
Institute of Sound and Vibration Research

SIGNAL PROCESSING & CONTROL GROUP

Wind Turbine Amplitude Modulation:
Research to Improve Understanding as to its Cause & Effect
Work Package WPB1 - Development of an Objective AM Measurement Methodology

The Measurement and Definition of Amplitude Modulations

by

Paul White

January 2012

Contents

Summary

There are a very large number of options in terms of how the degree to which is a signal is modulated might be measured and rated. These options cover the variety of ideas about what is to be measured, how the frequency dependence of the modulation is taken into account and what methods are used to extract parameters (or metrics) from the data.

Three main forms of methods were identified in this report, for application to the study of modulated noise from wind farms, with reference to similar techniques developed in other fields such as sonar research. These techniques all claim a degree of optimality and have been derived from basic principles, and are therefore considered more robust than methods requiring a degree of subjective evaluation.

Three forms of methods have been considered, and their performance evaluated on a range of representative artificial stimuli and field recordings of wind turbine noise. These three methods types considered and developed are:

1. Short term energy envelope analysis
2. Filter bank method
3. Periodic correlation method

The first method was found to work well, despite the theoretical limitations identified, although signal filtering was required in some cases to minimise the influence of secondary un-modulated noise sources, such as wind noise. The second method is similar to the first one but considers narrower components of the signal in isolation. It produced results which were smoother and somewhat more consistent in time, suggesting greater reliability, although some of the smallest modulation was sometimes not detected. Filtering of the part of the signal in which the modulation is significant was necessary in both cases, which effectively supposes this can be known *a priori*. The third method, although designed to work for a broader class of signals, failed in practice to detect modulation except for the strongest signals, and also had a significant computational load and so was not considered in further detail.

In each case, the magnitude of the modulation could be expressed by different values derived from the modulation spectrum produced by the analysis. The consideration of these potential AM magnitude metrics, resulting from the methods identified in this report, is separate to the consideration of their subjective meaning, which is investigated in other elements of the current research project.

1. Introduction

The work presented in this report is part of project funded by RenewableUK and entitled 'Wind Turbine Amplitude Modulation: Research to Improve Understanding as to its Cause & Effect'. The project comprises a total of six separate work packages. The outcome results of each of the work packages have separately resulted in their own dedicated final reports. A seventh work package, WPF, has produced an overarching final report in which the key findings across the separate work packages have been collated and discussed.

This is the final report of Work Package WPB2: '*Development of an Objective AM Measurement Methodology'*.

Wind turbine aerodynamic noise, by which is meant the noise produced by the rotating wind turbine blades, includes a steady component as well as, in some circumstances, a periodically fluctuating, or amplitude modulated (AM), component. However, AM may take different forms. One form of AM, commonly referred to as 'blade swish', is an inherent feature of the operation of all wind turbines. It can be explained by well understood mechanisms, it being the result of the directivity characteristics of the noise created by the air flowing over a turbine blade as it rotates. Because this type of AM is an inherent feature of the operation of wind turbines, whose origin can be explained and modelled, the present project adopts as its definition the term 'normal amplitude modulation' (NAM). The key driver for the project, however, is the recognition that some AM exhibits characteristics that fall outside those expected of NAM. Such characteristics include a greater depth of modulation, different directivity patterns or a changed noise character. For this reason the present project adopts as its definition the term 'other amplitude modulation', or 'OAM', for all observations of AM that lie outside that expected of NAM.

In recent years public concern has grown about the potential annoyance from wind turbine OAM noise. This concern has resulted in an increased interest to establish how AM, and in particular OAM, occurs, how it can be better defined and measured, and how it is generally perceived and responded to. It is the answers to these questions that the present project seeks to address.

This report specifically concentrates on the work considering appropriate metrics with which to measure amplitude modulation.

The observed fluctuations in wind farm noise are quasi-periodic, i.e. over short periods of time they fluctuations repeat themselves with an almost constant period. The study of modulated signals is an area of interest in a range of disciplines, but probably the area which has been the focus of most research activity is the field of communications, wherein modulations are used to code signals prior to transmission [1-5]. There are a variety of modulations which are used in communications systems, one of which is amplitude modulation that is, for example, used in older formats for radio transmission.

A second area, which is less widely studied but more closely related to this work, is passive sonar, i.e. listening to underwater sounds to detect, classify and localise acoustic sources. In particular, under certain circumstances, the propellers on vessels emit modulated sounds [6]. Methods for the detection of such sounds have been a primary tool for passive sonar for a great many years [7].

A similar area in which modulation processing is used is in the field of vibration based machine condition monitoring, in particular monitoring of bearings [8]. For example, imperfections in a bearing give rise to regular impulses as the fault makes contact with a surface. Within condition

monitoring the processing of amplitude modulated signals is referred to as envelop processing [9]. These fault related impulses vary on a cycle-by-cycle basis and are best modelled as a cyclostationary process, i.e. they can be modelled as amplitude modulated.

This report describes the various approaches to the objective measurement of modulations and then suggests a variety of metrics which can be assessed to determine which relate most closely with the results of subjective experiments.

2. Review of (Amplitude) Modulations in other Fields

Before considering the modulations in the context of wind farm noise this section will briefly discuss modulations in the fields of communications, passive sonar and condition monitoring. The open nature of research in the field of communications, allied to its everyday use, means that there more published material in this field. This is in contrast to the classified and rather specialist nature of passive sonar. Like communications, condition monitoring is an established research area in which processing based on amplitude modulation is an established methodology. This section initially reviews work in the area of communications and then describes the role of processing techniques based on amplitude modulation in passive sonar and condition monitoring. Finally it does briefly outline existing methods used in the analysis of wind turbine noise.

2.1 *Amplitude Modulated Communications Signals*

Historically one on the earliest forms of coding a signal prior to transmission through a medium is to use Amplitude Modulation (AM). In such applications the signal to be transmitted takes the form:

$$x(t) = A\left(1+\mu m(t)\right)\cos\left(2\pi f_c t + \phi_c\right) \tag{1}$$

where $m(t)$ is the message signal to be transmitted, f_c is the carrier frequency, which for a communications system is usually very much higher than any frequencies in the message signal, for example medium wave commercial radio stations use frequencies in the approximate range 100 kHz – 1 MHz to convey signals in the audio frequency band. The parameter A controls the overall amplitude of the signal, ϕ is the initial phase of the carrier signal and μ is often called the modulation index [2, 3, 5]. Typically the message signal has a large bandwidth relative to its centre frequency, i.e. it is a broadband signal such as music or speech. This is in contrast to the model we shall use for wind farm noise in which the modulation is quasi-periodic, i.e. relatively narrow-band, but the "carrier" signal is a broadband process.

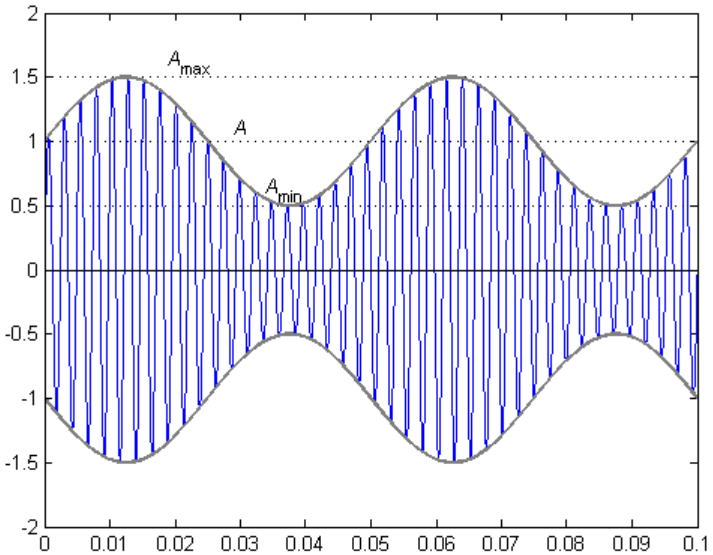

Figure 1: Sinusoidal modulation of a sine wave carrier signal. Blue line shows the modulated signal (centre frequency 410 Hz), solid grey line indicates the modulation function (frequency 20 Hz). A_{min} is the minimum amplitude, A_{max} the maximum and A the amplitude as defined in (1).

It is common practice in communications to consider the simple case where the message signal is itself sinusoidal, such that

$$m(t) = \cos\left(2\pi f_{\mathrm{m}} t + \phi_{\mathrm{m}}\right) \tag{2}$$

In such instances one can define the modulation index in terms of readily measureable quantities and it is tempting to consider these as the basis for metrics for wind-farm noise.

The minimum and maximum amplitudes are related to the modulation index μ as follows:

$$A_{\max} = A(1+\mu) \qquad A_{\min} = A(1-\mu) \tag{3}$$

There are several methods by which the modulation index can be measured. These include the modulation factor [4] defined as

$$M_{\mathrm{fact}} = \frac{A_{\max} - A_{\min}}{A_{\max} + A_{\min}} \tag{4}$$

and the percentage modulation [1]

$$M_{\mathrm{per}} = \frac{A_{\max} - A_{\min}}{2A}\,. \tag{5}$$

The percentage modulation is normally multiplied by 100 to yield a percentage from which its name derives. The factor of 100 has been omitted here for consistency with the other metrics.

For the simplified case of a sinusoidal message, as in (2), substituting the relationships in (3) into (4) and (5) one can see that both M_{fact} and M_{per} are equal to the modulation index μ. But when the message is not sinusoidal then the metrics (4) and (5) generally yield different values. Indeed for non-sinusoidal messages the concept of the modulation index itself becomes poorly defined.

The issues around defining the modulation index arise because there is no unique definition for the amplitude of an arbitrary message $m(t)$. To illustrate, consider a simple extension of (2) such that a single additional harmonic term is added to the message to give the following simplified formulation. This formulation does not include phase terms to avoid unnecessary algebraic cluttering.

$$x(t) = A\left(1+\mu\left(\alpha\cos\left(2\pi f_m t\right) + \beta\cos\left(4\pi f_m t\right)\right)\right)\cos\left(2\pi f_c t + \phi_c\right). \tag{6}$$

The message term in (6) has two degrees of freedom, but 3 free parameters (α, β and μ). Hence one can choose several equivalent parameterisations. Three example parameterisations are given below:

$$\begin{aligned}
&\mu\left(\alpha\cos\left(2\pi f_{\mathrm{m}} t\right) + \beta\cos\left(4\pi f_{\mathrm{m}} t\right)\right) \\
&= \mu_1\left(\cos\left(2\pi f_{\mathrm{m}} t\right) + \beta_1\cos\left(4\pi f_{\mathrm{m}} t\right)\right), \quad \mu_1 = \mu/\alpha,\ \beta_1 = \beta/\alpha \\
&= \mu_2\left(\alpha_2\cos\left(2\pi f_{\mathrm{m}} t\right) + \cos\left(4\pi f_{\mathrm{m}} t\right)\right), \quad \mu_2 = \mu/\beta,\ \beta_2 = \alpha/\beta \\
&= \mu_3\left(\alpha_3\cos\left(2\pi f_{\mathrm{m}} t\right) + \sqrt{1-\alpha_3^{\,2}}\cos\left(4\pi f_{\mathrm{m}} t\right)\right), \quad \mu_3 = \mu\sqrt{\alpha^2+\beta^2},\ \alpha_3 = \alpha/\sqrt{\alpha^2+\beta^2}
\end{aligned}\,. \tag{7}$$

In (7) the first parameterisation ensures that the fundamental component has an amplitude of one, whereas the second parameterisation ensures the harmonic's amplitude is one, whilst the third parameterisation arises from when the overall root mean square (rms) level of the signal is constrained to be one.

Each of the new formulations in (7) depends on two parameters, with the message term being normalised according to different criteria. The modulation indices (μ_1, μ_2 and μ_3) are all different with no single parameterisation having any theoretical advantages over the others. This ambiguity over what defines the modulation index for non-trivial message signals means that in order to use the modulation index one must specify precisely how the message term is normalised.

2.2 *Passive Sonar*

The noise from wind farms shares a closer relationship to the underwater noise from the propellers than to communication signals. This is not only because both are acoustic signals, but more importantly both are examples of broad-band processes modulated by narrow-band signals, as opposed to communications signals which are narrow-band signals modulated by broad-band signals.

For passive sonar the model (1) is a little misleading; it is more appropriate to adopt a model of the (simplified) form:

$$x(t) = \left(1 + \alpha \cos\left(2\pi f_{\mathrm{m}} t\right)\right) n(t). \tag{8}$$

In this case $n(t)$ is a broad-band noise process and α plays a role which is the equivalent of the modulation index.

In passive sonar the usual objectives are to detect the presence of a modulated signal and commonly to estimate the modulation frequency f_{m}. The most commonly used technique to detect and measure the modulation frequency is the so-called DEMON (DEModulation Of Noise) processing [7, 10]. It is not common practice in passive to measure the depth of modulation.

2.3 *Condition monitoring*

The use of envelope analysis, which is essentially the same as DEMON processing in passive sonar, is widespread in the field of condition monitoring. As in the case sonar case the primary objective for most condition monitoring systems is to determine whether there is an amplitude modulation, i.e. to detect a fault, and to determine the frequency of that modulation. In the case of bearings the modulation frequency can yield important information regarding the location of the fault.

2.4 *Monitoring of wind turbine noise modulation*

This Section concludes with an overview of some of the methods that have been proposed for the use of assessing wind turbine noise. These include methods described in [11-17], many of these techniques can be regarded as adaptions of the principles which are developed herein, such methods are referred to in the text in the appropriate location. These proposed techniques seek to exploit the inherently periodic nature of the AM from turbines. Many such methods exploit $\mathrm{LA}_{\mathrm{eq}}$ measurements, since these are readily obtainable from standard sound level meters. One method which differs from these techniques, and the other methods in this report, is outlined in [17] and is referred to as the "Den Brook" method. This is based on applying tests to short (2 s) blocks of $\mathrm{LA}_{\mathrm{eq,125\ ms}}$ data and looking for periods in which peak-to-trough levels differ by 3 dB(A). These periods are then counted over a

1 min period. This method fails to exploit the most striking feature of AM which is its periodic structure.

3. Formulations of Depth of Modulation

The acoustic signals we seek to characterise possess periodic fluctuations and the goal it to obtain a measure the strength of such modulations. A primary objective of this work is to develop a measure of the strength of such modulations which can then be correlated with the results of subjective tests. The quantity in question we refer to as the *depth of modulation*, which is intended as a generic term.

Based on the discussions in Section 2.1 one might consider using a method based on the modulation index in order to assess the depth of modulation, but such metrics are not the only options available. For example one can consider using the ratio

$$\frac{A_{\max}}{A_{\min}} = \frac{1+\mu}{1-\mu}. \tag{9}$$

This ratio is related to the modulation index via a monotonic function (in the region $0<\mu<1$). Whilst this ratio is an appropriate measure of modulation depth, it is not equal to the modulation index. This ratio has been proposed for use in measuring modulation in wind farms [13].

For an acoustic signal there are several domains in which the signals can be represented. The depth of modulation can be measured in any of these domains and an ideal metric would be insensitive to the choice of domain. The specific domains one can consider for an acoustic signal include: the linear pressure and the Sound Pressure Level (SPL). In terms of wind-farm noise several authors have considered using the L_{eq} (or its A-weighted equivalent) which is a time-averaged for of the SPL [11, 14], where that temporal averaging is conducted in accordance with a standardised method.

In order to illustrate the issues around defining the depth of modulation consider a simple example of a signal formed using (1), illustrated in Figure 2. The message signal, $m(t)$, is assumed to be a sinusoid, as in (2), in this example the value of μ is 0.5. The data are referred to as pressure, for this numerical example there is no precise physical meaning intended. In each instance the maximum amplitude, $A_{\max}$, minimum amplitude, $A_{\min}$ and the time averaged amplitude, A, are shown. Note that in Figure 2 c) the value of A shown is the time average of the dB value, which is not what constitutes perceived good practice in acoustics, the "correct" approach is to time average the squared values and then convert that to a dB scale. In the following discussion these two averages will both be discussed and will be denoted A_{dB} when the averaging is performed on the dB values and A_{lin} when the averaging is performed on the squared pressure.

The following considers three measures for the depth of modulation, each of which has been proposed for measuring wind farm noise. The first is the ratio of the maximum and minimum amplitudes, as in (9), which will be denoted R_{mm}, which is expressed on a dB scale in the different domains as follows:

Pressure: $20\log_{10}\left(\frac{A_{\max}}{A_{\min}}\right)$, Squared pressure: $10\log_{10}\left(\frac{A_{\max}}{A_{\min}}\right)$, dB domain: $A_{\max} - A_{\min}$

The second is the ratio of the maximum value to the time value, denoted R_{ma}, which is expressed as:

Pressure: $20\log_{10}\left(\frac{A_{\max}}{A}\right)$, Squared pressure: $10\log_{10}\left(\frac{A_{\max}}{A}\right)$, dB domain: $A_{\max} - A$

There are two forms for R_{ma} in the dB domain depending upon which form of averaging is employed, i.e. whether A_{dB} or A_{lin} is used. These two options will be denoted (inelegantly) as $R_{\mathrm{ma,lin}}$ and $R_{\mathrm{ma,dB}}$.

Finally a measure which is the ratio of half the difference between the maximum and minimum value and the time averaged value, denoted here as R_{mma} and expressed as

Pressure: $20\log_{10}\left(\frac{A_{\max}-A_{\min}}{2A}\right)$, Squared pressure: $10\log_{10}\left(\frac{A_{\max}-A_{\min}}{2A}\right)$,

There is no simple direct equivalent for R_{mma} in the decibel domain, since it is based on subtraction in a linear domain.

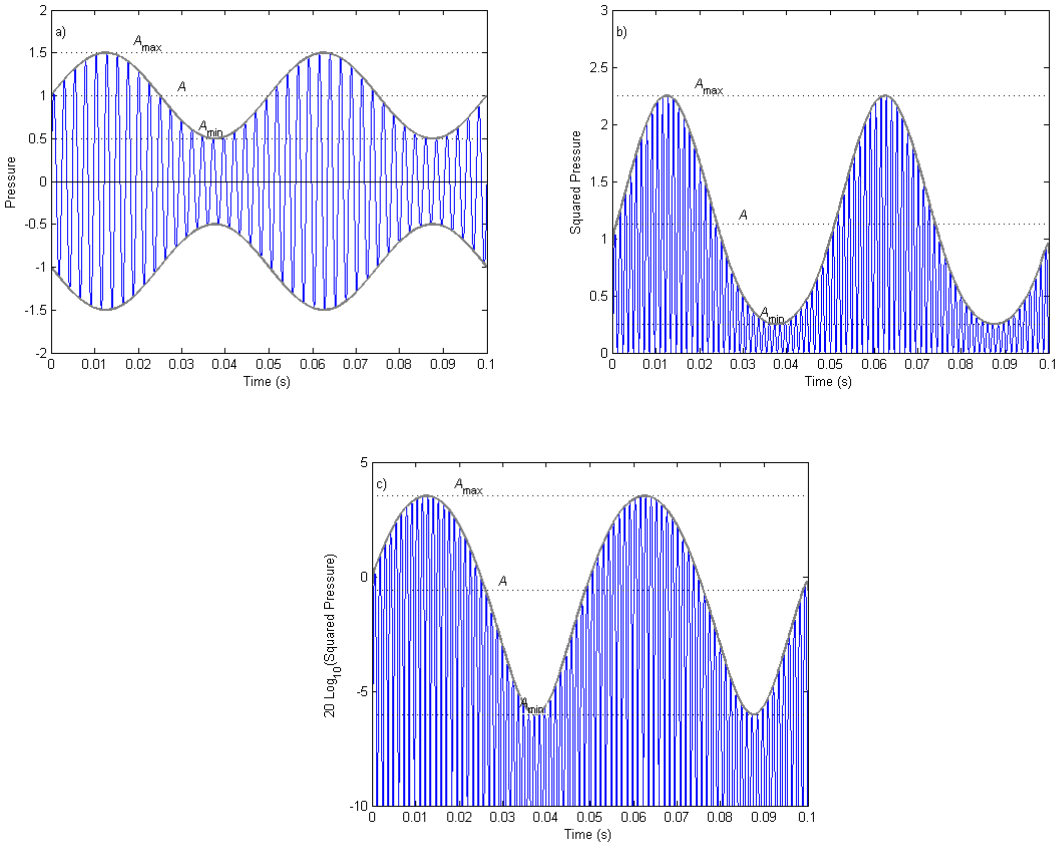

Figure 2: Modulations in various domains. a) Pressure wave form, b) Squared pressure waveform and c) Pressure on a (unreferenced) decibel scale

The results of applying the various metrics to the data shown in Figure 2 are shown in Table 1.

	Pressure (dB)	Squared Pressure (dB)	dB Domain (dB)
R_{mm}	9.5	9.5	9.5
R_{ma}	3.5	3.0	$R_{\text{mma,lin}}$: 1.0 $R_{\text{mma,dB}}$: 4.1
R_{mma}	-6	-0.5	N/A

Table 1: Table of metrics for depth of modulation for different

From Table 1 it is clear that the different metrics generally yield different values when computed in different domains. The metrics, R_{ma}, R_{mm} and R_{mma} aim to compute subtle different quantities so one might reasonably expect that they generate different values. One also needs to appreciate that different values are (generally) obtained when the metrics are computed in the different domains. The exception is R_{mm} which is consistent between all domains; a property that can be regarded as a significant benefit. However, measuring the minima and maxima in a time series, as is proposed in the "Den Brook" method [17], is prone to greater uncertainty than measuring a time averaged quantity.

These are not the only similar reasonable metrics which can be considered. There are few physical principles which allow one to justify the use of one metric in preference to another. For this project this judgement may be based on how well the metrics correlate with subjective preference, and how they can be applied to realistic signals, rather than any physical arguments.

3.1 *Effect of bandwidth*

A final consideration when assessing the depth of modulation is the analysis bandwidth. Measures of depth of modulation are always ratios. Loosely they relate the magnitude of the fluctuations to the steady background level. Commonly for wind farm noise the majority of the modulated sound occurs in a low frequency band, for the sake of this discussion assume the band is 100 Hz - 1 kHz. The measured depth of modulation will depend on the bandwidth of the recording processed. For example, if in the above case the acoustic pressure is measured only over the band from 100 Hz to 1 kHz then all of the energy in the modulated signal is captured. Whereas if the measurement is made over a wider band, say 1 Hz to 44 kHz, then this signal includes noise from the frequency band which is distinct from the modulated band. This additional noise serves to mask the modulation and so results in a lower depth of modulation. To further illustrate this point consider the signal divided into a set of frequency bands, the depth of modulation in each band is different and depends on the signal to noise ratio of the modulation in that particular band. In practice modulated energy is spread across a range of frequencies and the signal to noise ratio in each will be different. Determining which bands should and should not be included in a measure of depth of modulation is not a trivial problem. However to ensure that a consistent metric is achieved this selection should be performed according to objective criteria.

4. Sinusoidally Modulated White Noise Model

The initial approach adopted here is to consider models for modulations and fit those models to measured data sets. In doing so one explicitly obtains estimates for the parameters of the model from which an estimate of the depth of modulation can be extracted. This requires one to consider the form of the model to be fitted. Selecting the form of the model usually entails a compromise. As the complexity (and the realism) of the models is increased, the problem of fitting that model to the data becomes more challenging and the associated algorithms used typically require a commensurate increase in computational resources.

This section considers the simplest (non-trivial) model of a modulated random process, based on (8): white Gaussian noise modulated by a sine wave. This model is not expected to be a particularly accurate representation of noise from wind farms: in particular the fact that the signals are white, so have a flat spectrum, is not realistic. However, the resulting algorithm developed for this signal forms the basis of a more general approach described later.

A discrete time model for the signal, $x(n)$, can be written as

$$x(n) = w(n)\left(1+\mu\cos\left(2\pi f_0 n+\phi\right)\right) = w(n)\left(1+\alpha\sin\left(2\pi f_0 n\right)+\beta\cos\left(2\pi f_0 n\right)\right) \tag{10}$$

where $w(n)$ is a Gaussian white noise process with variance σ^2. In terms of acoustic signals this serves as a potential model for the acoustic pressure, since the model anticipates that the amplitude of signal will be symmetrically distributed about zero.

Assuming that $\alpha^2+\beta^2=\mu^2<1$, ensures that the amplitude of oscillation is such that the signal's variance is never zero, i.e. there is always some stationary (unmodulated) noise present.

The formulation (10) differs from the early representation (8) in three ways: it incorporates an initial phase term, ϕ, in the modulation, the model is expressed in terms of discrete time (*n* instead of *t*) and, consequently, the modulation frequency is denoted as f_0, as opposed to f_m, to indicate that it now represents a normalised discrete time frequency.

The instantaneous variance of $x(n)$ is given by

$$\sigma_x^2(n) = \sigma^2\left(1+\alpha^2\sin^2\left(2\pi f_0 n\right)+\beta^2\cos^2\left(2\pi f_0 n\right)+(\alpha+1)\beta\sin\left(2\pi f_0 n\right)+(\alpha+1)\beta\cos\left(2\pi f_0 n\right)\right)$$

$$\sigma_x^2(n) = \sigma^2\left(1+\frac{\alpha^2+\beta^2}{2}+(\alpha+1)\beta\sin\left(2\pi f_0 n\right)+(\alpha+1)\beta\cos\left(2\pi f_0 n\right)+\frac{\beta^2-\alpha^2}{2}\cos\left(4\pi f_0 n\right)\right) \tag{11}.$$

According to (10) and (11) it is evident that a process which is sinusoidally modulated in terms of linear quantities will be not sinusoidally modulated in terms of its variance. Emphasising that the choice of domain in which the modulation is measured generally affects the values obtained for the depth of modulation.

It transpires that adopting a sinusoidal model for the modulation of the variance, as opposed to a sinusoidal modulation for the signal itself, (10), leads to a mathematically more tractable solution. Accordingly an alternative model for modulated white noise is to assume that the variance which is sinusoidally modulated, i.e.

$$\sigma_x^2(n) = \sigma^2\left(1 + a\sin\left(2\pi f_0 n\right) + b\cos\left(2\pi f_0 n\right)\right) \tag{12}$$

Note that the variance parameter σ^2 in (12) is not equivalent to that in (11), although these parameters play essentially the same role, which is why they are represented by the same symbol here.

In order to estimate the depth of modulation from a set of measured data one needs to fit the model (12) (or (11)) to a data set, this allows one to estimate the parameters *a* and *b* which will form the basis of a measure of depth of modulation.

This fitting can be performed in a variety of ways but a common approach which is asymptotically efficient is to use the principle of maximum likelihood [18, 19]. Maximum likelihood is a general principle used to fit models to data and has a long history dating back to R.A. Fisher in 1912 [20]. It requires one to find the model which maximises $p(\mathbf{x}|\theta)$ [1] (the probability density function of the data given the model parameters, which is termed the likelihood). Loosely speaking the method of maximum likelihood aims to identify the parameters which are most likely to have given rise to the data. One can argue that it is more logical to seek to identify the most likely set of parameters given the data, i.e. to maximise $p(\theta|\mathbf{x})$. This can indeed be achieved by invoking the principles of Bayesian statistics [21] and modifying the maximum likelihood estimator accordingly. However in this work we shall consider the maximum likelihood approach.

The application of the maximum likelihood principle to the estimation of the parameters of modulated white noise was first detailed by [10] and the following description mirrors that development.

For a Gaussian process the likelihood, $l(x(n))$, for a single sample $x(n)$ is given by

$$p(x(n)|a,b,f_0) = l(x(n)) = \frac{1}{\sigma_x(n)\sqrt{2\pi}} e^{-\frac{x(n)^2}{2\sigma_x^2(n)}} \tag{13}$$

The likelihood for a set of *N* samples uncorrelated samples, arranged in a column vector $\mathbf{x} = \left[x(0), x(1), \cdots, x(N-1)\right]^t$ can be written as

$$p(\mathbf{x}|a,b,f_0) = l(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\prod_{n=0}^{N-1}\sigma_x(n)} e^{-\sum_{n=0}^{N-1}\frac{x(n)^2}{2\sigma_x^2(n)}} \tag{14}$$

Forming the logarithm of the likelihood, i.e. the log likelihood, $L(\mathbf{x})$, one obtains

$$L(\mathbf{x}) = \log\left(l(\mathbf{x})\right) = -\frac{N}{2}\log(2\pi) - \sum_{n=0}^{N-1}\log\left(\sigma_x(n)\right) - \frac{1}{2}\sum_{n=0}^{N-1}\frac{x(n)^2}{\sigma_x^2(n)} \tag{15}$$

The optimal (maximum likelihood) estimates for the modulation parameters *a*, *b* and f_0 are obtained by maximising (15) with respect to these parameters. The first term on the right-hand side of (15) is constant with respect to the modulation parameters, so does not need to be considered in the maximisation process.

[1] Where $\mathbf{x}$ represents the set of measurements and θ represents the model parameters.

The second term $\sum_{n=0}^{N-1} \log\left(\sigma_x(n)\right)$ is the sum of an oscillating expression. Assuming *N* is sufficiently large, so that the summation extends over many cycles of the oscillations, it is approximately constant with respect to *a*, *b* and f_0 [10].

Hence to approximately maximise (15) one can consider maximising

$$L(\mathbf{x}) = -\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x(n)^2 \left(1 + a\sin\left(2\pi f_0 n\right) + b\cos\left(2\pi f_0 n\right)\right)^{-1} \tag{16}$$

where the model for a sinusoidal modulation of the variance, (12), has been included. Following the approach taken by Lourens and du Preez [10] and assuming that the modulation is small so that both *a* and *b* are small, i.e. $|a|,|b| \ll 1$, one can use the approximation

$$\begin{aligned} L(\mathbf{x}) &\approx -\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x(n)^2 \left(1 - a\sin\left(2\pi f_0 n\right) - b\cos\left(2\pi f_0 n\right)\right) \\ &= \frac{1}{2\sigma^2}\left[-\sum_{n=0}^{N-1} x(n)^2 + a\sum_{n=0}^{N-1} x(n)^2 \sin\left(2\pi f_0 n\right) + b\sum_{n=0}^{N-1} x(n)^2 \cos\left(2\pi f_0 n\right)\right] \end{aligned} \tag{17}$$

The optimisation of (17) presents an impasse as there is no analytical solution for any of the unknown parameters, *a*, *b* or f_0. Lourens and du Preez [10] suggest the adoption of the following estimates for the amplitude parameters *a* and *b*. They show that these estimators are unbiased but fail to show that they are optimal in any sense. This is a short-coming of their method and one can reasonably argue that it undermines their claim to have developed a maximum likelihood estimator.

The estimates of *a* and *b* use by Lourens and du Preez are denoted $\hat{a}$ and $\hat{b}$ respectively and are given by:

$$\begin{aligned} \hat{a} &= \frac{2}{N\sigma^2}\sum_{n=0}^{N-1} x(n)^2 \sin\left(2\pi f_0 n\right) \\ \hat{b} &= \frac{2}{N\sigma^2}\sum_{n=0}^{N-1} x(n)^2 \cos\left(2\pi f_0 n\right) \end{aligned}. \tag{18}$$

Using these approximate amplitudes, and neglecting terms which are independent of f_0, one can see that the maximum likelihood estimate of f_0 is obtained by maximising

$$\Psi(f_0) = \left(\sum_{n=0}^{N-1} x(n)^2 \sin\left(2\pi f_0 n\right)\right)^2 + \left(\sum_{n=0}^{N-1} x(n)^2 \cos\left(2\pi f_0 n\right)\right)^2 = \left|\mathrm{F}\left\{x^2(n)\right\}\right|^2 \tag{19}$$

where $\mathrm{F}\{\ \}$ denotes the operation of taking the Fourier transform.

Consequently the (approximate) maximum likelihood estimator of frequency is obtained by seeking the largest peak in the Fourier transform of the square of the signal. The amplitude of the modulation is obtained using the modulation amplitudes $\hat{a}$ and $\hat{b}$, with the overall amplitude being computed using $\sqrt{\hat{a}^2 + \hat{b}^2}$. Note that these estimated amplitudes are obtained from the real and imaginary parts of the Fourier transform $\mathrm{F}\left\{x(n)^2\right\}$.

An alternative derivation of (18) has been developed as part of this work so that the optimality of the solution is clear and does not rely upon the incomplete development found in [10]. This new derivation is based on a higher order approximation of (16). The algebraic representations in this development is simplified if the following vector notation is adopted

$$\begin{aligned} L(\mathbf{x}) &= -\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x(n)^2 \left(1+\mathbf{a}^t\boldsymbol{\theta}(n)\right)^{-1} \\ \mathbf{a} &= \begin{bmatrix} a & b \end{bmatrix}^t \quad \boldsymbol{\theta}(n) = \begin{bmatrix} \sin(2\pi fn) & \cos(2\pi fn) \end{bmatrix}^t \end{aligned} \quad . \tag{20}$$

The paper [10] used an approximation for $\left(1+\mathbf{a}^t\boldsymbol{\theta}(t)\right)^{-1}$ based on its series expansion, specifically to develop (17) from (16) the following linear approximation is used

$$\begin{aligned} \left(1+\mathbf{a}^t\boldsymbol{\theta}(t)\right)^{-1} &= 1-\mathbf{a}^t\boldsymbol{\theta}(t)+\left(\mathbf{a}^t\boldsymbol{\theta}(t)\right)^2-\left(\mathbf{a}^t\boldsymbol{\theta}(t)\right)^3+..... \\ &\approx 1-\mathbf{a}^t\boldsymbol{\theta}(t) \end{aligned} \tag{21}$$

The following development uses one more term in the expansion, specifically

$$\left(1+\mathbf{a}^t\boldsymbol{\theta}(t)\right)^{-1} \approx 1-\mathbf{a}^t\boldsymbol{\theta}(t)+\left(\mathbf{a}^t\boldsymbol{\theta}(t)\right)^2 \tag{22}$$

Which, when substituted into (16), this leads to

$$\begin{aligned} L(\mathbf{x}) &\approx -\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x(n)^2 \left(1-\mathbf{a}^t\underline{\theta}(n)+\mathbf{a}^t\boldsymbol{\theta}(n)\boldsymbol{\theta}(n)^t\mathbf{a}\right) \\ &= -\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x(n)^2 + \frac{\mathbf{a}^t}{2\sigma^2}\left\{\sum_{n=0}^{N-1} x(n)^2\boldsymbol{\theta}(n)\right\} - \frac{\mathbf{a}^t}{2\sigma^2}\left\{\sum_{n=0}^{N-1} x(n)^2\boldsymbol{\theta}(n)\boldsymbol{\theta}(n)^t\right\}\mathbf{a} \end{aligned} \tag{23}$$

Accordingly the optimal amplitudes are obtained using

$$\begin{aligned} &\frac{\partial L(\mathbf{x})}{\partial \mathbf{a}} = \frac{1}{2\sigma^2}\left\{\sum_{n=0}^{N-1} x(n)^2\boldsymbol{\theta}(n)\right\} - \frac{1}{\sigma^2}\left\{\sum_{n=0}^{N-1} x(n)^2\boldsymbol{\theta}(n)\boldsymbol{\theta}(n)^t\right\}\mathbf{a} = 0 \\ &\Rightarrow \mathbf{a} = 2\left\{\sum_{n=0}^{N-1} x(n)^2\boldsymbol{\theta}(n)\boldsymbol{\theta}(n)^t\right\}^{-1}\left\{\sum_{n=0}^{N-1} x(n)^2\boldsymbol{\theta}(n)\right\} = \mathbf{R}^{-1}\mathbf{p} \end{aligned} \tag{24}$$

The matrix **R** can be simplified by examining its individual elements. Specifically

$$\begin{aligned} &\mathbf{R} = \left\{\sum_{n=0}^{N-1} x(n)^2\boldsymbol{\theta}(n)\boldsymbol{\theta}(n)^t\right\} \qquad R_{1,1} = \sum_{n=0}^{N-1} x(n)^2\sin^2(2\pi fn) \\ &R_{1,2} = R_{2,1} = \sum_{n=0}^{N-1} x(n)^2\sin(2\pi fn)\cos(2\pi fn) \quad R_{2,2} = \sum_{n=0}^{N-1} x(n)^2\cos^2(2\pi fn) \end{aligned} \tag{25}$$

These elements of **R** can be approximated as

$$\begin{aligned} &R_{1,1} = R_{2,2} \approx \frac{N\sigma^2}{2} \qquad R_{1,2} = R_{2,1} \approx 0 \\ &\mathbf{R} \approx \frac{N\sigma^2}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \Rightarrow \quad \mathbf{R}^{-1} \approx \frac{2}{N\sigma^2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{26}$$

Hence

$$\mathbf{a} = \begin{bmatrix} a \\ b \end{bmatrix} = \frac{2}{N\sigma^2} \begin{bmatrix} \sum_{n=0}^{N-1} x(n)^2 \sin(2\pi fn) \\ \sum_{n=0}^{N-1} x(n)^2 \cos(2\pi fn) \end{bmatrix} \tag{27}$$

This agrees with the result (18) proposed (without a demonstration of optimality) by Lourens and du Preez [10]. Thus the above development confirms that the original solution did indeed represent an approximate maximum likelihood solution, but the development presented here provides a more sound theoretical basis for that result.

The resulting likelihood value is given by

$$L(\mathbf{x}) = -\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} x^2(n) - \frac{\Psi(f_0)}{N\sigma^2} \right). \tag{28}$$

In order to estimate the depth of modulation the values of σ^2 and f need to be estimated and applied in (27) to obtain values of **a**. As represented by (19) the modulation frequency, f_0, is estimated by locating the maximum of the Fourier transform of the square of the signal. This general approach has been applied in the field of wind farm noise by several authors [11, 12, 14]. The value of σ^2 is obtained by noting from the model (12) that σ^2 represents the time averaged energy, i.e. the overall rms value of the signal, and as such can be estimated in a standard manner.

4.1 *Variations of the white noise model*

The above model can be regarded as the basis of many approaches. There are some simple variations that might be considered. These are very briefly discussed in the follow subsections.

4.1.1 Temporal smoothing of the energy

For instance one might consider using temporal averaging of $x(n)^2$ over some time window prior to taking its Fourier transform. This form of averaging is that employed when computing L_{eq}s. Such averaging removes high frequency components in $x(n)^2$. This is identical to attenuating high frequency values of the cost function $\Psi(f_0)$, see (19). For such an algorithm smoothing offers little benefit for the applications of interest here, since the modulation rates for wind farm noise are low-frequency and so are unaffected by such smoothing, unless that smoothing is extreme and acts over periods which are significant relative to a period, i.e. of several hundreds of ms. There is a practical advantage that can accrue from such smoothing, namely that after smoothing the data can be down-sampled to a lower sampling rate and hence the computational burden of the method can be reduced [15]. However, in its basic form this is a very efficient algorithm and computational loading is not normally a significant factor when considering this routine.

4.1.2 Filtering prior to computation of the energy

In many recordings the modulation only affects energy over a limited bandwidth. It is an obvious extension to the method described in Section 4 to apply some form of filtering to $x(n)$ prior to

computation of $x(n)^2$. Such filtering, if performed appropriately, will serve to concentrate the processing on the particular band of interest so can enhance system performance by rejecting out-of-band noise.

This filtered version of the algorithm forms the basis of DEMON processing in passive sonar and envelope processing in condition monitoring, see Sections 2.2 and 2.3 respectively.

One can opt to use a filter which is designed for the purposes of imposing a perceptual weighting, for example an A-weighting filter. This introduces a weighting which reflects the perceived importance of the modulated spectrum, it does not account for the perceptual effects of the modulation itself.

If an A-weighting is performed then this method is essentially a technique which is based on the Fourier transform of LA_{eq}, an approach which has been proposed [11, 14].

4.1.3 Using the analytic form of the signal

Some authors have suggested employing the squared magnitude of the analytic form of the signal [12, 22] (sometimes called the pre-envelope [23]) rather than simply squaring the signal. The analytic signal, $\kappa(t)$, which is complex valued, is created from the original (real-valued) signal using the Hilbert transform. Specifically

$$\kappa(t) = x(t) + i\mathrm{H}\{x(t)\} \tag{29}$$

where $i = \sqrt{-1}$ and H represents the Hilbert operator such that

$$\mathrm{H}\{x(t)\} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{30}$$

The analytic form of a signal is commonly used in a variety of signal processing applications based around narrow-band signals, e.g. [24, 25]. Indeed in condition monitoring it is common practice to process the pre-envelope, which is why such process in that field is referred to envelope processing, see Section 2.3.

There are distinct advantages to processing pre-envelope, $|\kappa(t)|^2$, rather than $x(t)^2$ in applications where the signals are narrow-band, but for wind farm noise the pre-envelope and the squared signal produce very similar results. To illustrate this Figure 3 compares the Fourier transform of the square of the signal and the Fourier transform of the pre-envelope for Gaussian white noise. One can see that in the low frequency region the two spectra are very similar. Typical wind farm modulation frequencies are very low-frequency, i.e. in the region where the two results are almost identical, suggesting there is no benefit to be gained by employing the analytic signal.

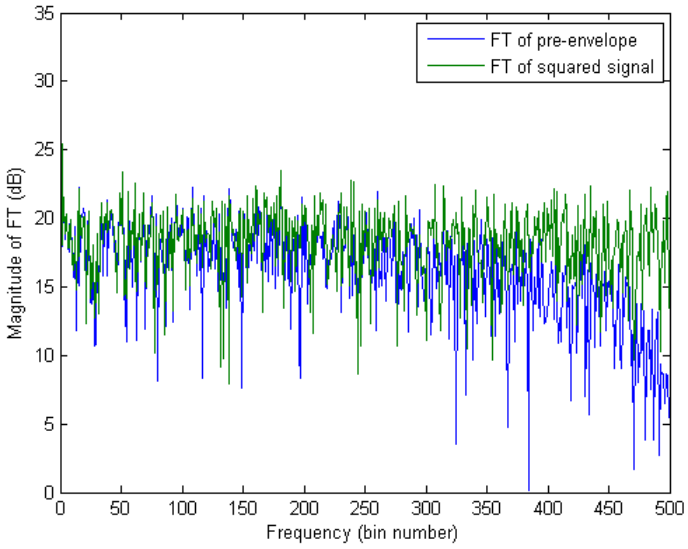

Figure 3: Fourier Transforms based on 1000 samples of Gaussian white noise: Blue - Fourier transform of the squared magnitude of the analytic signal (the pre-envelope), Green - Fourier transform of the square of signal.

4.1.4 Harmonics of the modulation frequency

In practice modulations are never exactly sinusoidal. The consequence is that the Fourier transform of the energy generates a set of peaks at harmonic frequencies. This can be regarded as energy which has leaked away from the fundamental frequency and so reduced its amplitude making it less detectable.

The problem of analysing harmonic periodic signals has been studied [26] and the optimal solution in that instance is to form the harmonogram which is created using the Fourier transform of the data. Specifically for each hypothesised frequency the sum of the Fourier transform at that frequency and the energy at the assumed set of harmonics is formed. In practice this method does introduce some practical difficulties since it requires one to define the number of harmonics believed to be present in the signal and the summing process tends to introduce a set of spurious peaks which can lead to incorrect estimation of the fundamental frequency.

Accepting the limitations of the harmonogram, which are (arguably) fundamental issues that are a consequence of assuming a model with a set of harmonics, then this approach can be extended to modulations. For detecting the presence of modulations the algorithm described in this section can be simply extended by including a summation over the assumed set of harmonics, specifically the new cost function, $\Lambda(f_0)$, based on that found in (19) is obtained as

$$\Lambda(f_0) = \sum_{k=1}^{K} \Psi(kf_0) \tag{31}$$

where K is the assumed number of harmonics. Typically because of the need for the K^{th} harmonic not to exceed the Nyquist frequency means that the cost function $\Lambda(f_0)$ would be evaluated for frequencies in the range $0 < f_0 < f_s/(2K)$.

The function defined by (31) is an optimal estimator of the modulation frequency and is a candidate metric for defining the depth of modulation for non-sinusoidal modulations (modulations with harmonics) it is not the only measure. The problem of determining what metric should be used for non-sinusoidal modulations is non-trivial. There are several potential approaches. One might consider using the amplitude (or power) of the largest single component, normally the fundamental frequency. Such a method effectively neglects the presence of harmonics and treats all modulations as if they were sinusoidal. Alternatively one might employ the total power in the modulation function, which leads to a method based on (31). Finally one might seek to use a metric based on the amplitude of the modulation waveform, for example forming the ratio of the modulation peak to its trough. Such a metric could be computed by reconstructing the modulation waveform the measured harmonics, but requires estimation of the modulation phase as well as amplitude at the harmonic frequencies.

5. Modulated Coloured Gaussian Noise

Measured wind farm noise usually consists of (at least) two parts: a stationary noise source and a modulated noise, neither of which is normally white. Consider the following model of the spectrum of such a noise.

$$S_{xx}(t, f) = S_{nn}(f) + S_{mm}(f)\left(1 + \sin\left(2\pi f_{\mathrm{m}} t + \phi\right)\right) \tag{32}$$

where f_{m} is the modulation frequency and ϕ is the initial phase of the modulation, $S_{xx}(f)$ and $S_{nn}(f)$ are the spectra of the measured signal and the stationary noise respectively. The quantity $S_{mm}(f)$ is the time averaged spectrum of the modulated process (the averaging occurring over many cycles of the process).

5.1 *Defining modulation depth for coloured noises*

In addition to the discussions in Section 3 regarding how to define the depth of modulation, the model (32) introduces an additional confounding factor into such definitions. The ratio $R(f) = S_{mm}(f)/S_{nn}(f)$ is a function of frequency which controls the degree to which modulation appears at a particular frequency. To extract a single value representing modulation depth across all frequencies one needs to consider how to combine the values to yield a single metric.

Assuming the ratio $R(f)$ can be computed (we shall shortly consider how this ratio might be achieved) then one approach is to compute a value by forming a weighted average across frequency in the form

$$M_d = \int_0^B W(f) R(f)\, df \tag{33}$$

where B is the bandwidth of the measured signal (for a digital signal this would be half the sample rate) and $W(f)$ is some weighting function. Choosing different forms for the weighting function generates a range of "reasonable" methods. For instance an A-weighting function might be considered to partly account for the frequency dependent characteristics of the human auditory system. Alternatively one might consider a flat (or uniform) weighting function. Such an un-weighted measure of depth of modulation might be applying to studies which are more based on understanding the physical properties of the acoustics field, as opposed to the human perception of that field (where an A-weighted value might be deemed more appropriate).

To compute the ratio $R(f)$ the spectra $S_{nn}(f)$ and $S_{mm}(f)$ need to be estimated separately. This can be achieved in a number of ways, one of which is to first compute the time averaged value of, $S_{xx}(t, f)$, which we denote, $\bar{S}_{xx}(f)$, so that, from (32), one can write

$$\bar{S}_{xx}(f) = \frac{1}{T}\int_0^T S_{xx}(t, f)dt = S_{nn}(f) + S_{mm}(f) \tag{34}$$

In some circumstances the stationary source may be measured in isolation (for instance the modulation strength might be time varying and there could be periods when the modulated signal is absent from recordings) allowing direct estimation of $S_{nn}(f)$. In cases when only measurements containing both sources are available then one can employ order statistics to estimate $S_{nn}(f)$ [27]. A similar approach, using minimum statistics, has been proposed for estimating the background noise spectrum in speech processing [28], however, employing a more general order statistics based method allows the estimation of $S_{nn}(f)$ with greater accuracy [27]. Once $\bar{S}_{xx}(f)$ and $S_{nn}(f)$ have been determined then using (34) one can compute $S_{mm}(f)$, or the ratio $R(f)$ directly using

$$R(f) = \frac{\bar{S}_{xx}(f)}{S_{nn}(f)} - 1 \tag{35}$$

In practice the ratio may be negative at some frequencies because of estimation errors and one should set any such values to zero. Combining this with (33) leads to

$$M_d = \int_0^B W(f)\frac{\bar{S}_{xx}(f)}{S_{nn}(f)}df - 1 \tag{36}$$

in which it has been assumed that $\int W(f)df = 1$ which would normally be the case for a suitable weighting function.

5.2 *The multi-band approach to coloured noise*

One solution to the problem of analysing modulations in which the noise and the modulated process are both coloured, i.e. to fit model (32) to data, is to adapt the method described in Section 4 [29]. This is based on the observation that if a signal is filtered around a sufficiently narrow frequency band then that signal can be approximated as white over the bandwidth of the filter. This approach requires one to apply a bank of filters to divide the signal into a set of narrow-band components. Each of these narrow-band signals can then be analysed under the assumption that they are modulated white noise

processes immersed in white noise, as in Section 4. The results from each band can then be combined in order to create an overall output.

Formerly, if the filter bank consists of K filters then applying it to the signal $x(t)$ creates a set of signals $x_1(t), x_2(t), \ldots, x_K(t)$, representing the outputs of each of the filter banks. If one assumes that each of the filter outputs is independent then the combined likelihood can be written as

$$L(\mathbf{x}) = \sum_{k=1}^{K} L(\mathbf{x}_k) \tag{37}$$

where the vectors $\mathbf{x}$ and $\mathbf{x}_k$ represent the set of samples from the signals $x(t)$ and $x_k(t)$ respectively. The assumption of independence is generally not valid for modulated processes since the modulation tends occur across multiple frequency bands, so that the levels in the bands are inter-dependent, hence the signals in those bands also have inter-dependences. Measuring and correcting for the dependence between bands requires considerable effort and adaptations to account for this have not been developed. Methods for detecting amplitude modulation by exploiting the fact that band will be co-modulated, i.e. there will be correlation between different bands [16]. This requires one to determine which bands to compute the correlation between and requires that more than one band be modulated, which, in the low SNR limit, will, in general, not be true. For the development here the assumption of independence underlying (37) is invoked, it is done so on a pragmatic basis rather than a theoretical one.

Based on (28) and employing (37) one can write

$$L(\mathbf{x}) = -\frac{1}{2}\sum_{k=1}^{K} \frac{1}{\sigma_k^{\,2}}\left(\sum_{n=0}^{N-1} x_k(n)^2 - \frac{\Psi_k(f_0)}{N\sigma_k^{\,2}}\right) \tag{38}$$

where the subscript k is used to denote quantities computed within the k^{th} sub-band. Eliminating the terms in (38) which do not depend on the model parameters means that the problem of estimating the modulation frequency reduces to

$$\arg\max_{f_0} \sum_{k=1}^{K} \frac{\Psi_k(f_0)}{\sigma_k^{\,4}} = \arg\max_{f_0} \sum_{k=1}^{K} \frac{\left|\mathrm{F}\left\{x_k(n)^2\right\}\right|^2}{\sigma_k^{\,4}} \tag{39}$$

Note that if the signal is pre-whitened, such that $\sigma_k^{\,2}$ is the same for all frequency bands then (39) can be simplified further to

$$\arg\max_{f_0} \sum_{k=1}^{K} \left|\mathrm{F}\left\{x_k(n)^2\right\}\right|^2 \tag{40}$$

Using (39) or (40) the modulation frequency can be estimated. Having determined the modulation frequency the amplitude of the modulation $\mathbf{a}_k$ can be computed for each sub-band, using a band equivalent form of (27). Finally an overall measure of depth of modulation can be obtained by combining the results from each of the individual bands. The ideas in Section 5.1 also applying in this case and the combination can be extend to include some weighting function, the most obvious choice for which is the A-weighting function.

This approach is not specific to any form of filter bank, the output of each filter is incorporated into (39) or (40). There are obvious candidate forms of filter bank which might be considered including constant bandwidth, octave, 1/3rd octave and perceptual filter banks.

However some forms of filter bank can be more efficiently implemented than others. For example if a constant bandwidth filter bank is used then the implementation can be achieved via a spectrogram. Since spectrograms can be implemented via the fast Fourier transform (FFT) they are computationally efficient. Similar an octave filter bank can be efficiently implemented using the wavelet transform [30] and this would provide a route to a suitable practical implementation of such a solution. Indeed this approach has been adopted by Lee *et al*. [15] for the analysis of wind farm noise, albeit that they did not combine data across frequency bins to yield a single metric.

6. Methods based on a periodic correlation model

Amplitude modulated signals fall within a broader class of signals, namely those which have periodic correlation functions [31]. Specifically if we consider the definition of a correlation function

$$r(t,\tau) = \mathrm{E}\left[x(t)x(t+\tau)\right] \tag{41}$$

where $\mathrm{E}[\ ]$ denotes the expectation operator and $r(t,\tau)$ is the time-dependent correlation function [24, 32]. This definition of a correlation function is the general one that does not rely upon an assumption of stationarity (modulated processes are non-stationary). A periodically correlated process is one for which

$$r(t+T,\tau) = r(t,\tau) \quad \forall t,\tau \tag{42}$$

where *T* is the period of the process. Using this model Kirsteins *et al*. [31] developed a maximum likelihood estimator for the modulation frequency (or more generally the frequency associated with the periodicity in the correlation function). To obtain the solution requires considerable effort, but it leads to an estimator which is computed by evaluating

$$\arg\max_{f_0} \Gamma(f_0), \qquad \Gamma(f_0) = \sum_{\tau=0}^{\tau_{\max}} \left| \sum_{m=0}^{N-1} x(m)x(m+\tau)e^{-2\pi i f_0 m\Delta} \right|^2 \tag{43}$$

where Δ is the sampling interval and is equal to $1/f_s$ in which f_s represents the sampling frequency. It is worth noting that in [31] a more general model containing harmonics is considered and the resulting algorithm is a combination of (31) and (43).

The quantity $\Gamma(f_0)$ is not particularly amenable to ready interpretation. To provide some understanding consider inner-most summation, which can be expressed as:

$$e^{\pi i f_0 \tau\Delta} \sum_{m=0}^{N-1} x(m)e^{-\pi i f_0 m\Delta} x(m+\tau)e^{-\pi i f_0 (m+\tau)\Delta} \tag{44}$$

Comparing this to the cross-correlation function for discrete complex signals, $r_{uv}(\tau)$, which is defined as

$$r_{uv}(\tau) = \frac{1}{N}\sum_{n=0}^{N-1} u(n)^* v(n+\tau) \tag{45}$$

where * denotes complex conjugation. Then by defining

$$u(n) = x(n)e^{\pi i f_0 n\Delta} \quad \text{and} \quad v(n) = x(n)e^{-\pi i f_0 n\Delta} \tag{46}$$

one can write (43) in the form

$$\Gamma(f_0) = \sum_{\tau=0}^{\tau_{\max}} \left| r_{uv}(\tau) \right|^2 \tag{47}$$

The signals $u(n)$ and $v(n)$ which have been introduced in this development, themselves have a straightforward interpretation. Specifically $u(n)$ represents the original signal up-shifted in frequency by $f_0/2$, whereas $v(n)$ represents the signal down-shifted by the same amount. The quantity $\Gamma(f_0)$ then represents the sum of the square of this cross-correlation function across all lags.

Equation (47) provides a computationally efficient method for evaluating $\Gamma(f_0)$. This because the correlation function can be evaluated using FFTs and the fact that $u(n)$ and $v(n)$ are conjugates of each other (see (46)) means that only one FFT needs to be evaluated.

7. Results

These methods are initially demonstrated on one of the test signals used in the listening tests conducted by the University of Salford Acoustics Research Centre as part of Work Package B2 of this project. These test signals are constructed using the method outlined in the appendix to this report. In Section 7.1 the performance on a single file is considered with the performance over a broader range of signals being considered in Section 7.3. The following section presents results obtained from a set of field recordings of amplitude modulated sounds made in the vicinity of wind turbines.

7.1 *Performance of methods on a single test file*

The results here have been obtained by applying the methods to the complete stimulus data sets which are of 20.25 s duration. In practice one would expect to apply these methods in a short-time framework wherein the data is divided in to small windows (or blocks). Each block is then processed individually. The block length needs to be selected according to the stationarity of the modulation process. Generally longer windows will yield more accurate parameter estimates, assuming the modulation is consistent (stationary) throughout the block. Excessively long block sizes may lead to parameter estimates which are unrepresentative, since the characteristics of the modulation may evolve within the block.

The file chosen for the initial assessment is a recording of stimulus 21 used in the final listening tests (target modulation 6 dB(A))[33]. This data contains modulations that are clearly discernible and are relatively strong for this particular data set. By using data with clearly discernible modulations one can understand the features that such modulations produce. Figure 4 shows time series data for the chosen data.

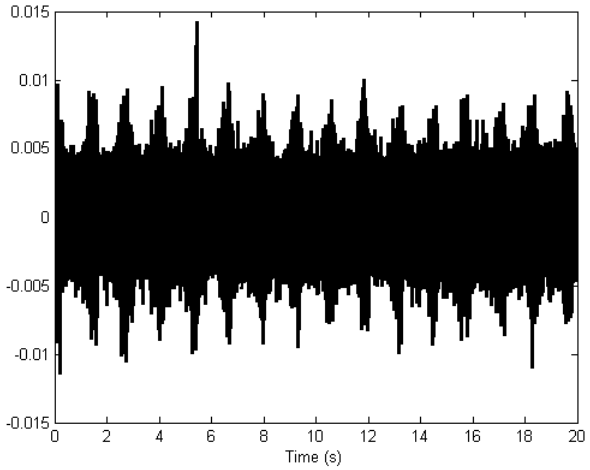

Figure 4: Time series data for stimulus 21

From Figure 4 the oscillations in amplitude can be clearly seen. Counting the number of cycles within the 20 s section of data suggests there are approximately 16 cycles of the modulation process within that period. This suggests a modulation frequency of approximately 0.8 Hz.

Figure 5 shows the result of computing the short time energy for two instances, in the first where there is no temporal smoothing and in the second case the data is smoothed using a Hanning windowing function of 100 ms duration. The quantities in Figure 5 are essentially L_{eq} measurements.

Visual examination of these energy time-series suggests that the smoothed curve (the lower frame) the modulation function is more apparent than in the unsmoothed data (the upper frame). This might lead one to suspect that there is an advantage to processing the smoothed data as opposed to the unsmoothed sequence; in fact there is comparatively little advantage since the noise affecting the unsmoothed data primarily lies in a frequency band which is very different to the modulation rate, as discussed in Section 4.1.1. This is illustrated in Figure 6**.**

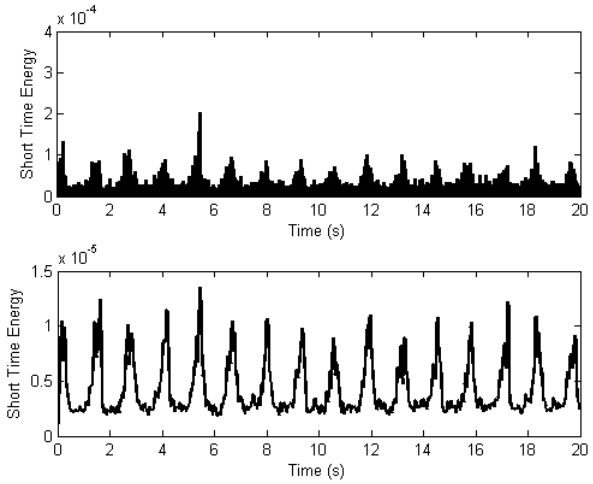

Figure 5: Short time energy. Upper frame: no smoothing, lower frame: smoothing with a Hanning window of 100 ms duration.

Since we are interested in detecting periodic oscillations in plots like those shown in Figure 5 then it is natural to consider taking Fourier transforms of those traces: which is precisely what the theory outlined in Section 4 suggests is optimal if the signals are white noise processes. Taking Fourier transforms of the data in Figure 5 yields the results shown in Figure 6 (note these transforms are shown on a reduced frequency scale 0-30 Hz to focus the reader's attention on the region where modulation rates and their harmonics are likely to appear for wind turbine data).

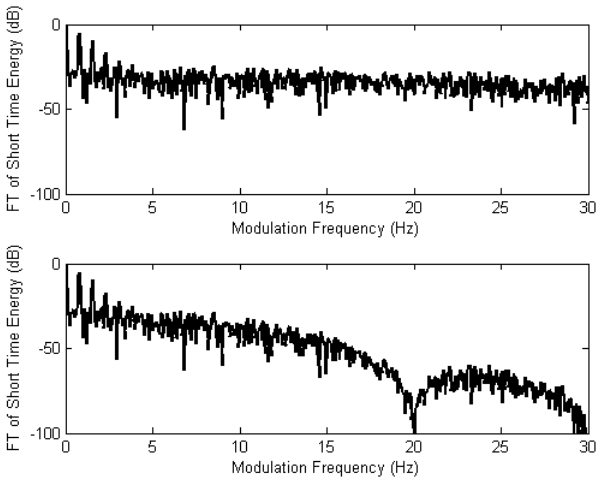

Figure 6: Fourier transforms of the short time energies shown in Figure 5. Upper frame: no smoothing, lower frame: smoothing with a Hanning window of 100 ms duration.

Comparing the two frames in Figure 6, one can see that the effect of smoothing is to attenuate the data in the high modulation rate region. In this case above a frequency of about 10 Hz; which corresponds to the reciprocal of the smoothing period (100 ms). In the band 0-10 Hz where the majority of components for the wind farm modulations are likely to appear the smoothing has no effect.

Consequently if one is processing these short time energy traces in the frequency domain (as is optimal, by virtue of the arguments in Section 4) then applying smoothing to the short time energies is of little benefit.

In order to examine the peaks in Figure 6 (upper frame) greater detail Figure 7 shows the same data on an expanded frequency axis, in the range 0-10 Hz. There is a large value at 0 Hz. This is a consequence of the fact that the average value in the short time energy (Figure 5) is not zero, this peak is a measure of the average energy in the time-series.

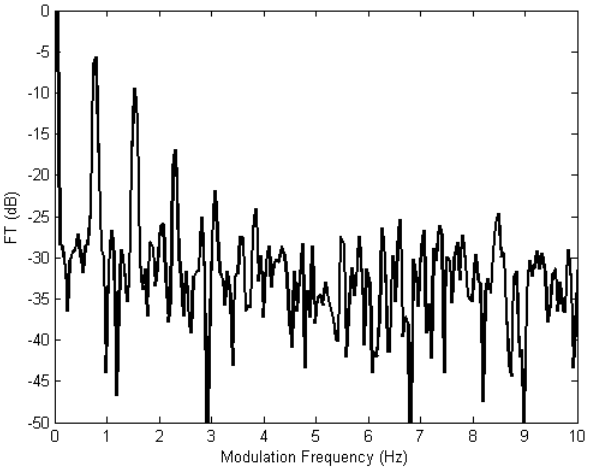

Figure 7: Fourier transform of short time energy for data shown in Figure 4, on an expanded frequency axis.

The peak at a frequency of a little below 1 Hz is measured to be at a modulation frequency 0.79 Hz which corresponds closely with the value estimated from the raw time series. This peak represents a fundamental frequency for which the harmonics are also evident (at least four are readily observable in this data set). This harmonic set indicates that the modulation is not sinusoidal: this is apparent in the short time energy traces (Figure 5 in particular the lower frame) where the shape of the energy waveform is more peaked than one expects from a sinusoidal signal. The value of the largest peak represents the power of the modulation only at the fundamental frequency. Alternatively summing three or four harmonic peaks which can be observed in the plot, would yield an estimate of the total modulation power, which could be used as single metric. As stated previously such a metric would not equate to a trough-to-peak measure of AM.

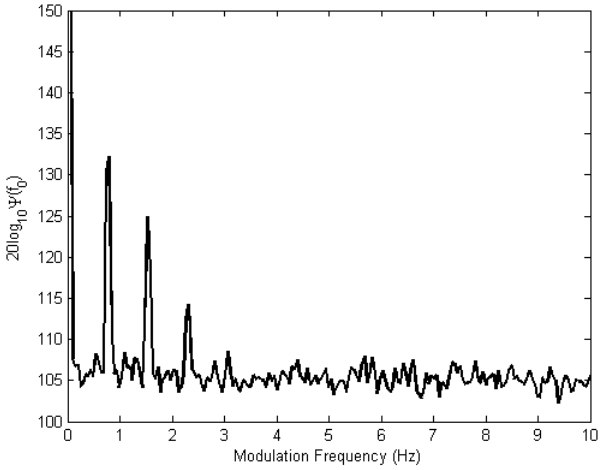

Figure 8: Output of the filter bank method for data shown in Figure 4

The filter bank method discussed in Section 4.3 is applied to this same data and the result is shown in Figure 8. In this instance a constant bandwidth filter bank has been applied, using a short-time Fourier transform (spectrogram) and a uniform weighting function used. This result is noticeably less variable (noisy) that the corresponding data for the Fourier transform of the short time energy, as seen in Figure 7. In Figure 8 one can again see a series of peaks, the fundamental frequency is measured to be 0.79 Hz (in agreement with other measures) and a series of harmonics can once again be observed. Despite the reduced variability in the data seen in Figure 8 it is not clear that more harmonics can be seen here than in Figure 7. This suggests that whilst the filter bank method produces a plot that is more visually appealing it is not clear that it in fact contains more useful information.

The reason for the reduced variability in the filter bank approach would seem to derive from the fact that it is the result of averaging data across many frequency bands.

Finally the method of Kirsteins, as summarised by (43), is applied to the time series in Figure 4. This routine is particularly slow on long data sets. This signal is stored at a sample rate of 48 kHz, whereas the majority of the energy is at a much lower frequency. Further this method is particularly sensitive to noise, even if that noise is in a frequency band which is distinct from that where the modulation lies. Hence before applying this method, the data is down-sampled to a sample rate of 4 kHz. The result is shown in Figure 9.

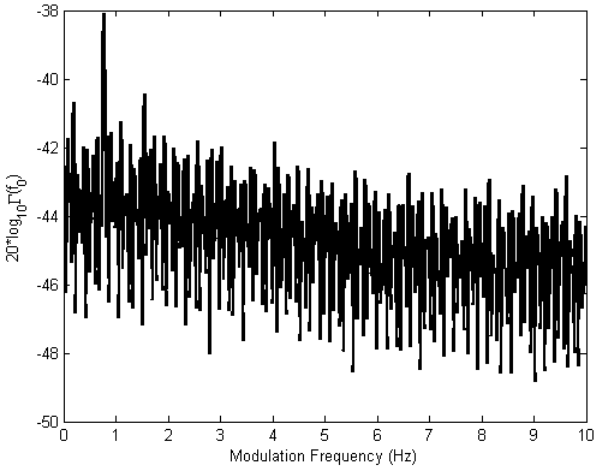

Figure 9: Result of applying Kirsteins' algorithm to the time series data in Figure 4, after it has been down-sampled to 4 kHz.

From Figure 9 one can see that there is a definite peak at approximately the correct frequency. In this case the peak is actually measured to be at 0.77 Hz, somewhat different to that obtained from the other methods. The resolution of this approach is not clear, the spectrum shown is evaluated with a frequency spacing of 0.01 Hz, so the observed difference in estimated modulation rate corresponds to only a few sample points.

In this plot only one harmonic component in the data can be observed at about 1.5 Hz. Away from the main peak there appears to be considerable noise with significant oscillations.

7.2 *Performance across a range of modulation depths*

The subset of the Salford stimuli data set considered consists of times series with different levels of modulation. The recordings are named with numbers (13-24) and the higher numbered files are those with greater levels of modulation (target modulation depth (MD) of 1 to 12 dB(A)). The following sub-sections consider the analysis of four such files, in particular those numbered 13, 17, 20 and 24 (respectively target MD of 1, 3.5, 5 and 12 dB(A)) and document the results on the whole set.

In these tests in order to facilitate a fair comparison between the methods all the sample sets are down-sampled by a factor of 12 to make them compatible with the results for Kirsteins' method.

7.2.1 Measures based on Fourier analysis of short time energy

Figure 10 shows the results of Fourier transforming the short time energy for each of the four signals. In each case a peak close to 0.79 Hz can be identified: for the lowest level of modulation (File 13) the peak value in the spectrum is at 0.74 Hz and in all the other cases it is at 0.79 Hz. This method creates a set of frequency bins of width, $\Delta f=1/T$, where T is the signal's duration, in this instance this means that $\Delta f=1/20.25=0.049$ Hz, so the difference between the result for File 13 and the others corresponds to a single frequency bin.

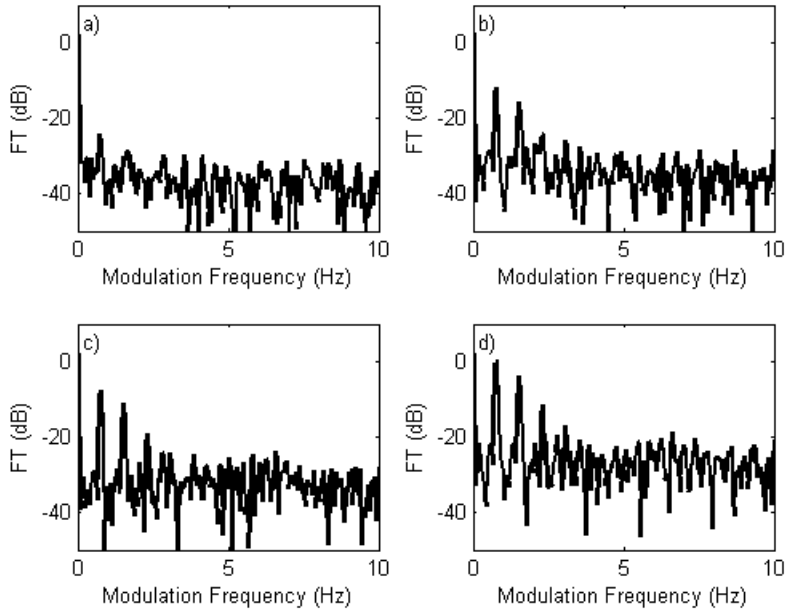

Figure 10: Fourier transforms of energy traces for four test files a) File 13, b) File 17, c) File 20, d) File 24

As the modulation level increases then the height of the peak also increases. The height of this peak relates to the amplitude of the fundamental frequency in the modulation. So one might consider using the following as a metric of depth of modulation, which is denoted D_M.

$$D_M = 10\log_{10}\left(\max\left(\Omega(f_0)\right)\right), \qquad \Omega(f_0) = \left|\mathrm{F}\left\{x(t)^2\right\}\right|^2 \tag{47}$$

The maximum is taken over a limited frequency range to encompass realistic modulation rates, in these simulations the range used is 0.15 Hz[2] to 20 Hz.

Further, as discussed in Section 4.1.2 then one can apply a filter to $x(t)$ prior to squaring it, so that the processing focuses on the frequency band where the modulation is strongest. In this instance a low-pass filter with a cut-off at 800 Hz is applied. This filter captures energy in the modulated band. The results for the four example signals are shown in Figure 11.

Comparing the unfiltered and filtered results (Figure 10 and Figure 11 respectively) there are only very small differences in the plots. This is largely because the signals actually contain comparatively little energy above 800 Hz, so the filter's effect is quite small. Had the recordings contained more high frequency noise one might anticipate that the effect of the filter would have been more noticeable.

[2] The value 0.15 Hz corresponds to 3 cycles in the measurement period (20.25 s).

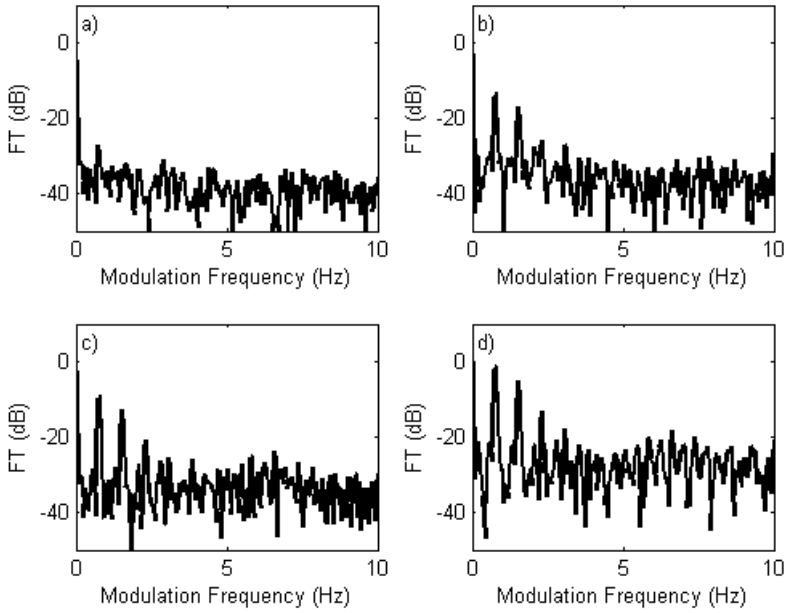

Figure 11: Fourier transforms of energy traces, computed after the data is low-pass filtered at 800 Hz, for four test files a) File 13, b) File 17, c) File 20, d) File 24.

The metric (47) is a reasonable measure when comparing similar data sets, but is not scale invariant, i.e. if the signal's amplitude is altered, then this measure of depth of modulation will change. It seems intuitive that depth of modulation is scale invariant, i.e. it is a relative measure. In order to accommodate this data is pre-scaled prior to processing. This is achieved by computing the mean value of the smallest 25% of values of $x(t)^2$. The idea being that the smaller values of $x(t)^2$ occur at times when the modulation is small, so that this measure is a measure of the stationary noise component. The choice of 25% involves a compromise: using a larger percentage would mean that data is included which contains significant contributions from the modulated energy is present, using a lower percentage increases the variability of the estimated energy.

Table 2 summarises the results of applying this metric to the full range of stimuli considered. The measure of modulation depth (D_M) shows an increasing trend for higher stimulus numbers, which is consistent with the manner in which the stimulus is constructed. The results obtained on the filtered data are somewhat greater than those obtained without the filter: they tend to be consistently 2 dB greater. One might speculate that this might convert into an ability to detect modulations in noise at lower Signal-to-Noise Ratios (SNR), but there is no direct evidence of this.

Stimulus Number	D_M (dB) Unfiltered	D_M (dB) Low Pass Filtered (800 Hz)
13	0.8	2.3
14	4.3	6.7
15	8.8	10.6
16	12.6	14.6
17	13.8	15.5
18	16.3	18.0
19	17.0	19.0
20	18.7	20.5
21	20.9	22.6
22	23.4	25.0
23	27.3	28.9
24	32.7	34.0

Table 2: Results of the metrics based on Fourier analysis of the short time energy to the Salford stimulus set

7.2.2 Measures based on filter bank analysis

As in Section 6.2.1 the range of stimuli considered is analysed using the filter bank method. Figure 12 shows the results of applying the filter bank method to the four examples from the Salford data.

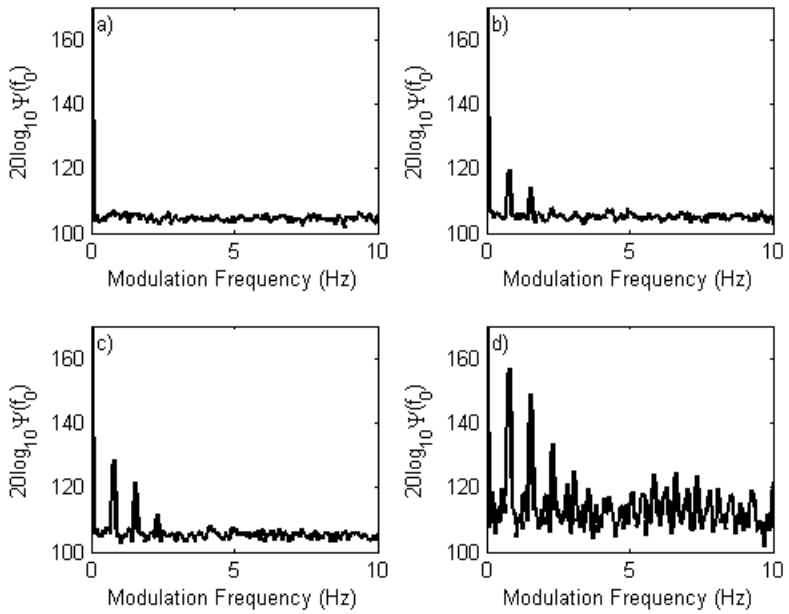

Figure 12: Output of the filter bank method for four test files a) File 13, b) File 17, c) File 20, d) File 24

From Figure 10 one can clearly see a peak close to 0.79 Hz in each instance, with the exception of the first example (frame a)). In that case in fact the maximum value occurs at 0.98 Hz, it is very close to the noise floor an so does not produce a discernible peak: there is also a smaller visually imperceptible peak at 0.79 Hz. Consequently we conclude that this method has failed to detect the modulation in the first example (File 13). Further analysis of these stimuli reveals that there is a detection made in File 14, i.e. the largest peak occurs at 0.79 Hz when that files is analysed.

Since the peak height increases with modulation depth then we consider using this as the measure of depth of modulation, as outline in Section 7.2.1. The data is normalised by the energy estimated from the lowest quartile of the data. Applying this method to the full data set one obtains the results shown in Table 3. Once again the general trend is for increasing measures of modulation depths with increasing stimulus number.

Stimulus Number	D_M (dB)
13	***6.2***
14	8.6
15	12.3
16	17.0
17	18.9
18	21.3
19	22.4
20	24.9
21	29.2
22	32.8
23	34.9
24	38.8

Table 3: Results of the metrics based on the filter bank method applied to the Salford stimulus set, the value in bold italics represents the point at which the detection failed.

7.2.3 Measures based on Kirsteins' method

The procedure described in the previous two sections is repeated using Kirsteins' method, i.e. the method described by (43). Firstly the method is applied to the four example data sets and the results are shown in Figure 13.

It is evident from Figure 13 that this method fails to detect the modulation for the first two examples: Files 13 and 17, since no clear peaks are generated. In the two examples where there is a clear peak (Files 20 and 24), it appears at a modulation frequency of 0.77 Hz.

This method is also used to compute estimates of the modulation depth for all of the files, in manner mirroring that of the previous sections. The results of this analysis are shown in Table 4. Note that the values in italics in that table indicate peaks at frequencies which are different to the modulation rate for this data, i.e. the algorithm is not detecting the modulation. This algorithm only detects the modulation in 5 of the recordings, compared to the other methods which only fail to detect the modulation on a single occasion.

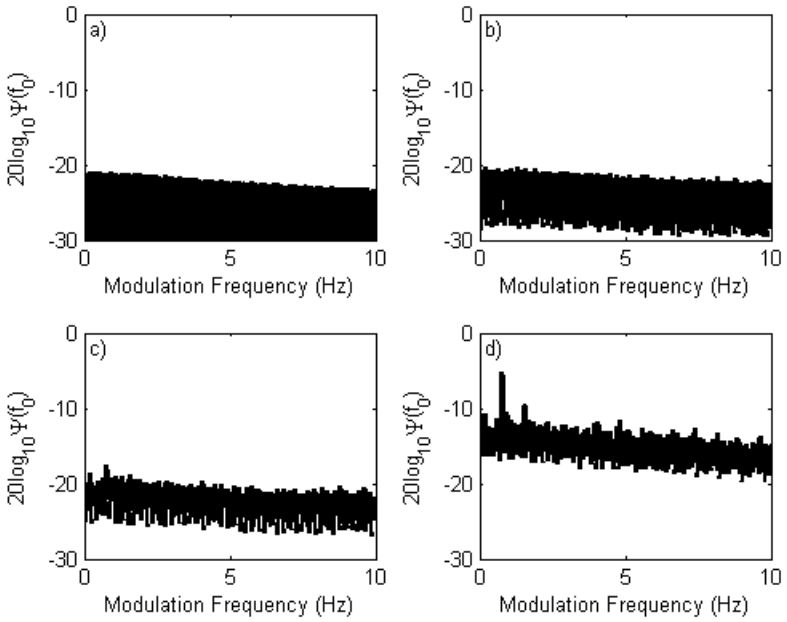

Figure 13: Output of Kirsteins' method for four test files a) File 13, b) File 17, c) File 20, d) File 24

Stimulus Number	D_M (dB)
13	***17.0***
14	***15.8***
15	***15.6***
16	***14.8***
17	***13.9***
18	***13.6***
19	***13.4***
20	15.4
21	17.9
22	20.8
23	23.2
24	27.0

Table 4: Results of the metrics based on Kirsteins' method applied to the Salford stimulus set, the values in bold italics represent the points at which the detection failed.

7.3 *Performance on field data*

This section discusses performance of the various algorithms on recordings from wind farms as opposed to the synthetic stimulus data. These recordings were provided as an output from Work Package C of the current project [34]. Some of these recordings are considerably longer than the stimulus data, see Table 5. This requires modification to the way in which the algorithms are applied to data sets. Such modifications are also necessary if the algorithms are applied to an incoming data stream, i.e. they are to be applied in an on-line mode. The approach adopted is a standard segmentation approach. The data are divided into blocks of a fixed length (10 s is typically used here) and the algorithm is applied to each block individually. The blocks are overlapped by a fixed

percentage (75% being used here). This overlapping increases the number of time points, so reducing the interval between output points, creating plot which has a smoother appearance. However, increasing the overlap increases the computational load without enhancing the inherent temporal resolution: it is a form of interpolation. The output from each such block is either the spectrum (as in Figure 7, Figure 8 and Figure 9) for each block or the estimated modulation depths. In the former instance one obtains data as a function of both time and modulation frequency, in a manner akin to a spectrogram, whereas in the latter case one obtains a time-series showing modulation depth as a function of time.

File number	Filename	Duration (s)
1	DTI LFN study - Site 1	570.0
2	File 1 - extract, last 20s	19.9
3	DTI LFN study – Site 2 - external	82.6
4	DTI LFN study – Site 2 - internal	81.2
5	Van Den Berg sample (JSV article)	172.6
6	Web-sourced audio (extract)	23.1
7	"Other AM" example 1	60.0
8	"Other AM" example 2	60.0
9	"Other AM" example 3	60.0
10	"Other AM" example 4	60.0
11	"Other AM" example 5	60.0

Table 5: Durations and names of files used in the assessment of algorithm performance

For the purposes of this discussion all the data is resampled to a sample rate of 3 kHz prior to processing.

7.3.1 Performance on a single file

Initially consider the data from the recording of File 1, which contains 570 s of data at an original sample rate of 6.4 kHz. Spectrograms of this data, after down-sampling to a sampling rate of 3 kHz, are shown in Figure 14. The spectrograms are plotted on a dB scale but with an unspecified reference level.

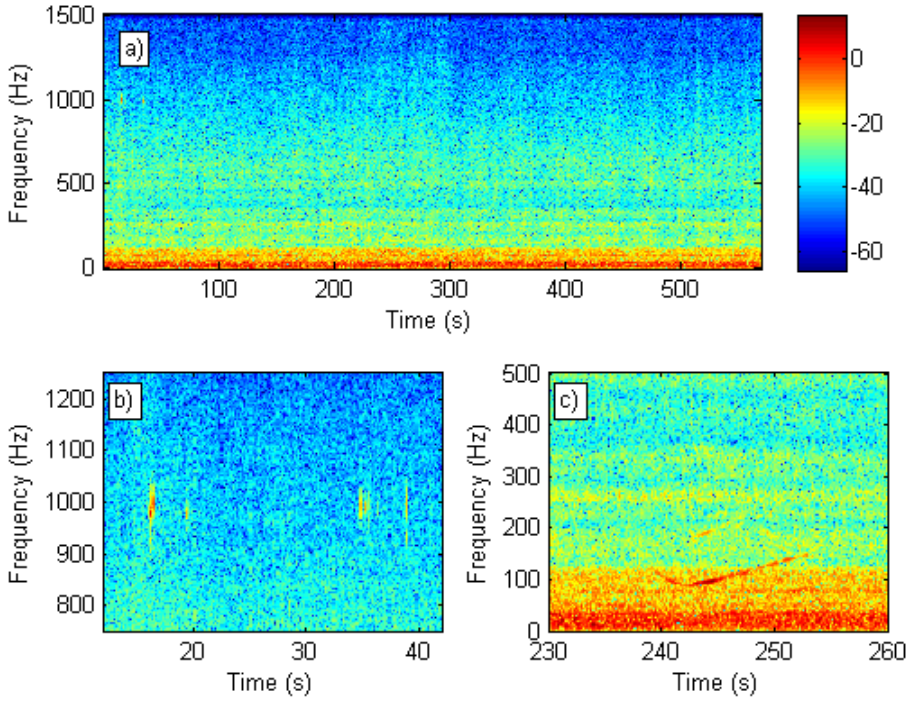

Figure 14: Spectrograms from File 1 a) Spectrogram of the full time-series b) Expanded view of spectrogram about region in which audible "squeaks" appear c) Expanded view of spectrogram about region where low-frequency frequency modulated signal occurs.

Figure 14 a) shows the spectrogram of all 570 s of data. On such a time-scale the general features of the time series are visible, specifically the rather consistent nature of signal is evident. It is clear that the frequency band up to 100 Hz contains a high level of noise. There are two clear events that happen in the data, for which the spectrogram is plotted in a smaller region around the events. Between 15 and 45 seconds there are series of audible "squeaks", the source of which is not clear, they produce short duration bursts of energy visible in the spectrograms close to 1 kHz; at least four such events are evident in Figure 14 b). The second event occurs roughly between 240 and 250 seconds, it takes the form of a narrow-band frequency modulated sound centred on 100 Hz. This is less readily audible, due to its low-frequency content.

This file is processed using three algorithms: the Fourier transform of the squared signal, the Fourier transform of the signal after band-pass filtering between 200 Hz and 1 kHz and the filter bank algorithm. The results are shown in Figure 15 in the form of images, each vertical line in the image represents the spectrum formed using the various algorithms for a data segment centred on the corresponding time. The values in the image are displayed on a decibel scale (unreferenced).

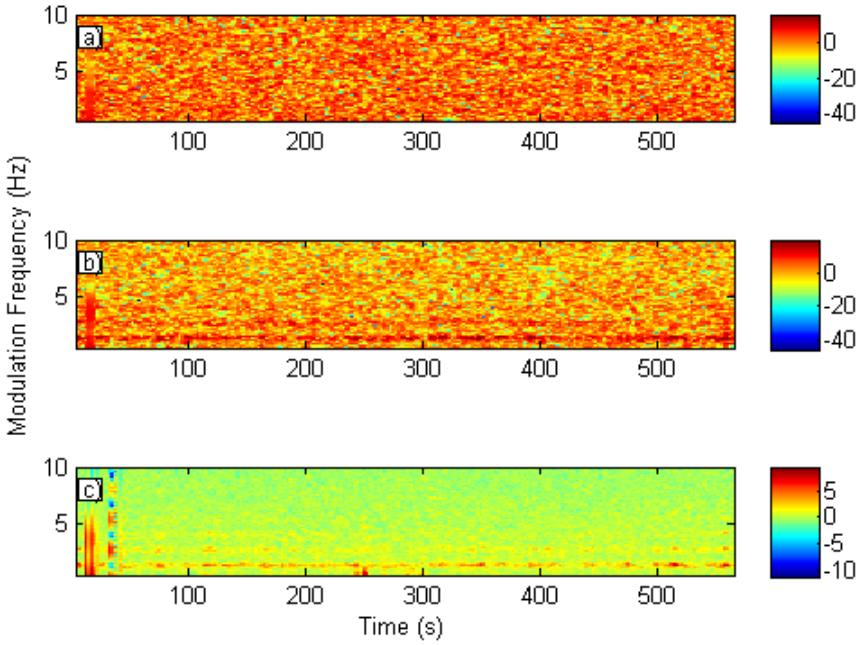

Figure 15: Results of applying three different modulation algorithms to File 1. The plots show the modulation spectra computed for 10 s segments, overlapped by 75%. a) The Fourier transform of the unfiltered energy, b) The Fourier transform of the energy after band-pass filtering between 200 Hz and 1 kHz, c) The results for the filter bank algorithm.

It is clear from Figure 15 a) that without filtering the Fourier transform of the square of the signal fails to detect a modulated component. This is in contrast to the results with filtering (in the band 200 Hz – 1 kHz), shown in Figure 15 b), where a band of increased values at a modulation frequency of slightly more than 1 Hz can be seen. Harmonics of this modulation can also be (faintly) seen in this plot. At approximately 20 s there is a vertical band of high values, this corresponds to the time at which squeaks occur (see Figure 14 b)). In Figure 15 c) the modulation is also clearly apparent at a frequency slightly in excess of 1 Hz. Also the effects of the squeaks between 10 and 50 s are clearly seen. Comparing Figure 15 b) and c) the modulation frequency is probably most obvious in Figure 15 c) suggesting some level of performance enhancement when using the filter bank method for this data set.

The same results are presented in an alternative form in Figure 16. These plots show the peak value of the spectrum computed for each 10 s data segment. Comparing Figure 16 a) and b) then it is clear that the values in Figure 16 b) are generally larger than those in Figure 16 a), suggesting that the modulation is more detectable in the filtered data than the unfiltered data. When considering these results with those in Figure 16 c) one has to takes some care since the absolute levels are not comparable.

There are peaks in Figure 16 c), for example the peak at approximately 520 s, which would appear to correspond to period of increased levels of modulation when the spectrogram is examined.

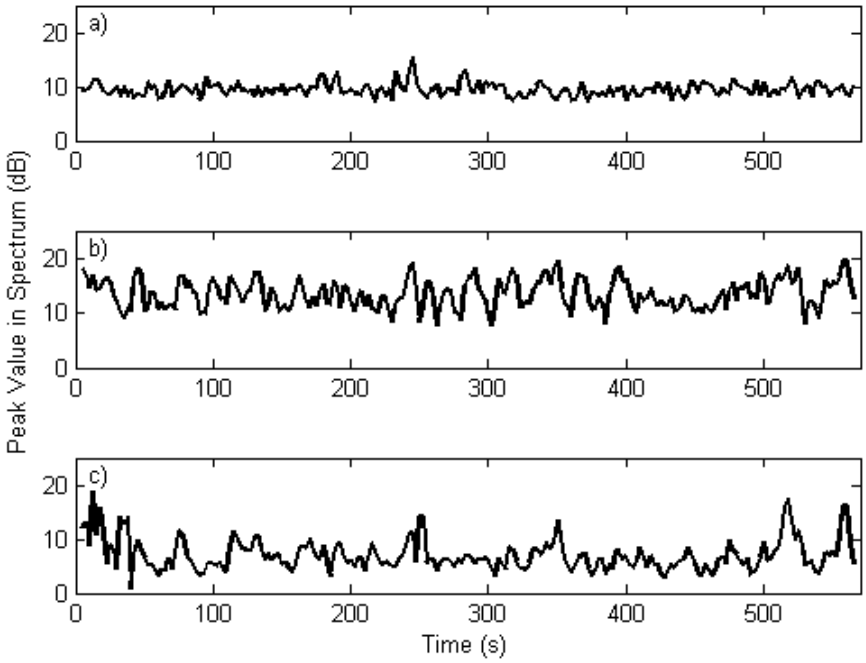

Figure 16: Results of applying three different modulation algorithms to File 1. The plots show the modulation spectra computed for 10 s segments, overlapped by 75%. a) The Fourier transform of the unfiltered energy, b) The Fourier transform of the energy after band-pass filtering between 200 Hz and 1 kHz, c) The results for the filter bank algorithm.

Figure 17 shows another aspect of these results. In particular it depicts the estimated modulation rate for each 10 s block. The upper frame, Figure 17 a), shows the results obtained using the FT of the unfiltered energy. The estimated modulation rates fluctuate considerably. This is a consequence of the fact that the maximum values in the corresponding spectra, as shown in Figure 15 a), do not occur at consistent modulation frequencies. This is in contrast to the data in Figure 17 b), where many of the points are close to 1 Hz (a realistic blade pass frequency for the turbine). The peaks away from this frequency correspond to point in time when the data in the corresponding 10 s window generates a peak due to noise which is greater than the peak due to the modulation rate. Finally, Figure 17 c) shows a greater level of consistency, most of the points are close to 1 Hz. This suggests that the filter bank method has achieved a more consistent metric of modulation in this data set.

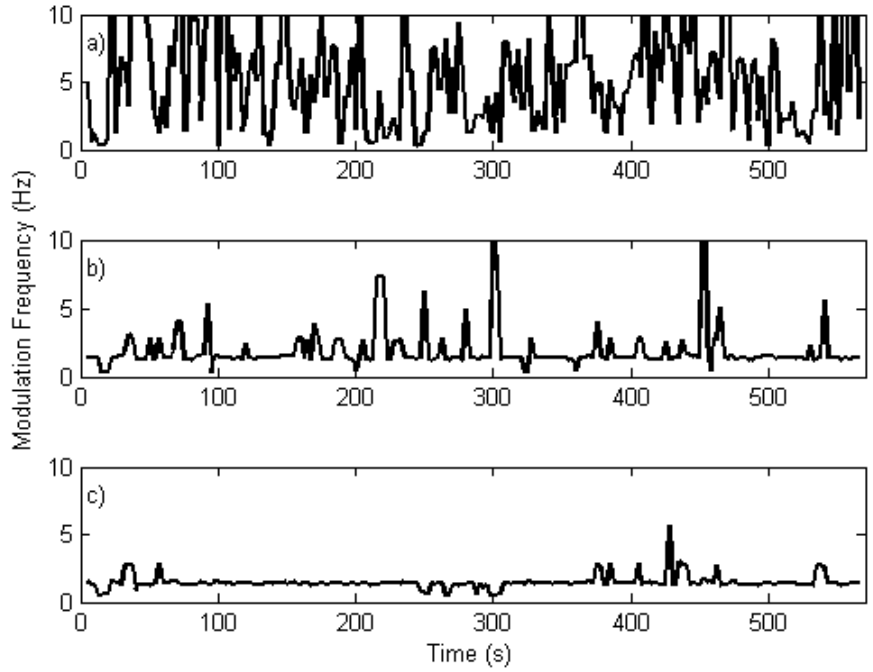

Figure 17: Estimated modulation frequencies for File 1. The plots show the modulation spectra computed for 10 s segments, overlapped by 75%. a) The Fourier transform of the unfiltered energy, b) The Fourier transform of the energy after band-pass filtering between 200 Hz and 1 kHz, c) The results for the filter bank algorithm.

7.3.2 Performance on the set of files

In order to examine the performance of the algorithms on the remaining 10 files, listed in Table 5, a full analysis, similar to that conducted in the previous subsection is not practical. In this section only the results from the filtered energy algorithm and the filter bank method are considered. The analysis depicting the estimated modulation depths and modulation rates as a function of time, as in Figure 16 and Figure 17 are shown.

The results from these analyses are depicted in the following in a sequence of figures. These figures are drawn in a consistent format, each showing the analysis of five files, so that two figures represent the analysis on the full dataset. The modulation depths and estimated modulation rates are shown side-by-side to allow for visual identification of regions where the detection is reliable (i.e. when the modulation rate is in the region of 1 Hz). The fact that some files are very short in duration means that the segment based approach is not very suitable for them as only a small number of segments are produced. However this form of analysis is still applied in order to maintain consistency.

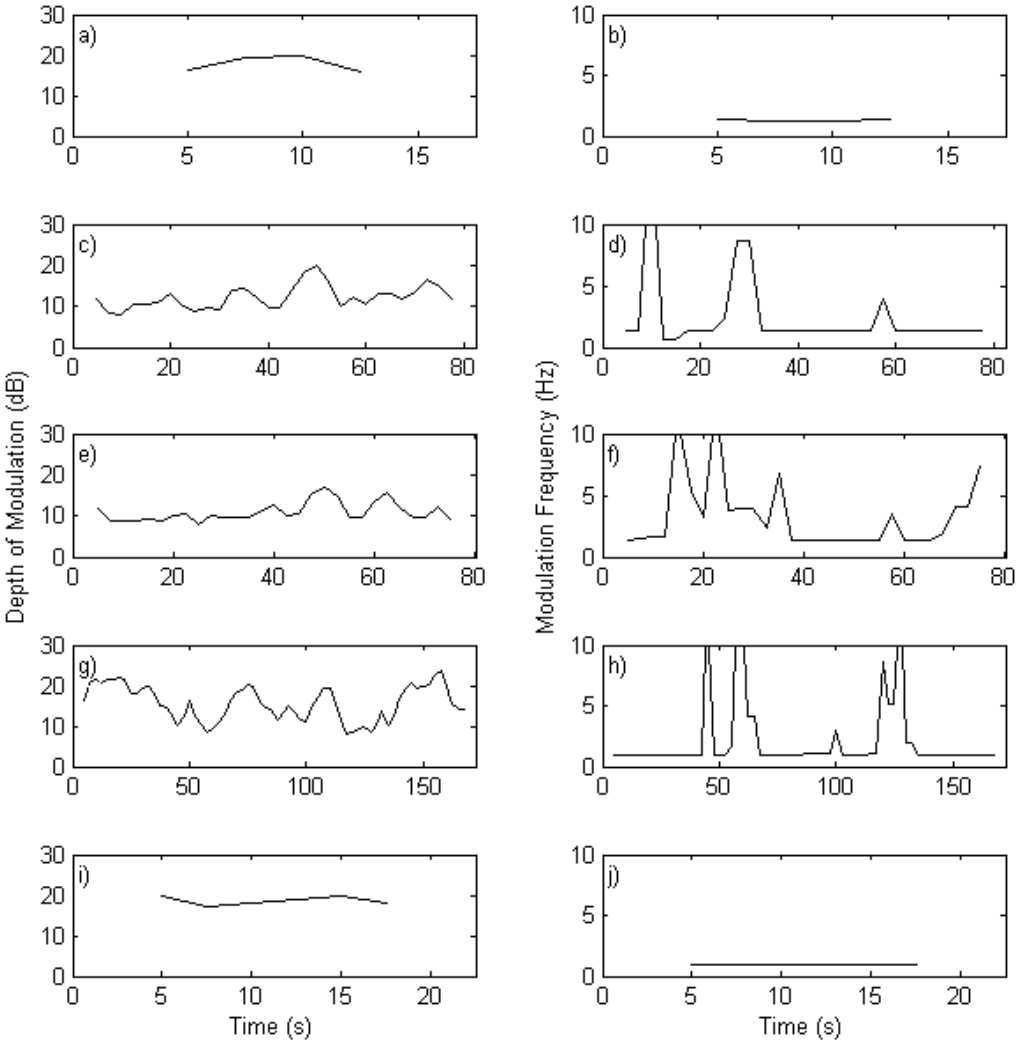

Figure 18: Results using the filtered energy algorithm (filter pass band 200 Hz – 1 kHz). Left hand column shows the depth of modulation, right hand column the estimated modulation frequency. File 2: a) and b), File 3: c) and d), File 4: e) and f), File 5: g) and h), File 6: i) and j).

The results for Files 2 and 6 shown in Figure 18 suggest good performance, in both case the modulation depth is high and the corresponding estimated modulation frequency is consistent and realistic, i.e. the modulation frequency is in the region of 1 Hz. Although both these recording are short, approximately 20 s. The results for Files 3-5 illustrate a rather less impressive level of performance, other methods also fail to strongly detect modulations in these files, see Figure 19, suggesting that this data is weakly modulated. The algorithm estimates realistic modulations frequencies for periods of time, but is prone to generating rather wild values when the depth of modulation is low. For example the results for File 5 (Figure 18 g) and h)), at around 50 s and 120 s there are significant drops in depth of modulation resulting in erratic values for the modulation frequency, excluding these unrealistic modulation rates would result in more consistent values.

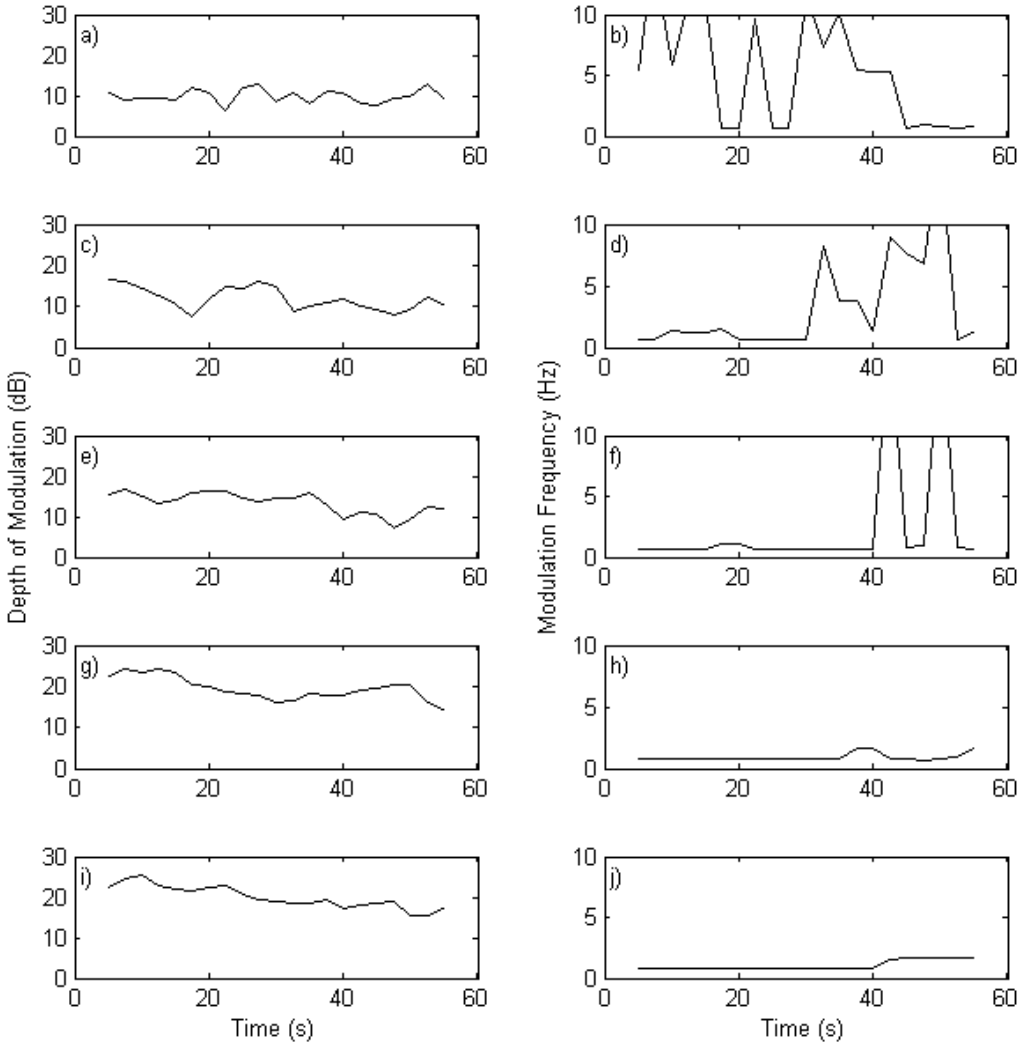

Figure 19: Results using the filtered energy algorithm (filter pass band 200 Hz – 1 kHz). Left hand column shows the depth of modulation, right hand column the estimated modulation frequency. File 7: a) and b), File 8: c) and d), File 9: e) and f), File 10: g) and h), File 11: i) and j).

The results in Figure 19 show results for recordings which were supplied as examples of "Other AM", of varying levels, measured in the far-field of a wind farm. In general the estimated depth of modulation for File 7 is low and consequently the estimated modulation rates are largely unrepresentative. In Files 8-11 there is all show a similar pattern of behaviour; at the start of the recording modulation depth is large and the estimated modulation frequency close to one, with the depth of modulation generally decreasing with time and the modulation frequency becoming less reliable. Files 10 and 11 have clear modulations that can be easily detected in the filtered waveform and in the spectrogram.

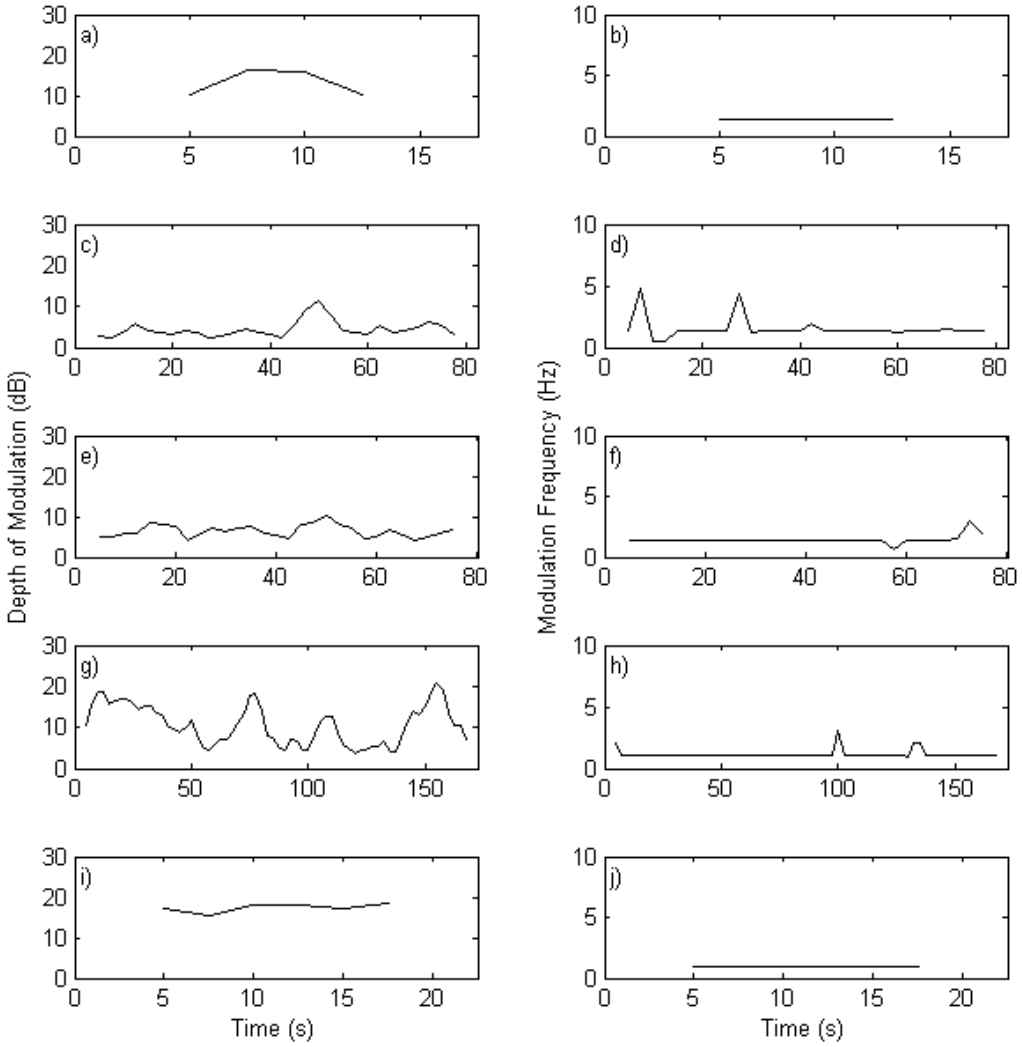

Figure 20: Results using the filter bank method. Left hand column shows the depth of modulation, right hand column the estimated modulation frequency. File 2: a) and b), File 3: c) and d), File 4: e) and f), File 5: g) and h), File 6: i) and j)

The results in Figure 20 should be compared to those in Figure 18 for the filtered energy algorithm. In general the estimated modulation depths show the same trends, albeit the that the two algorithms are not expected to yield values for the depth of modulation that are the same, i.e. comparisons of actual numerical values should not be considered. The performance on File 2 and 6 qualitatively matches that shown in Figure 18 a), b) and i), j). For Files 3-5 this algorithm seems to yield a more consistent estimate of modulation frequency, suggesting a more appropriate measure of depth of modulation. For example comparing the performance on File 4 (frames e) and f)) one sees that the filtered energy algorithm fails to detect a modulation prior to 40 s (with the possible exception of the

first segment), whereas the filter bank method tracks this modulation successfully throughout that time.

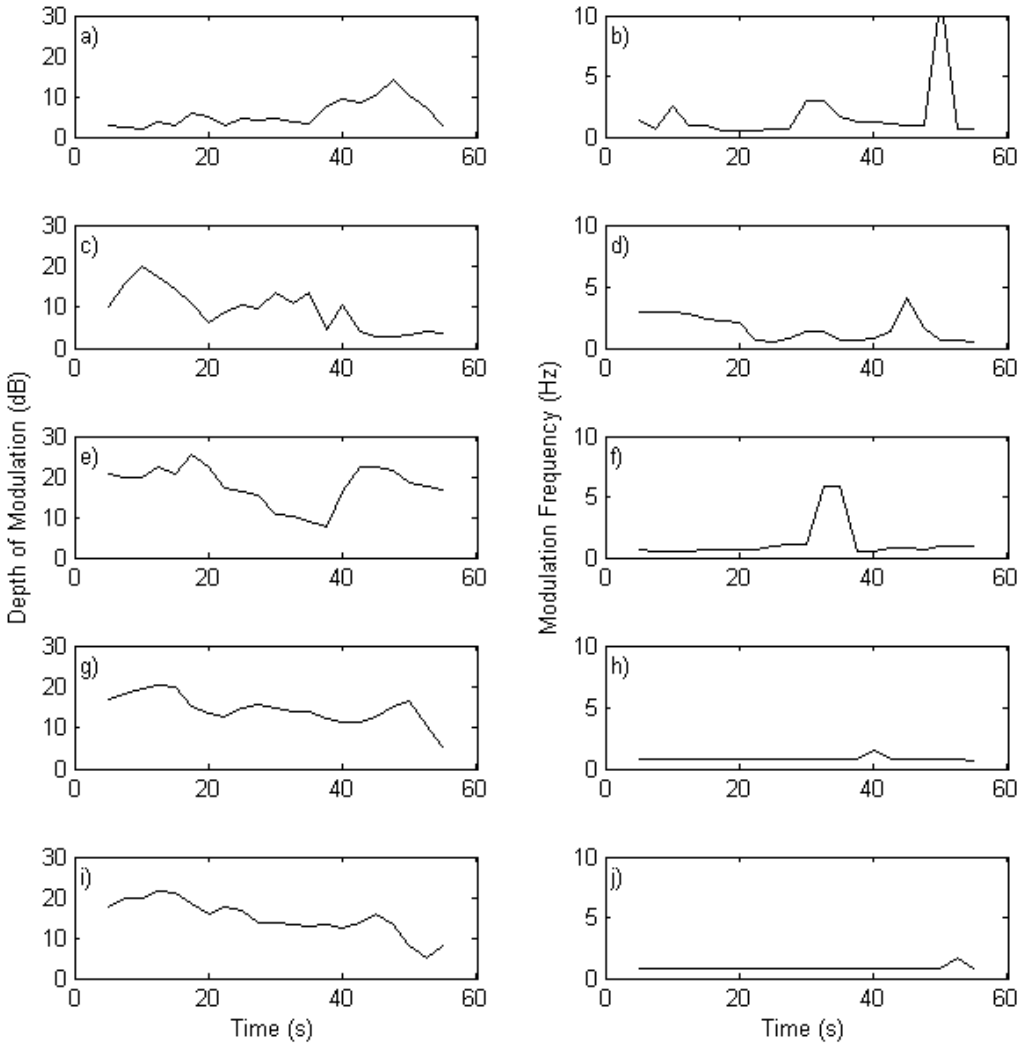

Figure 21: Results using filter bank method. Left hand column shows the depth of modulation, right hand column the estimated modulation frequency. File 7: a) and b), File 8: c) and d), File 9: e) and f), File 10: g) and h), File 11: i) and j).

Figure 21 should be compared to those in Figure 19. These results are interesting in that they would appear to suggest that the filter bank method, for these data, is performing somewhat less well that the filtered energy method. In general the modulation frequencies are rather larger than 1 Hz (not always unrealistically so) but the trends suggested involve turbine dynamics which probably are unrealistic (halving or doubling of turbine rotation rates in tens of seconds). Examination of the data in File 9 reveals that the start of the recording is contaminated by bird noise (carrion crows and pheasant calls).

These calls have energy which resides primarily above 1 kHz, so the filtered algorithm removes them, whereas the filter bank method is affected by their presence. More problematic are the pheasant calls at the start of File 8. Audibly these are completely different to the swish of a turbine blade, but they do appear as a sequence of (almost) periodic bursts of energy. Figure 22 shows a spectrogram of this section of data, at the original sample rate of 6 kHz. The periodic clucking is seen as vertical lines between 1 and 2.5 kHz. There are 16 clucks, approximately regularly spaced, within a 5 second interval, corresponding to a clucking rate of approximately 3 Hz. Inspection of Figure 21 d) reveals that the estimated modulation rate is also approximately 3 Hz, indicating that it is indeed these sounds which have been detected. From the spectrogram one can also see that there is some energy in these signals in the 1 – 1.5 kHz band, which is included in filter bank method, but is removed from the filtered energy approach, this explains why the latter is unaffected by these sounds.

There is a similar burst of pheasant clucking which occurs at 30 s in File 9. This event is quieter and harder to analyse, but it is at a noticeably high rate that that shown in Figure 22 and consequently is the likely cause of the peak in the estimated modulation frequency seen in Figure 21 f).

In this instance the problem can be readily removed from this algorithm by only using frequency bands up to 1 kHz.

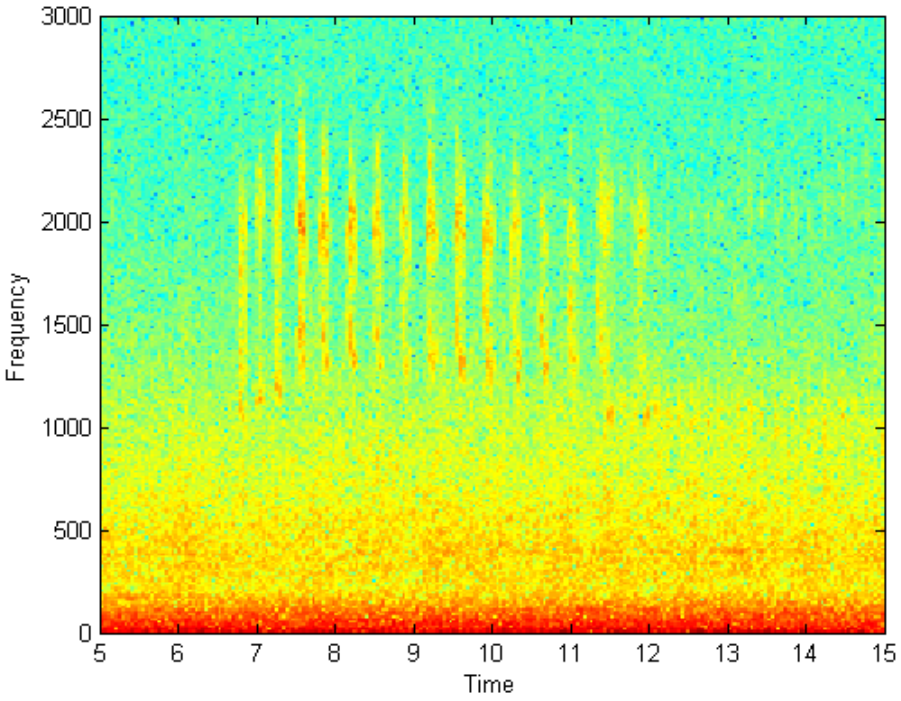

Figure 22: Spectrogram of data segment in File 8. Data contains the clucking of a common pheasant (*Phasianus colchicus*).

8. Conclusions

There are a very large number of options in terms of how the degree to which is a signal is modulated might be measured. These options cover the variety of ideas about what is to be measured, how the frequency dependence of the modulation is taken into account and what methods are used to extract parameters from the data. All of this variety exists without considering the additional problems associated with making these measures subjectively meaningful.

The majority of this report considers the algorithms which can be used to estimate the parameters of the modulation. There are three forms of method that have been considered, all of which claim a degree of optimality and have been derived from basic principles.

The methods based on short term energy (akin to L_{eq} based methods) are optimal if the noise and modulation process are both Gaussian white processes and the modulation is sinusoidal. These methods are simple to implement and can be adapted using various ways. Their performance is surprisingly good over the stimulus data set.

These short term energy methods can be adapted to deal with non-white processes by applying a filter bank and treating the output of each filter bank as a separate white noise signal. This can be efficiently realised using spectrograms. The performance of these methods on the stimulus set used for research on subjective response is similar to the short-term energy methods. The spectra produced by this method do look smoother, but the very smallest modulations may fail to be detected.

The results from the field recordings suggest that in general the method based on the FT of energy without filtering does not perform particularly well. In these examples the large amount of un-modulated low-frequency noise tended to disrupt the algorithm's performance. Applying appropriate filtering prior to computation of the energy can focus the processing on the frequency band where the modulation is significant. This does require one to identify the suitable band *a priori.*

The filter bank method seems to offer a small advantage over the FT of filtered energy. It tends to produce results which are somewhat more consistent in time, suggesting greater reliability. In these tests it is affected more by the presence of external sounds, in this case bird calls, which happen to fall within the frequency band that it employs, but is rejected from the filtered energy approach.

Finally the method of Kirsteins *et al*. [31] has been considered. This approach is optimal for a broader class of processes, not just amplitude modulated but for periodically correlated processes. This approach seems to require a high SNR in order to function adequately and in the stimulus data set considered it was unable to detect the modulation in most instances. Further the computational load for this method is significantly greater than that needed for the other methods.

Appendix: Construction of Test Stimuli

To facilitate the subjective listening tests a set of synthetic stimuli were desirable. The content of such stimuli are completely known, with a fully variable set of input parameters, and one can be confident that the subjects are not responding to unknown factors in the data. However, such stimuli need to be sufficiently close to real stimuli as to elicit the same response.

To this end a model of modulated wind farm noise was developed. This model was constructed by modulating Gaussian noise with a defined spectral shape. The parameters of the model were then fitted to recordings of wind turbine AM in the far-field, in an attempt to mimic that dataset.

The basic model uses shaped Gaussian noise. This noise is filtered using filters whose frequency response functions have the form:

$$\begin{aligned} H(f) &= e^{-(f-f_c)^2/B_l^2} \quad f < f_c \\ &= e^{-(f-f_c)^2/B_u^2} \quad f > f_c \end{aligned} \tag{48}$$

where f_c is the centre frequency of the modulated signal. The spectrum is asymmetric about this centre frequency, having a bandwidth of B_l below the centre frequency and a bandwidth of B_u above.

This filtered noise is then modulated with a time waveform which has the form:

$$m(t) = \sum_{n=0}^{N_p} p(t - nT_p) \tag{49}$$

where T_p is the interval between modulation pulses, N_p is the number of pulses in the waveform and the $p(t)$ shape of each pulse. The pulse waveform $p(t)$ is selected to be an asymmetric Gaussian shape, similar to that used in (48), specifically

$$\begin{aligned} p(t) &= e^{-t^2/D_l^2} \quad t < 0 \\ &= e^{-t^2/D_u^2} \quad t > 0 \end{aligned} \tag{50}$$

where D_l and D_u define the pulse durations before and after the peak respectively. The model can be summarised as:

$$x(t) = \alpha m(t) \times (n(t) * h(t)) + b(t) \tag{51}$$

in which α is a measure of the depth of modulation, $h(t)$ is the impulse response of the filter whose frequency response is defined by (48), * denotes convolution , $n(t)$ is a Gaussian white noise process and $b(t)$ represents the ambient (or masking) noise, which can be synthetic or a sample of real world data.

This model is parameterised via seven parameters: the centre frequency (f_c) and the two bandwidths (B_l and B_u), the inter-pulse interval T_p, the durations D_l and D_u and a parameter to control the depth of modulation α (this parameter represents the ratio, expressed in dB, of un-weighted peak root-mean-square (r.m.s.) level of the modulation and the r.m.s. level of the masking noise).

References

[1] L. W. Couch, *Digital and analog communication systems*: Pearson/Prentice Hall, 2007.

[2] A. R. Hambley, *An introduction to communication systems*: Computer Science Press, 1990.

[3] B. P. Lathi and Z. Ding, *Modern Digital and Analog Communication Systems*: Oxford University Press, 2009.

[4] I. Otung, *Communication engineering principles*: Palgrave, 2001.

[5] R. E. Ziemer and W. H. Tranter, *Principles of communications: systems, modulation, and noise*: Wiley, 2002.

[6] D. Ross, *Mechanics of underwater noise*: Peninsula Publishing, 1987.

[7] R. O. Nielsen, *Sonar signal processing*: Artech House, 1991.

[8] J. R. Stack, *et al.*, "An amplitude Modulation detector for fault diagnosis in rolling element bearings," *Industrial Electronics, IEEE Transactions on,* vol. 51, pp. 1097-1102, 2004.

[9] R. B. Randall, *et al.*, "The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals," *Mechanical Systems and Signal Processing,* vol. 15, pp. 945-962, 2001.

[10] J. G. Lourens and J. A. du Preez, "Passive sonar ML estimator for ship propeller speed," *Oceanic Engineering, IEEE Journal of,* vol. 23, pp. 448-453, 1998.

[11] J. Bass, "Investigation of objective AM assessment methodologies," RES 25 Feb 2010.

[12] J. N. McCabe, "Detection and quantification of amplitude modulation in wind turbine noise," presented at the Fourth Internation Meeting on Wind Turbine Noise, Rome, 2011.

[13] D. McLaughlin, "Measurement of amplitude modulation frequency spectrum," presented at the Fourth International Meeting on Wind Turbine Noise, Rome, 2011.

[14] G. Lundmark, "Measurement of swish noise. A new method," presented at the Fourth International Meeting on Wind Turbine Noise, Rome, 2011.

[15] S. Lee, *et al.*, "An estimation method of the amplitude modulation in wind turbine noise for community response assessment," presented at the Third International Meeting on Wind Turbine Noise, Aalborg, 2009.

[16] J. Vos and M. M. J. Houben, "Analyse van opnamen van windturbinegeluid," TNO TNO-DV 2010 C014, February 2010 2010.

[17] J. Bass, "Investigation of the 'Den Brook' Amplitude Modulation Methodology for Wind Turbine Noise," *Institute of Acoustics Bulletin,* vol. 36, p. 7, 2011.

[18] H. L. van Trees, *Detection, estimation, and modulation theory*: Wiley-Interscience.

[19] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation theory*: Prentice-Hall PTR, 1993.

[20] R. A. Fisher, "On an absolute criterion for fitting frequency curves," *Statistical science,* vol. 12, pp. 39-41, 1912.

[21] J. V. Candy, *Bayesian signal processing: classical, modern, and particle filtering methods*: Wiley, 2009.

[22] A. Papoulis, *Signal analysis*: McGraw-Hill, 1977.

[23] J. Dugundji, "Envelopes and pre-envelopes of real waveforms," *Information Theory, IRE Transactions on,* vol. 4, pp. 53-57, 1958.

[24] J. K. Hammond and P. R. White, "The Analysis of Non-stationary signals using time-frequency methods," *Journal of Sound and Vibration,* vol. 190, pp. 419-447, 1996.

[25] B. Boashash, "Estimating and Interpreting the Instantaneous Frequency of A Signal .1. Fundamentals," *Proceedings of the IEEE,* vol. 80, pp. 520-538, 04 1992.

[26] M. Hinich, "Detecting a hidden periodic signal when its period is unknown," *Acoustics, Speech and Signal Processing, IEEE Transactions on,* vol. 30, pp. 747-750, 1982.

[27] T. S. Leung and P. R. White, "Robust estimation of oceanic background spectrum," in *Mathematics in signal processing IV*, J. G. McWhirty and I. K. Proudler, Eds., ed Oxford: Clarendon Press, 1998, pp. 369-382.

[28] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *Speech and Audio Processing, IEEE Transactions on,* vol. 9, pp. 504-512, 2001.

[29] P. Clark, *et al.*, "Multiband analysis for colored amplitude-modulated ship noise," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 2010, pp. 3970-3973.
[30] G. Strang and T. Nguyen, *Wavelets and filter banks*: Wellesley-Cambridge Press, 1996.
[31] I. Kirsteins, *et al.*, "Maximum lieklihood estimation of propeller noise modulation characteristics," presented at the Underwater acoustic measurements: Technologies and Results, Kos, Greece, 2011.
[32] L. Cohen, *Time-frequency analysis*: Prentice Hall PTR, 1995.
[33] S. von Hunerbein, *et al.*, "Work Package B(2): Dose response relation for the annoyance of AM noise," University of Salford, Manchester2012.
[34] M. Cand and A. Bullmore, "Work Package C (WPC): Collation and analysis of existing acoustic recordings," 2012.