University of Salford MANCHESTER

Acoustics Research Centre

AM Dose-Response Relationship

s.vonhunerbein@salford.ac.uk

http://www.acoustics.salford.ac.uk/profiles/vonhunerbein/

# Wind Turbine Amplitude Modulation: Research to Improve Understanding as to its Cause & Effect

## Work Package B(2): Development of an AM Dose-Response Relationship

Prepared for RenewableUK by:

Sabine von Hünerbein, Andrew King, Benjamin Piper, Matthew Cand

# Contents

# Table of Figures

# 1. Preface

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 AM Dose-Response Relationship’.

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.

# 2. Executive Summary

Amplitude modulated (AM) sound from wind turbines is difficult to characterise and there is insufficient knowledge about listener response to the characteristic physical properties (metric) of AM sound.

The first objective of the present Work Package was to test whether the AM metrics developed in Work Package B1 (WPB1) would provide a meaningful measure of AM 'value' that correlates with subjective annoyance ratings. The second objective was to quantitatively investigate the relationship between the AM value and a measure of average annoyance in the form of a dose-response relation.

Key to obtaining representative results was the design of stimuli which were representative of the spectrum and character of actual noise experienced by wind farm neighbours exposed to AM. Extensive work was done to obtain this, based on input from Work Package C, test signals were synthesised for a characteristic range of wind turbine sounds, with a wide range of input parameters. This model is described in an annex to WPB1. It was considered by the project team that the artificial stimuli obtained were representative of AM experienced in the field.

In a first phase, sensitivity tests were undertaken to find the AM parameters that listener response was most sensitive to. 80 test sounds of at least 20 second duration were presented via calibrated headphones in a quiet room and 11 volunteers were asked to score the annoyance on a numeric 11 point scale. The outdoor sounds included test wind turbine noise at typical levels, with varying AM characteristics and some natural background noise.

Then, a final set of tests was undertaken in a quiet listening room with a sound reproduction that mimicked the outdoor directivity of one wind turbine in the distance. 32 test sounds were generated for and presented to 20 participants. Two validation tests containing another 158 and 32 test sounds respectively were conducted to clarify results from the sensitivity tests in the better-controlled listening room. Participants rated annoyance directly as before, and also adjusted an un-modulated test sound in level to be equally annoying to the modulated test sound.

The sensitivity tests showed in accordance with previous literature that annoyance crucially depended on the A-weighted level of the test sound, as measured in $L_{Aeq}$, and to a lesser extent on modulation depth which is a measure of the modulation strength. Modulation depth was shown to be also best expressed in terms of A-weighting to give consistent results. The use of $L_{A90}$ as an alternative to $L_{Aeq}$ produced similar results at the low and medium modulation depths most often observed from wind turbines.

In the final test there were therefore 3 sets of test sounds with the constant $L_{Aeq}$ of 30, 35, and 40 dB(A) for which the estimated modulation depth was systematically varied from 0 to 12 dB(A) in increasing steps. After taking into account the effect of $L_{Aeq}$, which always dominated the annoyance rating, the modulation depth seemed to increase the annoyance rating slightly but consistently. However, the effect was not statistically significant because there was a large spread of ratings. This suggests that given a large enough group of participants it can possibly be shown that annoyance increases consistently (monotonically) with modulation depth. In contrast, the $L_{Aeq}$ level of the adjusted un-modulated wind turbine noise remained broadly constant as the modulation depth increased above about 3 dB(A). This answered the question of how much louder would an equivalent un-modulated sound have to be to be equally annoying to a modulated sound. The adjustments were on average 1.7 dB(A) for a 40 dB(A) test sound and 3.5 dB(A) at 30 dB(A). Validation tests at two additional levels of 45 dB(A) and 25 dB(A) confirmed this trend. When levels were measured as $L_{A90}$, results suggest that annoyance ratings were similar for modulation depths of up to 6 dB(A) and generally increased with both modulation depth and $L_{A90}$. Because results for sets of stimuli with constant $L_{A90}$ and changing modulation depth are not available simple average adjustments cannot be identified and further work would be necessary. A clear onset of annoyance at a particular modulation depth could not be found for either of the two rating methods.

In a validation test with a subgroup of 11 participants, the spectral characteristics of the test sound were changed to represent Mid-Frequency AM, often described as swish, as opposed to Reduced Frequency AM which is sometimes described as a “swoosh” or “whoomp”. Also garden noise was added at a low level to change the character of the sound for both types of AM sounds. For all four groups the results for both absolute annoyance ratings and un-modulated level adjustments appeared very similar. This suggests that the relative effect on annoyance is small as long as the garden noise does not reduce the audibility of the modulated sound.

In a last step the annoyance ratings were compared for 6 different metrics, four of them based on different physical definitions of modulation depth, using input from WPB1, and 2 using the perceptive measure of fluctuation strength. The comparison showed that the main effect of the physical metric is to change the range of modulation depths. The same stimuli would have a range of 0 – 12 dB(A) modulation depth in one metric but 4 – 32 dB in another metric. Fluctuation strength results showed a further step towards a metric that correlates with listener response but not even a perception based metric can ever account for contextual and attitudinal aspects of annoyance rating.

# 3. Definitions

Different types of noise are present where wind turbine noise is audible and the sound at the receiver location can be thought of as a combination of some or all of these types of noise.

**Wind Turbine Noise** (WTN) includes a steady component as well as, in some circumstances, a periodically fluctuating or **Amplitude Modulated (AM)** component or character. The report for Work Package C (WPC) has highlighted the difference between different instances of this modulation. Firstly, **Normal Amplitude Modulation** (NAM) can be explained by well-understood mechanisms, as noted in the Preface. The second type named **Other Amplitude Modulation** (OAM) appears to have other characteristics, such as increased modulation level and/or frequency content, and does not seem consistent with the available theories, although different potential mechanisms have been described in Work Package A2 (WPA2). When the abbreviation AM is used in this report, this covers both NAM and OAM.

The differing frequency characteristics sometimes identified for AM signals has also led to the definition within the current report of **Medium-Frequency Amplitude Modulation** (MFAM) and **Reduced-Frequency Amplitude Modulation** (RFAM): see Section 6.1.

In general wind turbine noise can also contain tones, which might sometimes be audible above the steady WTN, but this study will not focus or include any tones, as subjective response to tones has already been studied extensively.

The audibility of wind turbine noise at the listener position can be reduced or completely masked by environmental or ambient **Background Noise** (BN). Residents near wind turbines experience BN as any noise that is not originating from the wind turbines. A specific type of BN is **Vegetation noise** (VN). It is also commonly called **Garden Noise** (GN) which in the context of this study is the sum of vegetation and other outdoors background noise. But the term **Masking Noise** (MN) is used in the present study as noise that will mask the AM component of stimuli experienced, and therefore MN will represent either steady WTN on its own, or the combination of steady WTN and all or some of vegetation noise and background noise.

# 4. Context and Aims

In recent years the debate about the annoyance from wind turbine noise in general, and RFAM noise in particular, has increased as this has been reported by some wind farm neighbours to be annoying [WPC, 1]. It is plausible that AM noise may be more disturbing than steady noise, and there is an ongoing discussion on whether current guidelines for the assessment of WTN such as ETSU-R-97 [2] are sufficiently taking AM and in particular RFAM in noise into account. Therefore wind farm developers, planners and policy makers are interested to find out how and why AM in general and in particular RFAM occur as well as how this type of noise is generally perceived and how listeners respond to it. Unfortunately not much is known about the occurrence of RFAM. The other parts of the current Renewable UK were aimed at improving the understanding, measurement and prediction of AM. Details can be found in the project reports, [WPA2, WPB1, and WPC]. They informed this report which focuses on the response of listeners to AM.

The work aimed to develop a scientifically based procedure for the rating of AM effects. Previous work suggests that the following general behaviour might be observed:

- **threshold of onset of annoyance –** when the fluctuation becomes strong enough sound becomes more annoying than steady sound of the same $L_{Aeq}$
- **relation between AM characteristic parameters and mean annoyance score at AM values above the threshold.** Previous research suggests that annoyance might be observed to systematically increase with the increase of certain characteristic parameters

If one or both of these hypotheses can be validated then they can aid the development of guidance for any AM based planning condition. For example a relation between mean annoyance scores and fluctuation strength could potentially be used to define a ‘penalty’ procedure for AM, by matching the mean annoyance scores of AM noise with continuous noise. This would provide the basis for rating the AM characteristic of wind turbine noise if this is considered to be necessary.

# 5. Overview of annoyance related literature

As a background to the listening test design this chapter aims to give an overview over common investigation techniques for sound related human response studies. In the second part of the literature study a short overview over the current knowledge of the annoyance from AM sounds is given.

## *5.1 General methods used for sound related human response studies*

In sound related studies it is useful to recognise the systematic differences between physical, perceptual and affective measurements as laid out in the Filter model [3]. It describes how perceptual measurements are different from physical measurements because they are "filtered" by the senses like the ear and the capability of the brain to process sounds. Affective measurements are different from perceptual measurements because they are influenced by non-acoustic factors such as mood, context, emotion, background and expectation of the listener. Typical physical properties of AM sounds are equivalent A-weighted sound pressure level $L_{Aeq,}$ spectral content, modulation depth and modulation frequency. Perceptual measures in this context are loudness and pulsation whereas the affective measure that this report focuses on is annoyance.

Recent literature reviews [4, 5] have discussed the methods for assessing noise annoyance in the general context of environmental noise and for wind turbine noise in particular. The following paragraphs contain excerpts from [5].

Common methods to study the perception of sound are scaling magnitude estimations and paired-comparisons. In the first a participant assigns a numerical value to a test sound or "stimulus". This value quantifies the property (loudness, annoyance, etc). A stimulus can be a naturally occurring sound or a sound that is synthesised in a laboratory. Another method is paired comparison, whereby two stimuli containing examples of the property are presented and a two-way rating scale is adjusted to indicate the relative rating of the two stimuli. Alternatively, one of the two stimuli can be actively adjusted by the participant until the two are equally representative for the property being studied. While the method of magnitude estimations can be used in survey studies and laboratory experiments the paired comparison method is naturally restricted to laboratory environments where stimuli can be presented in a controlled way. Paired comparisons are more difficult to do with stimuli of long duration. However, a number of studies such as [6] and [7] have concluded that stimuli length of as short as 30 sec give comparable results to long stimuli. Also the listening duration for adjustment procedures is harder to control as the listeners need to be given the choice to re-listen to both stimuli.

### 5.1.1 Survey studies

At present, the majority of work focuses on measuring the environmental noise levels, either at the source and using propagation algorithms or at nearby residences, and acquiring annoyance ratings via surveys. These two measures are combined to create dose-response relations for noise levels (or any other characteristic) and community annoyance classified by source. [8] provides a good synthesis of 11 such examples for various forms of transport noise.

Survey studies have the advantage that they measure in the listener's natural environment. Therefore context and attitude can be taken into account. The disadvantages are that these studies are retrospective studies on an emission that already exists. Apart from the well-known problems with this method it is not applicable in a situation where there the occurrence of AM sound is doubtful or infrequent. This limits the available data base where a large number of participants would be necessary because of the source variability and other factors.

### 5.1.2 Laboratory experiments

Another way of studying environmental noise annoyance is to present either recordings of noise or similar, synthesised sounds to participants in the controlled environment of a laboratory such as an anechoic chamber or listening room/booth. Many of the physical properties of sound and environment and some personal variables can be controlled, thereby allowing accurate estimates of how acoustical parameters affect noise annoyance. To an extent, researchers can choose to study noise annoyance with respect to its non-acoustical factors, although never quite as realistically as in field studies, by including non-acoustic sensory stimuli typically associated with the noise source.

Additionally, an experimental design can either allow or restrict the influence of context by, for example, asking participants to imagine a particular scenario during exposure. For attitude, the participant can be explicitly informed of the source thereby allowing their expectations and previous experiences of the source to influence their ratings of annoyance. Alternatively, researchers can study noise annoyance purely from an acoustical perspective by limiting other sensory stimuli or keeping them constant, removing contextual cues and keeping participants naïve to the source.

However well designed, a laboratory experiment will never give the same absolute ratings as a survey study because the laboratory environment is incompatible with the natural environment where the noise occurs and the listeners are out of their usual context. Therefore relative annoyance measures will give a better impression when comparative results are useful.

Because this current project studied the affective response to AM sound as compared to steady sound, the laboratory environment was suitable for controlled comparisons. Firstly, because the stimuli can be well controlled in a way which has not been studied before. This is important because WPC has highlighted the difficulties in measuring this type of WTN even in situations where its occurrence was relatively prominent, and the significant variability encountered. Secondly, it was chosen to expose all listeners to the same noise in the same environment thereby making the experiment reproducible and thirdly to control the context as described above.

## *5.2 Previous work on the response to modulated sounds*

The known characteristics of the perception of modulated sounds can be described in terms of a basic psychoacoustics model proposed by Fastl and Zwicker [9]. In its simplest form the model can be applied to the amplitude modulation of pure tones. This is not representative of wind turbine noise as that is broadband in nature.  A more complex form of the model which is extended to broadband AM sounds in the presence of masking noise, is used because it may be applicable when considering typical wind turbine noise modulation. The corresponding model uses the perception parameter loudness as well as the measureable physical parameters modulation depth, modulation frequency and frequency deviation as predictors for a perception parameter called fluctuation strength. This model was developed based on a series of observations and experiments. It is conceivable that this subjective parameter could relate to annoyance although [9] does not provide a direct relation.

Fluctuation strength is reported to experience a maximum for modulation frequencies around 4 Hz. A basic unit called a vacil is therefore defined by the authors relative to the subjective perception of a 60 dB, 1 kHz tone which is 100% amplitude modulated at 4 Hz. The model of broadband sinusoidally modulated sound is given by the following equation:

$$F_{BBN} = \frac{5.8(1.25m - 0.25)[0.05(L_{BBN} / dB) - 1]}{(f_{mod}/5)^2 + (4/f_{mod}) + 1.5} \text{vacil}$$

where $L_{BBN}$ is the level of the broad band noise, $m$ is the modulation factor[1] and $f_{mod}$ is the modulation frequency. $m$ is here used in a way that is defined in [10]. In general modulation factor/depth is a parameter that has been defined in various different ways in the past and is known to be subject to substantial uncertainty as shown in WPB1 among others.

The following comments can be made on the results of this work:

- the fluctuation strength at a modulation frequency of 1 Hz[2] is approximately 50% the fluctuation strength at a modulation frequency of 4Hz;
- fluctuation strength increases with increasing overall level (maintaining 100% amplitude modulation as the overall level increases) with a 40 dB increase in overall level corresponding to an increase by a factor of approximately 2.5 in the fluctuation strength;
- fluctuation strength increases with modulation depth (maintaining the same overall level as the modulation depth increases) – for the example of a 60 dB overall level, the fluctuation strength is zero until a modulation depth of approximately 3 dB (modulation factor ~17%) after which it increases approximately linearly with the logarithm of the modulation depth until it flattens out at around 30 dB (modulation factor ~94%);
- the fluctuation strength (based on experiments with 70 dB level pure tones amplitude modulated by 40 dB at 4 Hz) shows an insignificant correlation with tone frequency

Lenchine [11] uses the model above to estimate how typical wind turbine noise would vary in fluctuation strength with modulation frequency and modulation factor: both parameters affect fluctuation strength within the same order of magnitude. Legarth [12] conducted listening tests using the model. Participants were asked to rate the fluctuation strength and modulation frequency of artificial stimuli. The author reports satisfactory correlation for the ability of the participants to identify fluctuation strength ("swishing sound"). They were not

[1] In analogy to AM as used in electronic communication the **modulation factor** is a value between 0 and 1 and defined as the ratio of low frequency modulation signal amplitude to un-modulated broadband signal amplitude.

[2] Note that modulation frequency is not the signal frequency. Therefore 1 Hz does not refer to infrasound.

good at rating the modulation frequency. Jiggins [13] reported a study of loudness perception of simulated broad-band sounds of increasing modulation depth. He found that modulated signals tended to be rated louder than their numerical values suggested.

It is important to note that while fluctuation strength seems to describe the perception of fluctuating sound well, the affective response to wind turbine sounds has to be measured by the degree of annoyance in response to the sound and there is no simple or direct link between the two.

## *5.3 Previous work on the annoyance of modulated wind turbine sounds*

Little is known about the effects of AM WTN sound on annoyance. A survey study on wind turbine noise annoyance [14] reports complaints about different types of AM. The authors state that most of the complaints were associated with reports of “swishing” or “lashing”, and with these there were roughly equal number of people annoyed as not-annoyed. There were many fewer reports of “thumping” or “throbbing”, but where this description was used four times as many people were annoyed as not annoyed.

Lee *et al.* [15] conducted listening tests on stimuli in which both the $L_{Aeq}$ and modulation factor parameters varied. The stimuli were generated from two different wind turbines which were recorded in the near field. They were then adapted by changing the masking noise to achieve the desired modulation factor. Annoyance is shown to vary with both parameters. In these results annoyance scales more clearly with $L_{Aeq}$ than with modulation factor. The findings are in agreement with low frequency noise studies such as [16] which is not related to wind turbines. However, for both studies it is difficult to judge how large the spread in the data was and therefore how reliable the results are because error bars are not shown. Also some detailed information on the exact nature of the stimuli was missing in both studies (such as how the $L_{Aeq}$ were achieved). In the absence of more detailed studies such results have been interpreted as preliminary dose-response relations.

Moorhouse *et al.* [17] studied the response of subjects to amplitude modulated low frequency tones using an $L_{A90}$-$L_{A10}$ criterion for tones and found that night time acceptability thresholds changed from a high level at low modulation depth / $L_{A90}$-$L_{A10}$ to a low level at higher values of that parameter. This study is designed to find whether a similar threshold behaviour can be found for AM sounds.

Members of the project research and the management team have observed that WTN AM display high temporal variability of parameters such as modulation depth and $L_{Aeq}$, and can also have varying spectral characteristics.

Based on this review of existing research on the subject, it was decided to design listening tests to derive a dose-response relation for the annoyance from WTN AM sounds using both scaling magnitude estimation and paired comparison methods. These tests would also evaluate the effect of different spectral and temporal characteristics of these signals. Initially, many modulation parameters were piloted in participant Sensitivity Tests to define a final set of stimuli.

# 6. Listening test design

## *6.1 Modulation Model & Parameters*

On the basis an analysis of AM data samples collected as part of other parts of this project WPC, the stimuli were designed using model described in the WPB1 Appendix. For better understanding of all variable properties of WTN AM sound a strongly simplified but representative model of a modulated sound signal is shown in Figure 6.1.

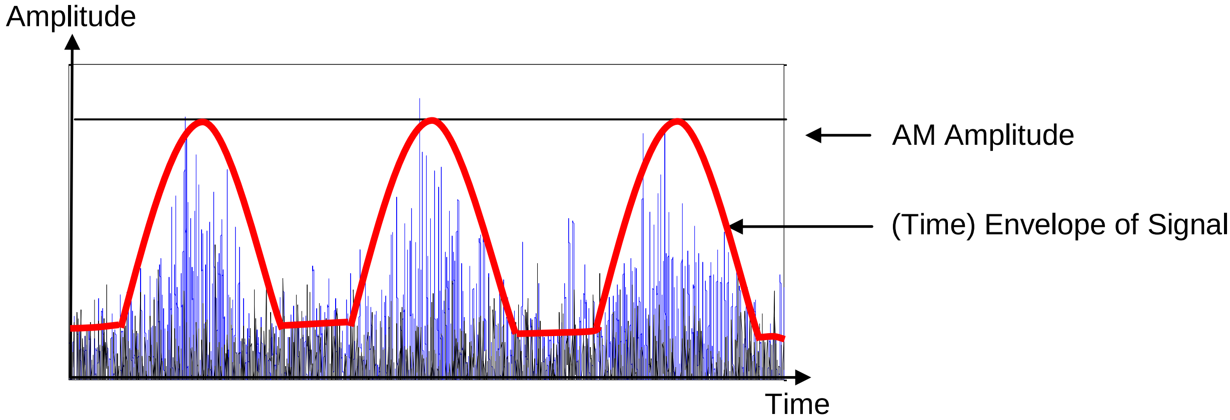



**Figure 6.1 Simplified schematic defining terminology of amplitude modulated signals.**

The signal printed in blue is the AM sound: the modulated wind turbine noise is modelled as pulses of a certain height (variable in the general case), profile shape, width and spectral content. The separation period between the pulses defines the modulation frequency. This varying signal is overlaid over the masking noise (in black), which can have different spectral characteristics (and can be different to the AM pulses in the general case, as represented by different colours in Figure 6.1). The effective depth of modulation is determined by the emergence of the pulses above the masking noise.  In the WPB1 model, both the amplitude of the pulses and their spectra were defined by potentially asymmetric Gaussian profiles (see WPB1 Appendix).

Figure 6.1 illustrates the importance of comparing the absolute amplitudes of a modulated signal with the amplitudes of the masking noise levels because if the two are comparable or the masking noise amplitudes are even higher, then the modulated sound is inaudible. The parameter that describes this ratio is the modulation depth which can be defined in different ways (WPB1). Different definitions can be considered, and the determination of a metric was considered from the outset to be dependent on the outcome of the subjective tests themselves. During the stimuli design, a preliminary measure of modulation depth proved useful to design a representative range of stimuli signals:

> **Modulation depth (MD)** derived from 100 ms averages of $L_{Aeq}$: The modulation depth is defined as the difference between the mean peak level and the mean trough level in the A-weighted RMS time series for any consecutive group of all pulses over the length of the test stimuli (Figure 17.8 and Figure 17.9, Appendix V). A-weighting is a common filter that takes the sensitivity of the human ear to different frequencies into account and is therefore perception related, and the $L_{Aeq}$ acts as a signal envelope. This is comparable to measures such as those proposed by [18].

Other measures of modulation depth were evaluated in WPB1 of the project to consider different metrics of "modulation depth" and how to best relate them to listener reaction.

Whilst the time envelope in Figure 6.1 is symmetrical about the maximum this is not the case for every naturally occurring signal. If the envelope is skewed to one side then the modulation may sound more impulsive. A strongly skewed envelope looks like a sawtooth with a certain slope and a vertical side. The slope can be characterised in terms of rise/fall time or the skew in terms of percentage. At present little is known about the occurrence of impulsive modulation and its impact on perception has not been quantified.

Other properties of the envelope are its width, and its repetition rate which is also called modulation frequency. The modulation frequency of wind turbine noise occurs at the blade passing frequency. For most modern large-scale wind turbines, this modulation frequency is generally between 0.5 and 2 Hz, which is the frequency range that the study has focused on.

Other key characteristics of AM signals, that cannot be easily represented in Figure 6.1, are measures of the frequencies that are contained in the different elements of the sound signal. This is called the spectral content. In WPC the spectral content of AM noise has been shown to vary from sample to sample. Because of the limited database of AM sound samples, generalisations on spectral content are difficult and the stimuli were therefore based on the available knowledge of those AM spectra. The observed far-field data fell broadly into two types of categories with differing spectral content:

A. Dominated by the 500-1000Hz region (mid-frequencies)
B. Dominated by the 200-600Hz region, with a slight-low frequency bias (peaking around 300Hz) (reduced frequencies)

Because of this difference, the first type is more likely to be described as “swish” whereas the second may be described as “swoosh” or “whoomp” because of the increased prominence of lower frequencies. To differentiate between these two spectral types, type A may be labelled **Medium-Frequency Amplitude Modulation** (MFAM) and type B **Reduced-Frequency Amplitude Modulation** (RFAM). These labels will refer in the remainder of this report to the different spectral types.

Corresponding examples of representative frequency spectra are shown in Figure 17.6 (Appendix V) and Figure 16.7 (Appendix IV). Note that the frequency spectra envelope are defined in this study using a "filter function" with its respective properties called centre frequency, bandwidth (BW) and symmetry/skewness.

AM sound from wind turbines can be intermittent and the duration of its occurrence can vary strongly in reality. The effect of intermittency (i.e. cumulative effect of variable occurrence of AM on annoyance over long periods) is difficult to test in a laboratory environment and has therefore not been included in this study. The study of subjective response over long periods (minutes or hours) would require extensive testing periods, even if restricting the range of variables in the stimuli, after which the tests would become increasingly impractical and unrepresentative of realistic situations. The need to consider a sufficiently extensive and representative range of different scenarios of length and repetitions would equally increase testing requirements beyond what was reasonably feasible as part of this project. This was particularly relevant for the paired comparisons discussed in Section 5.1.2.

Apart from AM of broadband noise another type of AM which is the amplitude modulation of tones within a wind turbine sound has also been observed in practice. However tonal modulation is outside the scope of this study. Precise and robust methods already exist to rate non-stationary tones, both in general [19] and for wind turbine noise [2].

## 6.2 Scenarios

Many of the reported complaints about AM were about sound heard indoors, which is not surprising as residents tend to be inside their dwellings at night when background levels are quieter and WTN (including AM if present) will tend to be more audible. It may therefore seem intuitively natural to try to reproduce indoors conditions for listening tests, especially as listening tests for this study were conducted in a listening room; although it should be noted that the characteristics of a listening room will be different to a typical dwelling room (in terms of reverberation and background noise levels). One advantage would be that naturally sounding $L_{Aeq}$ that meet the expectation of the participant can be produced.

There are several problems with indoor stimuli:

- Because every dwelling is different it is unclear how realistic or representative wall sound insulation loss or room acoustics models are. This would introduce further variables including room furnishings, open/closed windows, etc. This would therefore increase test length and uncertainty to an unreasonable degree. And these stimuli would not be representative for all indoors environments.
- Previous attempts at using indoor stimuli with the general research question of whether participant ratings are context dependent [5] did not yield conclusive results.
- The associated significantly lower noise levels associated (as both WTN and MN are reduced to the internal) would be increasingly challenging to test and reproduce even in a controlled and dedicated testing environment,
- The noise environment inside dwellings is often affected by internal sources (fridge, heating systems etc), as well as human activity, including sleeping noises at night.
- Because of these difficulties a limited number of suitable internally measured recordings were available
- Further complications arise through room acoustics in listening rooms: unrealistic stimuli can easily result, particularly for monaural signals.

Wind turbine noise is generally assessed outdoors at a free-field location for similar reasons. Therefore developing a metric based on outdoor measured noise was the preferred option within the research team. For the results of the study this means that the absolute annoyance ratings cannot be directly compared to studies such as [14] that are based on questionnaires about the perception of wind turbine noise in the listeners home environment. This is for two reasons, firstly the context of the home environment can never be fully reproduced in a laboratory. Secondly, outdoor sounds that are played back in an indoor environment at typical outdoor $L_{Aeq}$ sound often too loud to the listener (e.g. [5]). The relative ratings within the study will nonetheless give meaningful results.

Adding local vegetation noise to a WTN or MN stimulus seemed natural, because this type of noise occurs naturally and a stimulus without vegetation noise might sound unrealistic of an outdoor amenity area. Vegetation noise may naturally decrease the audibility of wind turbine noise, and it is the most likely source of masking in rural areas, therefore providing potentially more realistic annoyance ratings.

However, there are a number of counter arguments including that adding another type of noise complicated the tests and was not directly related to the task of establishing a dose-response relationship between amplitude modulation and annoyance. The representativeness of vegetation noise can be questioned because it varies with location and time. For example. in high wind shear conditions wind speeds close to the ground are low while the noise level produced at turbine height is high. Therefore, masking vegetation noise can be limited, so the presence of significant masking would be representative of some conditions, but not all. For the present study a large part of the tests therefore included both stimuli without and with vegetation noise to explore the potential effect of local vegetation noise on annoyance.

## *6.3 Stimulus Design*

Following extensive discussions and preliminary assessments within the project team, the synthesised stimuli were designed in three steps:

1. Firstly, synthesised representative un-modulated masking noise which consisted in some cases of WTN and in some cases of WTN plus garden noise in a certain fixed proportion. The garden noise was kept at a low level (1 GN : 5 WTN) so that it would not dominate the stimuli.
2. Secondly, the modulated part of the signal was added, with varying parameters, according to the model described in WPB1 and above.
3. And finally the stimuli were converted to the optimum format to be played back either over headphones or in the listening room.

The modulation model is based on an analysis of wind turbine noise recording made in the far field rather than near field recordings adjusted with propagation model. Whilst this approach provided reproducible stimuli which could be compared with recordings and enabled a systematic study of the response to certain AM parameters, the stimuli were simplified and could not represent all possible scenarios. However, the use of naturally occurring recorded sounds would encounter the same difficulties but would not assist the systematic control of modulation parameters. They would have the added complications contamination by background noise or other features of the turbine noise such as tonal noise. Recordings vary in distance to turbine; they are typically highly variable in time and artefacts can be introduced by the recording techniques. The use of carefully controlled synthesised stimuli in comparison to recordings was therefore considered to be the only realistic way to derive thresholds of annoyance onset and a systematic study of relations between AM characteristic parameters and mean annoyance score at AM values above the threshold. The resulting stimuli were judged by the project team to sound realistic compared to real WT sound at similar distances.

### 6.3.1 Step 1 – un-modulated masking noise

The previous stimuli design procedure that has been described in [5] was used. It can be summarised as follows:

- Recordings (in accordance with [20]) of 47 wind turbines have been analysed and an average spectrum compiled.
- The spectrum was "propagated" to an immission site that experienced a fixed sound level using the NORD2000 propagation model [21].
- A random white noise of constant amplitude was filtered using the spectrum and the phase randomised to produce the unmodulated WTN sound

The un-modulated WTN spectrum can be found in Figure 17.1 and was considered representative of typical un-modulated WTN in the far-field by the research team. The advantage of this procedure is that the use of average spectra and sound propagation model delivered a set of standardised stimuli. Importantly this did not contain any significant modulation or audible tonal content. Different distances from the turbine in flat terrain can within a limited range be simulated realistically by changing the $L_{Aeq}$ of the stimuli although a propagation model needs to be used where more accuracy is required. Propagation parameters such as ground impedance were well defined and therefore reproducible. The stimuli were based on recordings in the downwind direction and are therefore not representative for other directions.

The local garden noise was chosen to match a noise spectrum of 8 m/s wind through deciduous foliage (see Figure 17.1). This was generated in a similar way as for the WTN above, using a random process adapted to the derived spectrum shape. There again, whilst real recordings may sound more realistic, this would introduce similar concerns of reproducibility and bias with repetition of looped recordings. Both the un-modulated WTN and GN were combined to give the masking noise (MN).

### 6.3.2 Step 2 – modulated noise

AM stimuli were created by overlaying the resulting constant masking noise (MN) with AM pulses according to WPB1. The different modulation parameters which were varied are investigated throughout the different tests below and detailed in Sections 6.4 and 17.5.

### 6.3.3 Step 3 – playback adjustments

This step differed between the sensitivity tests which were conducted using headphones and the final tests which took place in the listening room. The two different reproduction systems were chosen to optimise the test procedure for the different tasks. Using headphones for the sensitivity tests allowed simultaneous testing of up to 10 participants giving flexibility to stimuli design and fast results. A disadvantage of that method was that monaural stimuli sound less realistic than ambisonic reproduction. In contrast, the listening room allows very accurate, realistic directional sound reproduction at sound levels down to just below 20 dB(A)[3], but thereby restricting the number of simultaneous tests to one participant. This latter test environment was judged by the research team to be most suitable to attempt to derive a dose-response relation between a more limited number of modulation parameters and annoyance, as determined from the preliminary sensitivity tests undertaken using the headphone arrangement.

*Multi-participant headphone tests*

To assess the sensitivity of participants to modulation parameters that can vary in wind turbine noise, two preliminary sensitivity studies were conducted (see below). A relatively fast test procedure was employed using the annoyance scale method and calibrated headphones, which allowed several participants to be tested in parallel.

---

[3] Noise floor measurements in the listening room which conforms to ITU-R BS 1116-1 are described in detail in [5].

To achieve a good approximation of a free-field signal at the ear a ‘transfer function’ taking the effect of the headphones and sound card[4] into account was derived and applied to the monaural stimuli. This was done via a 1/3 octave-band filter function implemented in Adobe Audition, firstly with a correction for the headphone/sound card combination[5], and secondly with a correction that compensated for the differences between omnidirectional microphone mono recordings and HATS[6] binaural measurement system as shown in Figure 16.1. This correction meant that close to free field response could be achieved. Using these readily available but not highly specified headphones also resulted in a slightly less accurate reproduction and a higher noise level than necessary for the quietest occurring wind turbine sounds compared to reproduction in a more controlled environment.

It should be noted though that the stimuli in the preliminary stages of the test were not designed to sound totally realistic through the headphones because they were based on monaural signals. They were however designed so that relative changes in modulation parameters would be clearly audible and therefore allow changes in annoyance ratings to be interpreted in terms of listener sensitivity. Absolute annoyance ratings are therefore meaningless for such a design and the results need to be compared to each other to allow meaningful conclusions.

The stimuli file names were randomised to ensure that participants could not guess the nature of the stimulus but the garden noise stimulus which served as a reference sound for these tests was clearly labelled for easy participant access.

*Listening room*

The stimuli for use in the listening room were auralised using a planar ambisonic reproduction technique consisting of a ring of 6 loudspeakers to introduce different directional reproduction for the wind turbine and the surrounding garden noise. Subwoofers were used in a CABS configuration [22]. This is a set of 8 woofers four on the front wall and four on the back wall of the room which achieves a local cancellation of low frequency room

---

[4] External soundcard Mbox 2 Digidesign specifications available at http://www.maudio.nl/UserFiles/Mbox2_DS_36078.pdf [20/3/2012]

[5] Beyerdynamic DT100, specification available at http://www.beyerdynamiconline.com/Datasheets/DT100_DB_E_a3.pdf [18/1/2012]

[6] Head And Torso Simulator (HATS), Brüel & Kjär, Type 4128, specifications available at http://www.bksv.com/products/telecomaudiosolutions/headtorso/headandtorsosimulatorhatstype4128c.aspx [18/1/2012]

modes. This reduces the effect of room acoustics and makes the reproduction as similar to outdoor sound as possible. The system was calibrated using a custom measurement system and modification of the stimuli signals sent to each loudspeaker. The overall approach, equipment and room used is similar to that used in a previous study of the response of subjects to tonal noise stimuli, described in [5].

Wind turbine sound at a listener position would be usually perceived as coming from a specific direction. This was achieved in the test design by playing the stimuli over one loudspeaker right in front of the listener. In contrast garden noise - where present - was played by all loudspeakers in the ambisonic ring to achieve an immersive effect.

The response of the system was evaluated using a sound level meter[7] and a low noise microphone[8]. The superior spatial performance of this reproduction technique was judged by the research team to be more important than the disadvantage of restricted calibration accuracy at frequencies above 600 Hz due to the interference of the ambisonic ring. When listening to the stimuli the characteristics of the broadband sounds were not audibly changed compared to other reproduction systems. For full details of the sound reproduction setup and calibration see Appendix IV: Sound Production System and [5].

## 6.4 Parameter specifications for different parts of the listening tests

The first important question to answer was which modulation model parameters (see Section 6.1) would most determine the affective response of a listener. This was following a process in which different stimuli models and a variety of comments made in the available literature and in subjective reports were reviewed within the project team. Two sensitivity tests were conducted to systematically study and potentially exclude some parameters from the test. Sensitivity Test I was conducted with a very limited number of varied modulation parameters and served mainly to test the stimuli design procedure and to compare rating sensitivity to these parameters using a simplified reproduction technique. Sensitivity Test II was then a more comprehensive and systematic rating exercise to find the modulation

[7] Svantek Svan 959, details available at http://svantek.com/pub/files/File/produkty/datasheet/SVAN959_2011.pdf [18/1/2012]

[8] B&K4179 microphone and B&K2660 preamplifier, details available at http://www.bksv.com/products/transducersconditioning/acoustictransducers/microphones/4179.aspx & http://www.bksv.com/Products/TransducersConditioning/AcousticTransducers/Preamplifiers/2660.aspx [18/1/2012]

parameters to which participants would react most sensitively. That second test enabled the choice of parameters that would be used for the Final Test. However, a few additional parameters were included in parts of the Final Test to validate and extend the results. The rationale for the choice of parameters will be discussed further in Sections 7 to 9 as each set of parameters has been based on the results of the previous tests.

### 6.4.1 AM parameters in Sensitivity Test I

While variation in all modulation parameters can possibly change the affective response of listeners the aim of the first test was to use a very simple modulation based on the present knowledge about AM WTN signals. An obvious choice was to systematically vary the modulation depth to find out whether there is evidence that the affective response might change in a continuous way with increasing modulation depth and whether a sudden onset of annoyance will be observed. The shape and width of the time envelope of amplitude modulated wind turbine noise (Figure 6.1) were parameters that were easily changed in the WPB1 model by at the time and were therefore included in the first test.

Increasing the peak amplitude of the AM pulses, through the increase of the input MD, achieved a progressive increase in modulation depth. It was thought to be important to keep the level and therefore the annoyance from the background noise constant. However, this meant that the stimuli get louder with increasing modulation depth, both of which parameters could contribute to changes in the affective response (see Section 5.2).

Two other parameters that might conceivably change the character of the stimulus and therefore the listener response were shape and width of the time envelope depicted in Figure 6.1 [23]. The chosen waveform was a saw-tooth-type shape with varying rise times. From analysis of wind turbine noise recordings data (particularly for OAM), this was judged by the research team to be a feature of some of the available recordings. This was modelled by an asymmetry percentage factor (% rise time).

A fixed but representative modulation frequency of 0.8 Hz was used. The duration of the modulation pulses was raised as another factor which may potentially influence subjective response, as it may be related to descriptions of “impulsiveness” of the signals. The pulse length was varied from very short pulses of 0.1 s duration to long pulses of 0.45 s. Table 6.1 gives an overview over the chosen modulation parameters. A garden noise was added for comparison. A detailed list of stimuli including all signal parameters is available in Table 17.3.

| | | |
|---|---|---|
| MD, dB(A) | 7 values | 1, 1.2, 1.5, 2, 4.1, 5.6, 8.0 |
| Envelope rise time In % of width | 9 values | 10, 20, 30, 40, 50, 60, 70, 80, 90 |
| Envelope Width (s) | 8 values | 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45 |

**Table 6.1 Stimuli design parameters for Sensitivity Test I. Main modulation parameters under investigation were the modulation depth, shape and width of the modulation pulse envelope. Other parameters were kept constant.**

### 6.4.2 AM parameters in Sensitivity test II

To further evaluate the number of variable AM parameters to establish meaningful dose-response relations and thresholds, a second preliminary sensitivity study was conducted including a further set of modulation parameters that were thought to be possibly influencing listener reactions. Also, the issue that had been observed in the first test where an increased modulation depth inevitably resulted in louder stimuli was considered by creating three independent sets of stimuli:

To differentiate between the annoyance from masking level (ML), peak level of the modulated part of the signal (highest value in blue curves) (MPL) and modulation depth MD were varied as shown in Figure 6.2 a) – c):

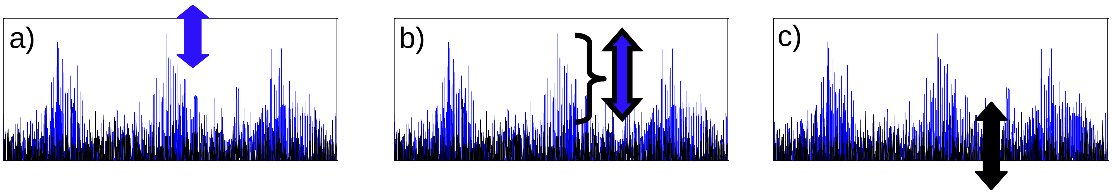



**Figure 6.2 Visualisation of parameter changes a) ML constant, MPL and MD change simultaneously; b) MD constant, ML and MPL change simultaneously, c) MPL constant, ML and MD change simultaneously**

The aim of this design was originally to separate the response to the parameters. All scenarios necessarily resulted in a variation in overall energy content of stimuli. Although this variation in level is known to affect annoyance, the effect might be compensated by the fact that participants could initially adjust the level of the masking garden noise to suit their preconception of how garden noise would sound.

Spectral shapes were varied by using

- masking noise which in addition to the WTN sound included two different types of garden noise  (artificial and real recording) to explore the influence of garden noise on annoyance, The amplitude ratio between WTN and GN was 5:1.
- differing values for bandwidth, peak frequency and frequency skew to create a range of spectra representative of AM recordings

Two modulation spectra labelled as ‘MFAM’ and ‘RFAM’ were used similarly to those defined in Section 17.2. However in this test, a preliminary spectral shape was used: ‘MFAM’ had a peak frequency at 600 Hz and a bandwidth of 350 Hz and ‘RFAM’ had a peak frequency of 300 Hz and a bandwidth of 400 Hz. The sets of stimuli generated by these conditions were tested with two types of garden noise to assess the effect on annoyance. The modulation spectra were also found to be asymmetric, based on analysis of measured spectra in WPC: this was also modelled using a percentage factor. Different skew factors and therefore spectral shapes were investigated as specified below. Finally, different modulation frequencies were tested (0.65 and 1.3 Hz), again based on observations from WPC.

In detail the following parameters were used:

| | | |
|---|---|---|
| MD, dB(A) | 4 different values | RFAM: 1, 2, 4, 6<br>MFAM: 5, 8, 12, 15 |
| Period (s) | fast and slow turbines = 2 values | 0.65 and 1.3 |
| Bandwidth (Hz), Peak Frequency (Hz), Frequency Skew (%) | 2 types used | 'RFAM': 350 (peak), 400 (width), 70 %<br>'MFAM': 600 (peak), 350 (width), 50 % |
| Frequency Skew (%) | 6 levels to investigate the role of frequency skew systematically | MFAM frequency skew of 67, 70 and 80 % corresponding to BW 300, 400, 500 Hz<br>RFAM frequency skew of 30, 50 and 61 % corresponding to BW 250, 350 and 450 Hz |
| Background Noise Type | WTN + 2 different GN spectra in an amplitude ratio of 5:1 | • WTN + GN based on the filtered synthesised garden noise<br>• WTN + GN based on a direct recording[9] |
| Total Level Change | Results from other changes | |

**Table 6.2 Stimuli design parameters for Sensitivity Test II**

The following parameters were fixed for all stimuli:

- Envelope width = 0.2 s
- Envelope rise time = 70% of pulse length

A detailed list of stimuli filenames with their respective parameters can be found in Table 17.4.

### 6.4.3 AM parameters for Final Test (incl. Validation Tests I and II)

The main purpose of the final test was to investigate from which modulation depth annoyance starts to increase and, when modulation exceeds that threshold, whether there is a continuous increase of annoyance with increasing modulation depth or whether annoyance

[9] The WTN is the same in both cases but the second type of GN is identical to GN used in [5: Part B].

assumes a constant value from a certain threshold. The stimuli were normalised to keep the overall $L_{Aeq,20s}$ of the stimuli constant at defined and calibrated levels, for the different modulation depths. Three main representative $L_{Aeq}$ values of 30 to 40 dB(A) were used to compare the well-researched response to increased sound level with the less well-researched response to modulation depth. The stimuli were limited to one type, RFAM, without garden noise. This main test contained 32 stimuli per participant.

For the purpose of designing a representative range of stimuli, the modulation depth was varied by changing the modulation depth (MD), as evaluated by averaged $L_{Aeq,100ms}$ readings, as presented in Section 17.4. The main test included values of MD between 2 to 12 dB(A). Modulation depth intervals were chosen to be less than 1 dB(A) below 5 dB(A) and larger intervals from 5 dB(A) up to 12 dB(A). The choice of high resolution at low MD was aimed at determining a threshold of the onset of annoyance as it was thought that AM signal characteristics might not be perceptible at low MD [9]. Whereas maximum modulation depths were reported by the research team to be occurring at a maximum of 10 dB(A) in reality, the range of the synthetic stimuli was extended to the higher value of 12 dB(A).

For validation purposes, (validation I), subsets of participants also listened to RFAM and MFAM with and without garden noise, and considering additional $L_{Aeq}$ levels and intermediate MD values. The number of additional stimuli per participant was up to 160 for this part of the test.

As part of these final tests, in addition to providing an absolute annoyance rating (*scaling magnitude estimation*), the participants compared the AM stimulus to an unmodulated wind turbine sound with the same spectral shape as the one that the modulated stimuli were based on. The participants adjusted the level of this unmodulated sound until it was equally annoying as the modulated wind turbine sound (*paired comparison method*). This Adaptive BroadBand Stimulus (ABBS) used the un-modulated broadband WTN noise, and was therefore identical to the un-modulated (0 dB) stimulus (MN without garden noise).

As concerns were raised within the project team that the effect of normalising for set $L_{Aeq}$ levels in the stimuli might reduce the relative effect of the increased modulation rate, a second validation test (II) therefore compared the results from the main test to a set of stimuli that had a constant masking noise (MN) level, with increasing modulation depth and corresponding increasing $L_{Aeq}$. By using the adaptive test method and normalising for the $L_{Aeq}$ of the test stimuli, the influence of the overall level on annoyance was limited. The additional number of stimuli for this part of the test was 32.

Detailed lists of stimuli for the first and the second subgroup of participants can be found in Table 17.6 and Table 17.7. The following parameters were used:

| Key parameters (main test in **bold**) | | |
|---|---|---|
| Modulation depth MD, dB(A) | 8 different values in main test plus 5 intermediate values for validation I | **0, 2, 3, 4, 5, 6, 9, 12**<br>1, 2.5, 3.5, 4.5 and 7 |
| Sound level of total stimulus $L_{Aeq}$ dB(A) | 3 values plus<br>2 for validation I | **30, 35, 40**<br>25 and 45 |
| Additional parameters for validation I (main test in **bold**) | | |
| Type of AM | 2 types | MFAM, **RFAM** |
| Type of Masking | 2 types | **WTN**, WTN + GN |
| Additional stimuli for validation II | | |
| Modulation depth MD, dB(A) | 8 different values in main test | 0, 2, 3, 4, 5, 6, 9, 12 |
| Sound level of masking noise, $L_{Aeq}$ dB(A) | 3 values plus | 30, 35, 40 |

**Table 6.3 Stimuli design parameters for Final Test**

## *6.5 Participant recruitment and screening*

### 6.5.1 Sensitivity tests

For the two sets of sensitivity tests, the participants were directly recruited from the staff and student population of the Acoustics Research Centre at the University of Salford by email and word of mouth. As these tests were designed to identify the physical modulation parameters that can be significantly distinguished, there was felt not to be a need to aim for a representative population sample. Because of the preliminary nature of the sensitivity tests participant details for these were not recorded. The expected higher number of expert listeners in an acoustic department can be an advantage in that situation. For the Sensitivity Test I, 5 students from local secondary schools who were on a work placement and 8 postgraduate students from the Acoustics Research Centre at the University of Salford volunteered as participants. And for Sensitivity Test II, 11 postgraduate students and staff members students from the Acoustics Research Centre at the University of Salford volunteered as participants. The number of participants in the sensitivity tests was large enough to test stimuli design and rating procedures and to decide which modulation parameters dominate the response. The number was too small to be used for statistical analysis.

### 6.5.2 Final tests

For the final tests the sampled population was initially contacted via an article placed on the main University of Salford website and the separate websites for the staff and students at the university in September and another article in late November 2011 to increase participant numbers. The articles are detailed in Appendix I. The articles explicitly mention wind turbine noise in an attempt to attract attention. While this has the potential disadvantage of the attitude towards the source affecting judgements this could not be avoided as the stimuli would have been likely to be identified as wind turbines by several participant in any case. The article mentioned that volunteers for participation were requested and that they would be paid. It also mentioned that participation was subject to a screening procedure. This procedure entailed the volunteer providing their names, age, nationality, occupation, sex, and previous listening test experience; then volunteers completed several multiple-choice questions about the type of area they live in (see Appendix II). Participant details were recorded but kept confidential. Non-leading questions were used to prevent responders to the advert from falsely claiming to belong to the population of interest.

In comparison to some of the studies referenced in Section 5, the noise sensitivity of volunteers was assessed in the screening procedure by the Zimmer and Ellermeier short noise sensitivity measure (Appendix II). This is a 9-item self-reported questionnaire that asked the participant to either strongly agree, slightly agree, slightly disagree or strongly disagree with statements about disruptions caused by everyday noises. This was deemed a useful measure as an individual's sensitivity to noise may influence how annoying they perceive sounds to be [28] and would therefore inform the choice of participants.

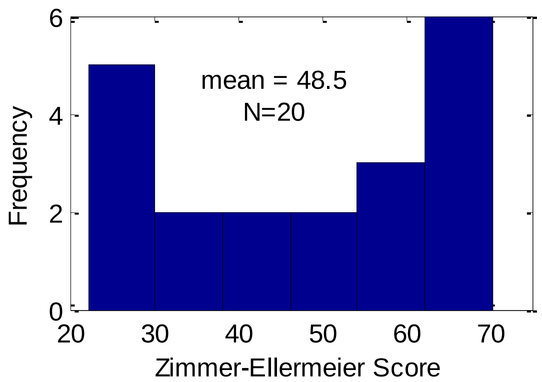


**Figure 6.3 Zimmer and Ellermeier noise sensitivity score distribution for participants (final tests).**

The sensitivity distribution in Figure 6.3 has a similar mean value as for the participants in [5] which was 49.8. The distribution is very different though with six participants being very sensitive to noise as seen by their high scores and 5 participants being very insensitive.

Another criterion for participant selection was that their hearing was not impaired. This aimed to recruit participants with 'normal' hearing for their age, rather than a sample of particularly sensitive or impaired hearing participants and was necessary because some stimuli were close to the hearing threshold. If not heard, the results for these stimuli would have been meaningless. Additionally to choosing participants on their assertion that they were of normal hearing, an audiometric test was performed for each participant confirming that all volunteers were of normal hearing and nobody needed to be excluded because of significant hearing loss.

A total of 20 volunteers, 8 female and 12 male, participated in the final tests. The age distribution is centred between 20 and 30 as shown Figure 6.4a) with only 6 participants in their mid-thirties and beyond.

A subgroup of 11 volunteers, 4 female and 7 male, participated in the first validation of the final test. The age distribution is mainly centred between 20 and 30 as shown in Figure 6.4b) with only two participants older than 40 years. Another subgroup of 9 volunteers 4 female and 5 male, took part in the second validation of the final test (Figure 6.4c).

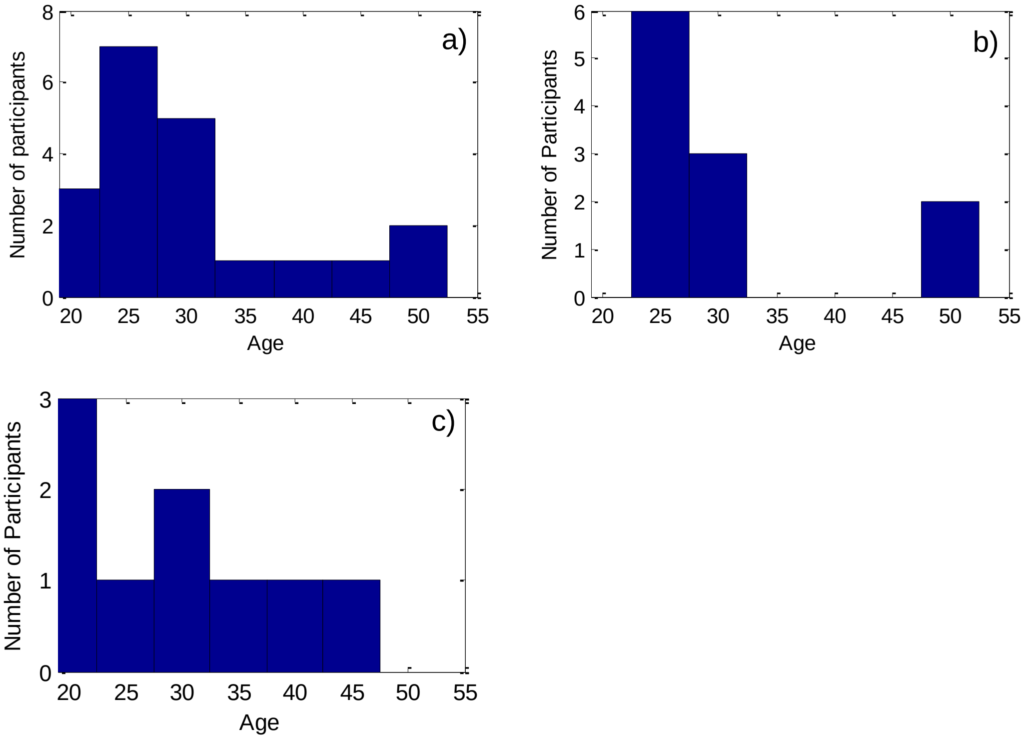



Figure 6.4: Age distribution of participant a) main test, b) subgroup 1, c) subgroup 2.

### 6.5.3 Participant protection

The data set from the screening process contained potentially sensitive personal information. Therefore it was stored in password protected spreadsheets on a secure server that was only accessible by project staff. No other copies were kept. Outsides these files participant information was made anonymous by the use of ID numbers; therefore the data could only be traced back to the participant via the protected spreadsheet. Informed consent forms as specified in Appendix I were signed by each participant in accordance with standard University procedures.

## *6.6 Participant briefing*

### 6.6.1 Sensitivity Tests

Participants were given the test sheets (Figure 18.1 and Figure 18.2), and told to put headphones on. Headphone positions were checked. Stimuli of 10 second length were presented in a looped configuration using the Windows™ Media Player. The listening time for each stimulus was therefore controlled by the participants. Participants were then instructed to listen to the garden noise stimulus and told that the stimulus represents the sound of wind in trees and bushes. There was no aim in these tests to closely represent the situation of a resident relaxing in his garden, as all were done in a standard university computer music room, and the tests therefore focused on the comparative effect of different stimuli parameters. Using the onscreen volume slider participants were instructed to adjust the level until the sound was reflective of conditions in which they would spend time in their garden relaxing, but still audible. The level adjustment resulted in a range of stimuli $L_{Aeq}$ shown in Figure 8.1. It was useful because it avoided the situation when a participant feft the stimuli to be very unrealistic and therefore more difficult to rate for annoyance which had been found to be the case in [5]. The volume was subsequently not changed so that the $L_{Aeq}$ of all stimuli would be traceable relative to each other. After Sensitivity Test I, it was recognised that the information of $L_{Aeq}$ for every participant was required for data analysis and participants were told make a note of the volume in Sensitivity Test II.

In both tests participants were then asked to rate on a numerical scale between 0 and 10 (Figure 18.1 and Figure 18.2) how annoying the garden noise was if they were sitting in their garden trying to relax after a hard day's work. The rating scale was similar but not identical to the standard rating scale [24] in that the maximum rating was described as "very annoying" in contrast to the "extremely annoying" suggested in the standard but in agreement with studies like [4]. The context information was given to make the ratings as similar as possible to ratings that would occur in the participant's home environment. The garden noise served as a reference to compare responses between a noise that is usually experienced as pleasant to responses to noises that are known to attract complaints. The annoyance rating of garden noise was introduced as a measure of how difficult participants found it to imagine a pleasant noise in a laboratory environment. This was in case they were not able to make the sound 'not at all annoying'.

The stimuli were then presented to participants in random order which is important to avoid fatigue bias[10]. Participants were asked to rate the stimuli on the same scale and in the same context as the garden noise. Therefore the response to garden noise became the reference to which all other responses could be compared.

Due to the random order of stimuli presentation a participant might listen to a number of, for example, very quiet stimuli to start with and then realise that their ratings for the very loud stimuli would not fit within the scale. Therefore participants were given the option to manually select stimuli to re-listen to any sounds and amend prior ratings. The importance of not

[10] An effect on the statistical data analysis from systematically different ratings between the beginning and the end of a test.

listening to the sounds in alphabetical order by manual selection was pointed out to participants.

### 6.6.2 Final Test

When volunteers first arrived they were instructed to do a standard audiometric test to ensure their hearing was adequate for participation. They were then shown into the listening room where they were handed an instruction sheet (Figure 18.3). When they were ready to start, stimuli were played back using a setup shown in Figure 15.1 and described in more detail in Appendices IV and V and [5] including pictures of the room in its final set up. The room was set up to try and approximate the feel of an outdoor amenity area used for relaxation. The participant was facing the direction from which the wind turbine sound was played from the front loudspeaker as shown in Figure 15.1. When garden noise was added, the sound was produced from all loudspeakers to make the directional impression as authentic as possible by making it immersive.

Participants used a touch screen to rate the annoyance of stimuli and adjust levels of the unmodulated ABBS to equal annoyance level as shown in the graphical user interface (Figure 6.5). The 20 second stimuli were looped and the total listening time for each stimulus was controlled by the participant. To test the different responses for sliding scales and for equally annoying ABBS the participants were asked to:

- first directly rate the annoyance of a sound on the sliding scale (*scaling magnitude estimation*) which resulted in annoyance ratings from 0 to 10 on a numerical scale;
- then for the same stimulus, adjust the ABBS level so that the (un-modulated) sound becomes equally annoying to the AM test sound (*paired comparison method*).

The rating scale in Figure 6.5 is an 11 point scale like the one used in the sensitivity tests but instead of discrete values a continuous slider was implemented for ease of use.

Participants started the test off with practice ratings until they were comfortable with the task and decided to go on to the main test. Stimuli were then presented in random order, which changed for each participant. The listening test procedure is outlined in further detail in Appendix II.

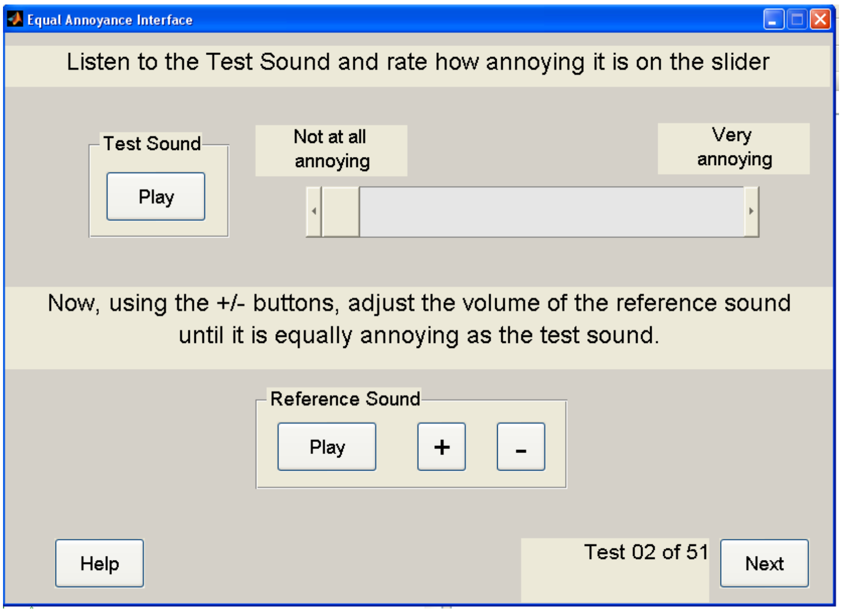



**Figure 6.5: Graphical user interface for final test. Participants were first asked to rate the annoyance of the AM test sound on the sliding scale in the top of the window and to then adjust the level of the ABBS (Reference Sound in the GUI) to match the annoyance of the AM test sound.**

# 7. Results of Sensitivity Test I

To assess the sensitivity of participants to a subset of the possible modulation parameters that can vary in the model of WTN AM produced, a fast test procedure using the annoyance scale method and calibrated headphones was designed, and two preliminary sensitivity studies were conducted.

In Sensitivity Test I results of participant annoyance ratings are shown as a function of a very limited number of modulation parameters to demonstrate the viability of the stimuli design and sound reproduction method (Section 6). Some initial conclusions on sensitivity to the modulation parameters are drawn.

The sensitivity tests were designed using the amplitude related model input parameter α defined in Section 6.1. However, a simple modulation depth (MD) metric, based on A-weighted 100 ms $L_{Aeq}$ averages peak-trough levels (see 17.4) seemed more useful (in a first instance) to provide context to participant responses, because this measure is expected to be more directly related to the frequency response of the ear. Although the test stimuli were not designed for equal distribution of this parameter, for consistency with the following sections, results are displayed as a function of MD.

It was observed that participants in their first task adjusted garden noise sound levels to very different values in their judgement of what sounded natural. The range was from just audible to medium levels. Because the level was not changed subsequently all other judgements were relative to this garden noise level.

In Figure 7.1 the mean annoyance score ranged between 3 and 4 and rose to a value of 7 for a modulation depth of 8 dB(A). Increasing annoyance scores with increasing modulation depth have also been found by other authors ([15] and [16]). This result is also in agreement with theory on psychoacoustic annoyance which increases with fluctuation strength [9].

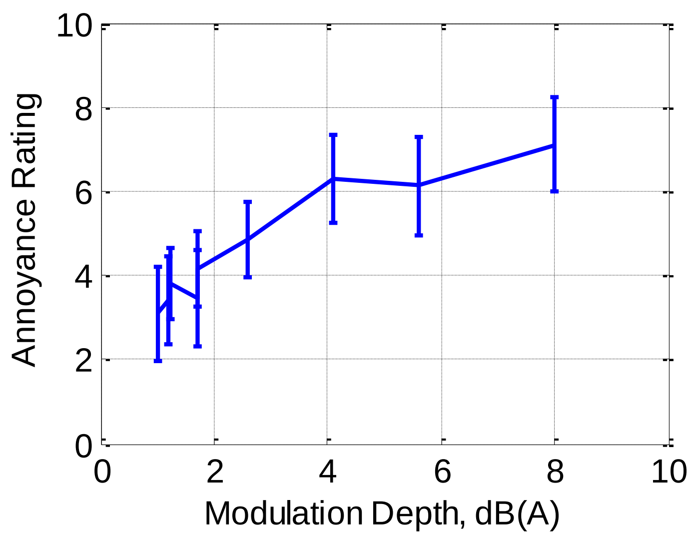



**Figure 7.1: Annoyance rating for changes in modulation depth (MD).**

The error bars in Figure 7.1 and all subsequent figures represent 95 % confidence intervals (CI)[11] . With a range of about two points on the rating scale they are relatively large in this case. There are three possible reasons for their size. Firstly the number of participants was small. Secondly every participant had chosen their individual reproduction volume. And lastly affective responses are generally influenced not only by the sound characteristics but potentially by many other factors as pointed out in Section 5. The error bars for very small

[11] A 95 % confidence interval is an interval in which a measurement or trial falls corresponding to a 95 % probability

and large modulation depths do not overlap. This suggests that the rise in annoyance is possibly significant.

Because of the stimuli design, signal energy/loudness increases with modulation depth. But it was noted that the rise in annoyance could be due either to the change in modulation depth or to the increase in signal energy/loudness. The error bars are also too large to conclude whether annoyance ratings start to increase from a particular modulation depth or whether the increase is continuous from the lowest modulation depth. [9] suggest that there should be a "threshold" modulation depth which is at the limit of perception and a continuous increase of annoyance above that threshold. However, a wind turbine related study [14] reports that Fastl and Zwicker's metric was not sufficient to explain annoyance ratings.

Figure 7.2 shows the annoyance ratings as a function of signal shape. The change in rise time proportion ranged from 10 – 90 %, from a sharp saw-tooth signal (<50%), to symmetric pulse (50%), to a saw-tooth in the other direction (>50%). The modulation depth of the stimuli was constant at 1.7±0.2 dB(A). Annoyance was rated at values between 3.8 and 4.5 but a trend is not obvious. This is clear from the 95% confidence intervals. . The average annoyance at 80% rise time was slightly higher than other values but the large error bars suggest that this is probably not reproducible. The rating could well be due to temporal masking effects where a change in modulation can take up to 200 ms to be fully perceived [9] and therefore the difference between a stimulus with a rise time of 10 % can possibly not be distinguished from a stimulus with a rise time of 20 %. The relatively small modulation depth might also have contributed to difficulties in distinguishing between stimuli of different envelope shapes. In spite of the small modulation depth used it seemed nevertheless unlikely that the envelope shape would influence annoyance strongly.

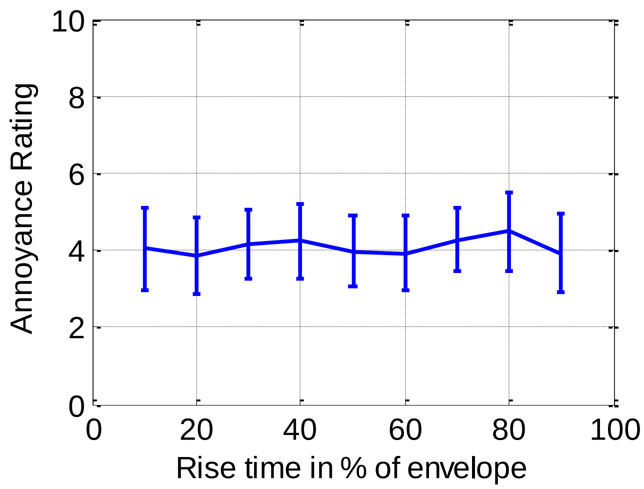



**Figure 7.2: Annoyance ratings as a function of signal shape.**

Figure 7.3 shows the statistics of annoyance versus pulse length. Most mean values were close to an annoyance rating of 4. The stimulus with 0.3 s pulse length was omitted because of an error in the signal design stage: that stimulus sounded quite different and could therefore not be compared with the others, and this was reflected in average participant ratings of 6. After exclusion of this erroneous stimulus, it can be seen that the effect of pulse length (and associated potential "impulsivity") does not appear significant. This might possibly partly be due to temporal masking effects.

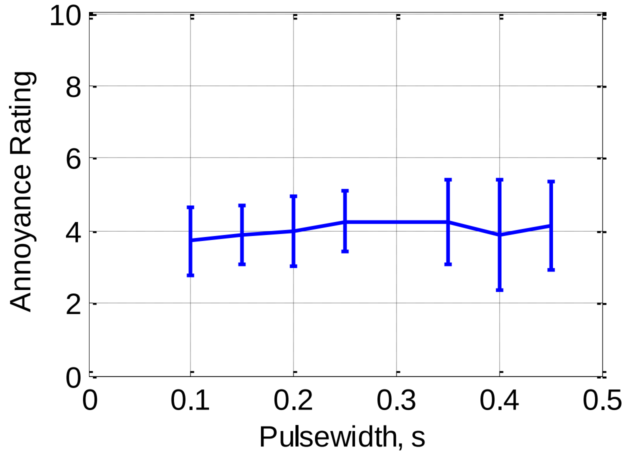



**Figure 7.3: The effect of pulse width on annoyance.**

## *7.1 Conclusions*

The preliminary sensitivity tests focused on the following parameters: modulation depth, pulse shape and pulse width. The annoyance ratings showed a possible systematic sensitivity to modulation depth/signal amplitude, but the effect from the increased signal energy was not separated from that of the increased modulation level in itself. A low sensitivity to pulse shape and/or pulse width was found.

It has been observed that realistic noise levels of garden noise were difficult to judge in a laboratory environment as the values between participants varied widely. Exact levels were felt to be useful and monitoring was therefore implemented for Sensitivity Test II. The range of $L_{Aeq}$ was additionally restricted to realistic levels which will be discussed in the context of Sensitivity Test II results in Section 8.

The test design for further stages of the project was aimed at finding out which other modulation parameters might affect annoyance, evaluating the effect of increasing signal energy, and whether sensitivity to modulation depth varies linearly with modulation depth or is governed by threshold behaviour.

# 8. Results of Sensitivity Test II

While Sensitivity Test I focussed on the temporal characteristics of the modulated signal, Sensitivity Test II was designed to address the effect of relative levels of signal components and the frequency content of the stimuli.

In contrast to Sensitivity Test I, participants made a record of the volume level when they adjusted the playback volume on a scale from 0-100% to a level at which they judged the garden noise to sound realistic. It should be noted that following the observations in the first test, the volume range available to participants had been reduced considerably in comparison to Sensitivity Test I to avoid unrealistic choices. The average adjustment for the GN was therefore 40±4.5% of the volume slider length with the lowest value at 30% and the highest occurring at 70% which indicates that the chosen range was well within the expected level for a garden and reproduces the finding that realistic noise levels have been difficult to judge. This was probably not helped by the fact that they can vary widely in reality. The average annoyance ratings for this garden noise were 2.95±0.64 with a minimum of 0 and a maximum of 6. All other stimuli were judged relative to the garden noise level.

The range of stimuli $L_{Aeq}$ that resulted from the adjustment of GN is shown in Figure 8.1. The figure also contains the 95% CI resulting from participant adjustments which span about 2 dB(A). This is small due to the limited choice available to participants. It can be seen that the quietest stimuli were reproduced at an average $L_{Aeq}$ of just above 28 dB(A). Note that the choice of x-axis in Figure 8.1 is only a convenient way to categorise the stimuli and that therefore the $L_{Aeq}$ are not per se a function of modulation depth.
The room background noise level in the headphones was measured using the HATS system and was between $L_{Aeq}$ = 30 - 35 dB(A). In effect, it has to be assumed that the annoyance ratings for the quietest stimuli were affected by the room background noise although the modulation was always audible. The average $L_{Aeq}$ values are therefore specified for all figures in this section.
The wide range of background levels is due to a number of factors, firstly the noise in the room depended on how many computers were operated at a particular time, secondly the volume of the external soundcards had to be adjusted manually to about 25% of the available volume range, thirdly the headphone efficiency had to be estimated and assumed to be identical for all headphones.

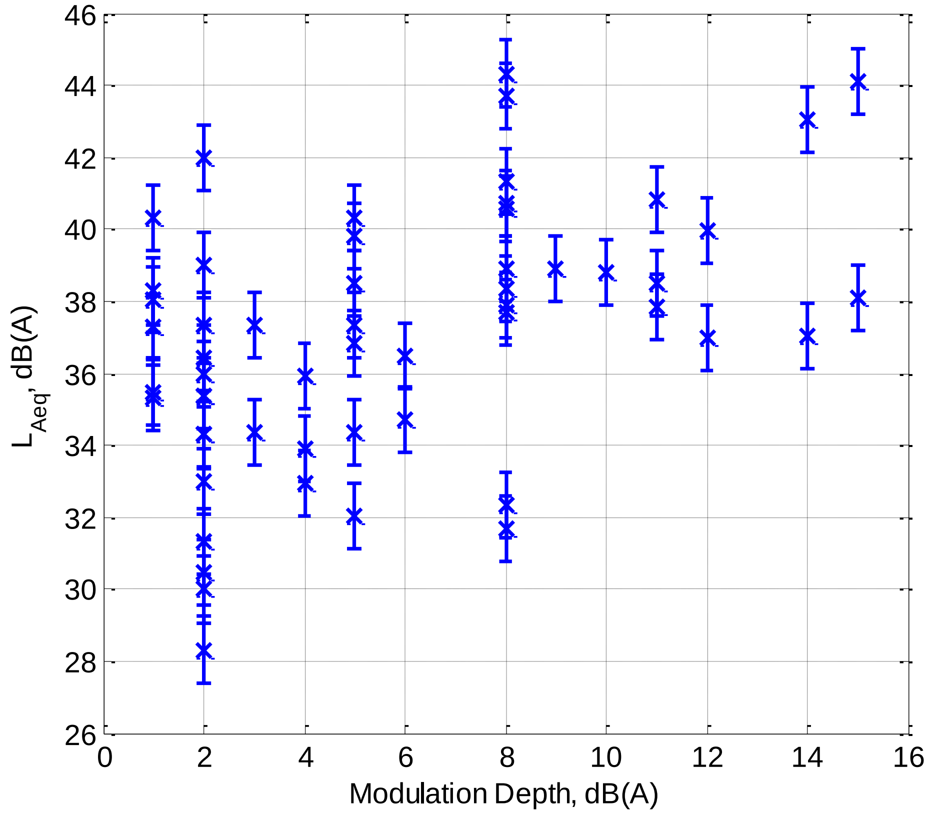



**Figure 8.1 Mean A-weighted stimuli levels after participant adjustment.**

In this test, three different ways of varying the modulation depth were investigated for two types of modulation: MFAM and RFAM (see Section 6.4.2). This involved varying in turns: the modulation peak level (MPL), masking level (ML) and their relative ratio (MD). MD was also kept constant in one set of stimuli to control for the effect of the overall $L_{Aeq}$. The masking noise type, modulation spectra (frequency range and shape) and modulation frequency were also varied.

### 8.1.1 Stimuli set I: Fixed masking noise level

Firstly, the modulation peak level (MPL) was increased with constant masking noise. Data are summarised in Table 8.1. In Figure 8.2, the mean annoyance score is shown as a function of MD for the two types of stimulus (MFAM and RFAM) and two types of GN. GN was part of the masking noise which also contained WTN with an amplitude ratio of 1:5, respectively.

In the figure the "unmodulated" stimulus was the respective type of GN on its own with a natural modulation depth of about 1 dB(A) for both types of GN. These GN stimuli served as reference sounds to compare with the AM stimuli.

The average annoyance for AM stimuli ranged between 3 and 4 and rose to a value of 8 for MD = 15 dB(A). 95% CI for the red curve span 0.3 - 1 points (half the error bar) on the 11 point rating scale for AM stimuli and 2 points for the GN (shown as MD = 0 dB(A)). Note however that increasing modulation amplitude contributed to the increase of annoyance toward higher MD because of the increased acoustic energy in the stimuli. This can be seen in Table 8.1. The table also shows that the high annoyance ratings for the GN coincide with large values of $L_{Aeq}$.

| | | **MD, dB(A)** | **$L_{Aeq}$, dB(A)** | **Rating** | **95% CI** |
|---|---|---|---|---|---|
| GN1 | MFAM | GN1 | 38 | 3.6 | 1.4 |
| | | 5 | 34 | 4.6 | 0.9 |
| | | 8 | 38 | 5.6 | 0.9 |
| | | 12 | 40 | 7.0 | 0.5 |
| | | 15 | 44 | 8.1 | 0.6 |
| | RFAM | GN1 | 38 | 3.6 | 1.4 |
| | | 1 | 34 | 2.6 | 0.7 |
| | | 2 | 34 | 4.9 | 0.7 |
| | | 4 | 36 | 6.2 | 0.6 |
| | | 6 | 36 | 7.1 | 0.7 |
| GN2 | MFAM | GN2 | 40 | 4.7 | 2.0 |
| | | 5 | 37 | 4.7 | 0.6 |
| | | 8 | 38 | 6.5 | 0.5 |
| | | 11 | 41 | 7.4 | 0.7 |
| | | 14 | 43 | 8.4 | 0.7 |
| | RFAM | GN2 | 40 | 4.7 | 2.0 |
| | | 1 | 35 | 3.9 | 1.0 |
| | | 2 | 36 | 5.0 | 0.8 |
| | | 3 | 37 | 6.1 | 0.7 |
| | | 5 | 38 | 7.4 | 0.3 |

**Table 8.1 Stimuli data for which modulation peak level (MPL) was varied and the masking noise kept constant.** $L_{Aeq}$ **95% CI = ±1 dB(A).**

Although the general trend both for MFAM and RFAM in Figure 8.2 is a similar total increase with MD/MPL, it seems that the MFAM sounds increase in annoyance faster with MD than the RFAM. It is therefore possible that the response to RFAM and MFAM is significantly different especially given that the relative stimulus $L_{Aeq}$ were comparable as seen in Table 8.1.

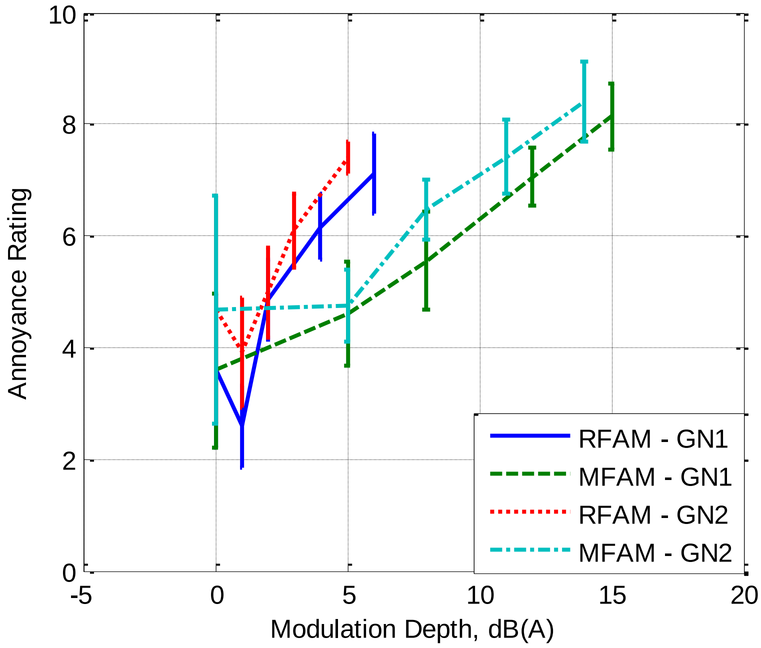
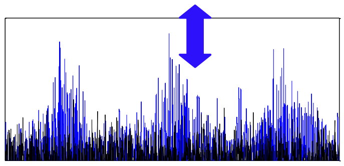



**Figure 8.2: Annoyance rating for fixed background noise level and increase in MD (and peak level) (as indicated by inset) for RFAM and MFAM compared.**



### 8.1.2 Stimuli set II: fixed modulation depth

Secondly, the stimulus $L_{Aeq}$ was increased at constant MD. Data are summarised in Table 8.2. In Figure 8.3, the mean annoyance score is shown as a function of $L_{Aeq}$ for the two types of stimulus MFAM and RFAM and two types of GN.

| | | MD, dB(A) | $L_{Aeq}$, dB(A) | Rating | 95% CI |
|---|---|---|---|---|---|
| GN1 | MFAM | 8 | 32 | 4.8 | 0.7 |
| | | 8 | 35 | 5.6 | 0.7 |
| | | 8 | 38 | 5.6 | 0.9 |
| | | 8 | 41 | 6.8 | 0.5 |
| | | 8 | 44 | 7.3 | 0.8 |
| | RFAM | 2 | 28 | 3.0 | 0.9 |
| | | 2 | 31 | 3.7 | 0.8 |
| | | 2 | 34 | 4.9 | 0.7 |
| | | 2 | 37 | 5.6 | 0.5 |
| | | 2 | 41 | 6.4 | 0.6 |
| GN2 | MFAM | 8 | 32 | 5.0 | 0.8 |
| | | 8 | 35 | 5.8 | 0.8 |
| | | 8 | 38 | 6.5 | 0.5 |
| | | 8 | 41 | 7.4 | 0.6 |
| | | 8 | 44 | 7.5 | 0.6 |
| | RFAM | 2 | 30 | 3.6 | 0.6 |
| | | 2 | 33 | 4.4 | 0.7 |
| | | 2 | 36 | 5.0 | 0.8 |
| | | 2 | 39 | 5.6 | 0.7 |
| | | 2 | 42 | 6.8 | 1.0 |

**Table 8.2 Stimuli data for which the modulation peak level (MPL) and the masking noise were varied simultaneously to keep the modulation depth constant for each AM type.**

$L_{Aeq}$ **95% CI = ±1 dB(A)**

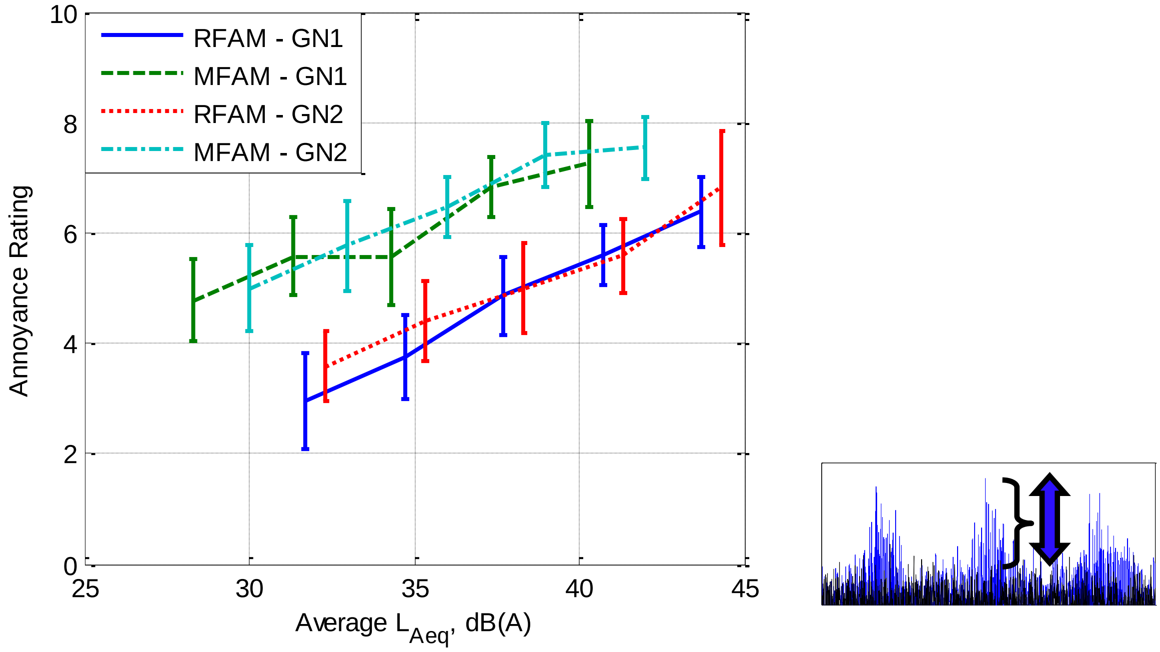



**Figure 8.3: Annoyance rating for fixed MD of 8 dB(A) for MFAM and 2 dB(A) for RFAM and increasing background noise and peak level (as indicated by inset) resulting in a systematic change of overall $L_{Aeq}$ in steps of 3 dB(A).**

The average annoyance for the 4 groups of AM stimuli consistently increased with $L_{Aeq}$. 95% CI for the red curve span 0.5 - 1 points (half an error bar) on the 11 point rating scale for AM stimuli. MFAM ratings are higher than RFAM ratings. However the MD levels for the stimuli types are different too. Both levels of modulation were the same before A-weighting but turned out to be significantly different in the MD metric after A-weighting with MD = 2 dB(A) for RFAM stimuli and MD = 8 dB(A) for MFAM stimuli. This difference results from the different frequency spectra for RFAM and MFAM. Indeed, the MFAM stimuli has significantly more energy in region 600-1kHz which is near the peak of the A-weighting filter: see Section 17.3. The ratings for the two types of GN were very similar for this set of stimuli.

### 8.1.3 Stimuli set III: Fixed peak level

In a third set of stimuli the peak level was kept constant and MD was increased by decreasing the masking noise level. Data are summarised in Table 8.3. In Figure 8.4, the mean annoyance score is shown as a function of MD for the two types of stimulus (MFAM and RFAM) and two types of GN.

| | | MD, dB(A) | $L_{Aeq}$, dB(A) | Rating | 95% CI |
|---|---|---|---|---|---|
| GN1 | MFAM | GN1 | 38 | 3.6 | 1.4 |
| | | 5 | 37 | 5.3 | 0.8 |
| | | 8 | 38 | 5.6 | 0.9 |
| | | 12 | 37 | 6.5 | 0.7 |
| | | 15 | 38 | 6.7 | 1.1 |
| | RFAM | GN1 | 38 | 3.6 | 1.4 |
| | | 1 | 37 | 4.2 | 0.9 |
| | | 2 | 34 | 4.9 | 0.7 |
| | | 4 | 33 | 5.1 | 0.9 |
| | | 6 | 31 | 5.2 | 0.7 |
| GN2 | MFAM | GN2 | 40 | 4.7 | 2.0 |
| | | 5 | 40 | 5.6 | 1.0 |
| | | 8 | 38 | 6.5 | 0.5 |
| | | 11 | 38 | 6.3 | 0.6 |
| | | 14 | 37 | 6.9 | 0.7 |
| | RFAM | GN2 | 40 | 4.7 | 2.0 |
| | | 1 | 38 | 4.6 | 0.9 |
| | | 2 | 36 | 5.0 | 0.8 |
| | | 3 | 34 | 5.9 | 0.6 |
| | | 5 | 32 | 6.0 | 0.7 |

**Table 8.3 Stimuli data for which the Modulation Depth was increased by decreasing the masking noise. MPL was kept constant.** $L_{Aeq}$ **95% CI = ±1 dB(A)**

Again the average annoyance for the 4 groups of AM stimuli consistently increased with $L_{Aeq}$. 95% CI are similar to the previous two graphs. In this graph the error bars overlap strongly and therefore the difference between RFAM and MFAM or GN1 and GN 2 is less clear than in the previous two graphs. For both RFAM and MFAM GN2 stimuli the increase levels off towards higher MD values. These were the stimuli with consistently decreasing $L_{Aeq}$ values. For the MFAM GN1 stimulus with its almost constant $L_{Aeq}$ the average annoyance ratings do increase steadily with increasing MD.

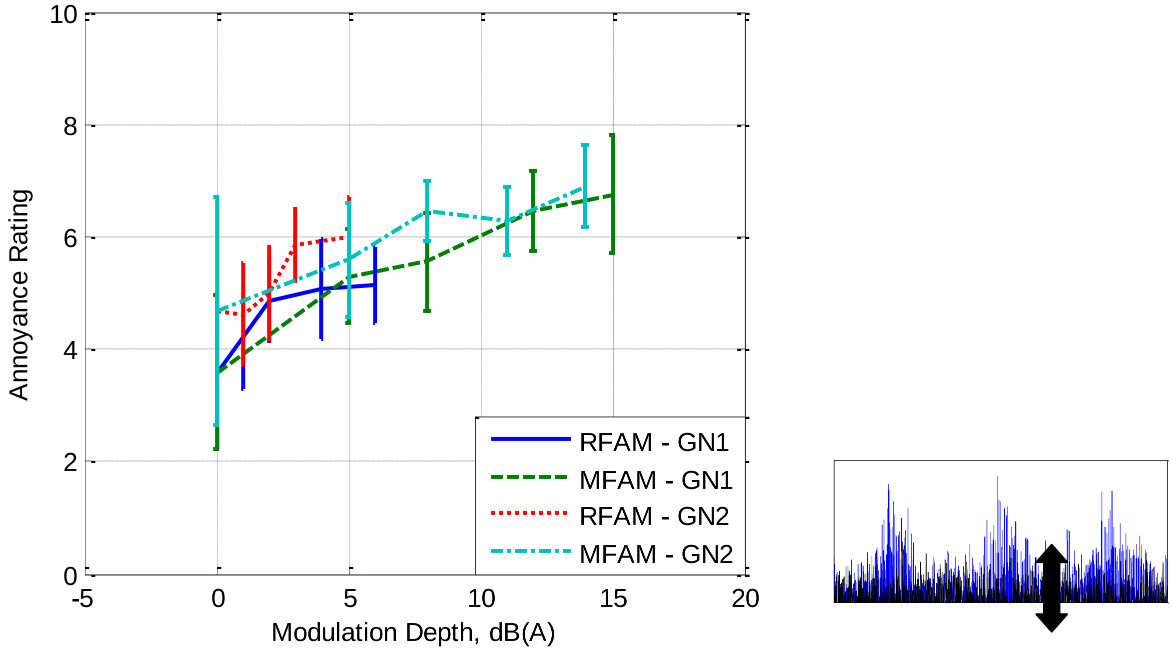



**Figure 8.4: Annoyance rating for fixed peak modulation amplitude and increasing MD by means of decreasing background noise (as indicated by inset) for RFAM and MFAM compared.**

### 8.1.4 Comparison of sets I - III

Comparing the three stimuli sets the responses to the two types of stimuli, MFAM and RFAM, were very different. Some reasons for this have been pointed out but the relative effects on annoyance of $L_{Aeq}$ for the first and third stimuli sets and of MD for the second set remained unclear. Further testing was therefore required under more controlled test conditions to clarify the effect of modulation type on participant response.

The situation is similar for the role of GN noise. Figure 8.2 also shows that there was a systematic difference in the annoyance ratings of the stimuli with two difference garden noise types. While the stimuli with GN2 were on average about 1-2 dB(A) higher than the stimuli with GN1 (Table 8.1 Stimuli data for which modulation peak level (MPL) was varied and the masking noise kept constant. **$L_{Aeq}$** 95% CI = ±1 dB(A).Table 8.1) this might not fully explain the increased annoyance ratings. Also the responses did not differ so clearly between the GN types in Figure 8.3 and Figure 8.4. It is therefore possible but not certain that the masking noise which was played at relatively low level (amplitude ratio between WTN masking and GN masking 5:1) could contribute significantly to the perception and the affective response to WTN. This would be expected as it is widely reported in the literature (e.g. [10] and [1]). Because GN varies widely both spatially and temporally it is however impossible to design a representative GN stimulus.

### 8.1.5 The role of $L_{Aeq}$ and MD compared

From the results in Figure 8.2 - Figure 8.4 it seems clear that major contributors to annoyance are the parameters $L_{Aeq}$ and MD. Therefore the average annoyance ratings for all stimuli have been plotted versus these parameters in Figure 8.5 a) and b) to get an impression about the strength of the correlation. When fitting a linear function through the data the slope of the line shows that the ratings increased on average faster within a typical range of $L_{Aeq}$ than within a typical range of MD values. The spread of the data around the fit was comparable for the two plots.

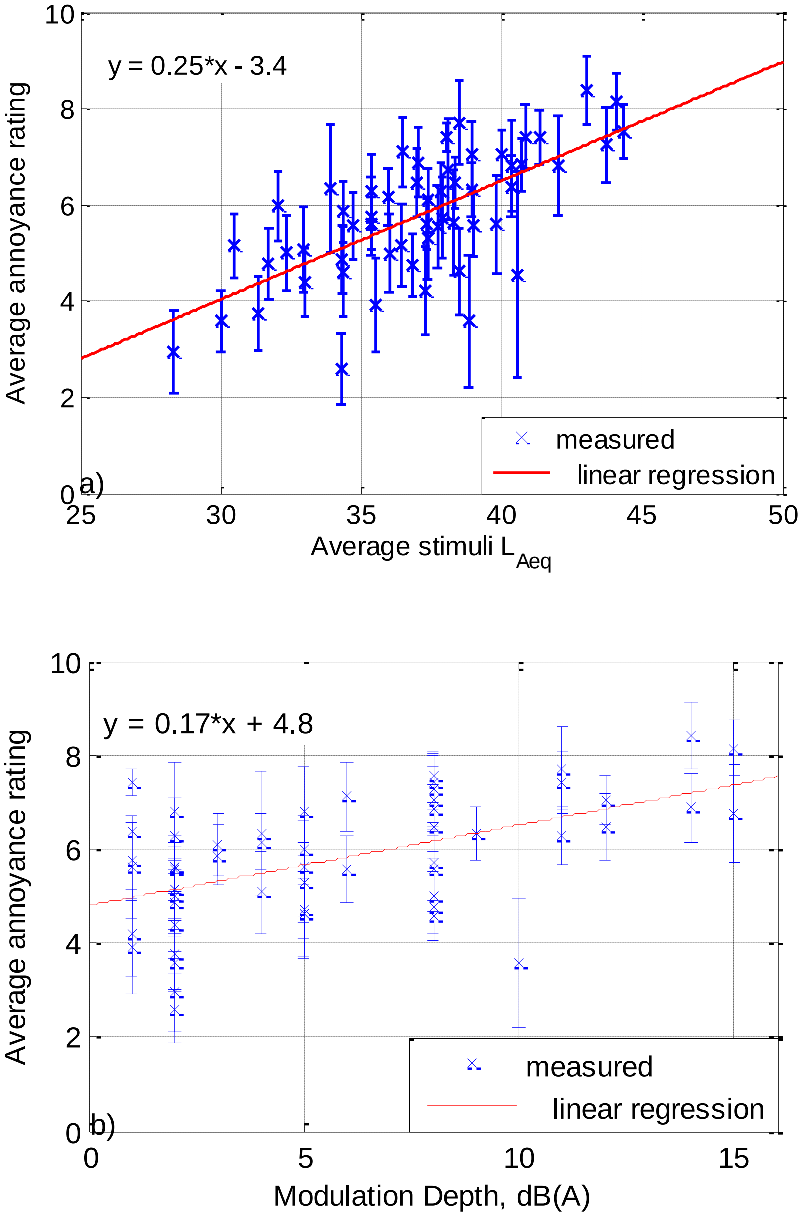



**Figure 8.5 Average annoyance ratings for all stimuli vs a) the average stimuli $L_{Aeq}$ and b) the Modulation Depth.**

### 8.1.6 Stimuli set IV: spectral characteristics

In a fourth set of stimuli the frequency content namely the bandwidth and the frequency skew were changed for both AM types as shown in Table 8.4. Figure 8.6 shows the average annoyance rating as a function of pulse bandwidth. The variation in the average ratings are as small as 0.6 points on the rating scale for each of the AM types. It is therefore concluded that there is no significant dependence for annoyance on the bandwidth/frequency skew of the modulation. MFAM ratings are higher than ratings of RFAM stimuli. This can be attributed to both the higher average $L_{Aeq}$ and the higher MD value of the MFAM stimulus.

| | BW, Hz | Skew, % | MD, dB(A) | $L_{Aeq}$, dB(A) | Rating | 95% CI |
|---|---|---|---|---|---|---|
| **RFAM** | 300 | 67 | 2 | 35 | 5.6 | 0.5 |
| | 400 | 70 | 2 | 34 | 4.9 | 0.7 |
| | 500 | 80 | 2 | 36 | 5.2 | 0.9 |
| **MFAM** | 250 | 30 | 10 | 39 | 6.3 | 0.6 |
| | 350 | 50 | 8 | 38 | 5.6 | 0.9 |
| | 450 | 61 | 8 | 38 | 5.7 | 0.8 |

**Table 8.4 Stimuli data for which bandwidth was varied.** $L_{Aeq}$ **95% CI = ±1 dB(A)**

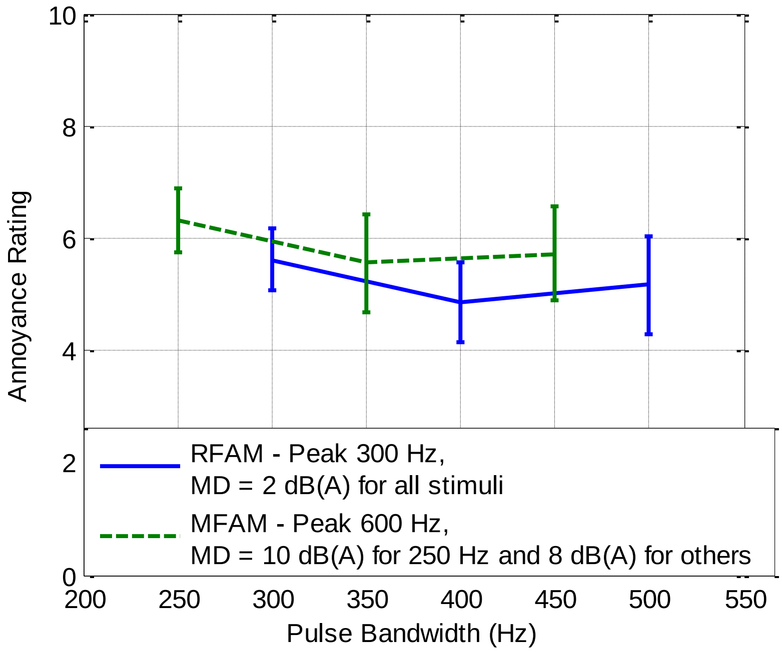



**Figure 8.6: Annoyance rating as a function of pulse bandwidth and frequency skew for RFAM and MFAM compared.**

### 8.1.7 Stimuli set V: modulation frequency

The last set of stimuli explore the effect of modulation frequency on annoyance ratings using two modulation periods of 0.65 sec and 1.3 sec. Test data are summarised in Table 8.5 Figure 8.7 shows the average annoyance rating as a function of MD for both RFAM and MFAM stimuli. The stimuli with shorter modulation period are rated consistently higher than the ones with a 1.3 sec period. In fact the contour lines are very nearly parallel. It is not surprising that the high frequency modulation which according to [9] increases fluctuation strength in this frequency range is therefore more annoying than the lower modulation frequency. The peak in fluctuation strength would be expected at a modulation period of 0.25 sec. The variation in the average ratings ranges from 0.7 to 1.1 and is therefore small for all stimuli. Because MD was varied by dropping the MN level the effects of MD and $L_{Aeq}$ might partly cancel each other out as seemed to be the case in Figure 8.4.

| Period, sec | | MD, dB(A) | $L_{Aeq}$, dB(A) | Rating | 95% CI |
|---|---|---|---|---|---|
| 0.65 | RFAM | 1 | 38 | 6.4 | 1.6 |
| | | 2 | 35 | 7.0 | 1.3 |
| | | 4 | 34 | 7.1 | 1.9 |
| | MFAM | 5 | 40 | 7.6 | 1.1 |
| | | 9 | 39 | 7.8 | 1.1 |
| | | 11 | 38 | 8.5 | 1.2 |
| 1.3 | RFAM | 1 | 37 | 5.0 | 1.2 |
| | | 2 | 34 | 5.6 | 1.2 |
| | | 4 | 33 | 5.8 | 1.1 |
| | MFAM | 5 | 37 | 6.1 | 1.4 |
| | | 8 | 38 | 6.3 | 1.0 |
| | | 12 | 37 | 7.2 | 1.1 |

**Table 8.5 Stimuli data for two modulation frequencies (here expressed as modulation period).** $L_{Aeq}$ **95% CI = ±1**

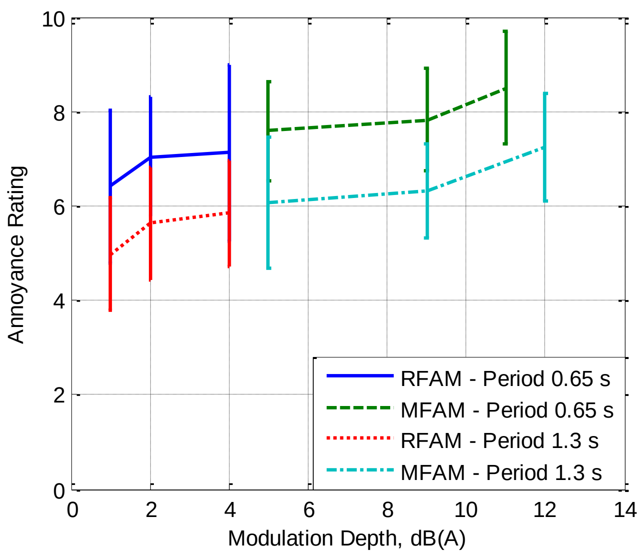



**Figure 8.7: Annoyance rating as a function of MD for two modulation frequencies and for RFAM and MFAM compared.**

Figure 8.7 shows a distinct difference between the higher and lower modulation frequency but not a strong dependence on MD for the two different types of modulation. Within the model from [9] the fluctuation increases as the modulation frequency get closer to 4 Hz. Therefore it is not surprising that the annoyance is significantly higher for shorter modulation periods, *i.e.* higher modulation frequencies.

## *8.2 Conclusions*

Although it seemed possible that in principle signal properties Modulation Depth, Modulation Peak Level, and Masking Level could be separated in their effect on annoyance, the main effects on annoyance were seen from varying the stimuli parameters $L_{Aeq}$ and MD. Therefore a more straightforward method seemed to be to fix the $L_{Aeq}$ for each set of stimuli. This method would also allow a direct comparison with previous studies. The exact bandwidth of the amplitude modulation seemed to have little effect on annoyance and it was therefore decided that these parameters would not be included in the final tests. Because the evidence on garden noise and AM type was not clear it was decided to initially focus on RFAM stimuli without GN but to include validation stimuli to part of the test for one type of garden noise and for MFAM. A systematic analysis of the role of modulation frequency would have been possible but because it is probably sufficiently described by the existing Fastl and Zwicker's metrics for fluctuation strength it was decided to focus on modulation at a fixed frequency of less than 1Hz, representative of typical of large modern wind turbines.

For the final tests it was therefore decided to use RFAM stimuli at 3 $L_{Aeq}$ levels of 30, 35, 40 dB(A) and 12 different MD values (1, 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, 7, 9, 12 dB(A)). The validation studies included also a subset of MFAM stimuli and one with garden noise masking.

# 9. Results of Final Test

The final tests were conducted under tightly controlled conditions in a calibrated listening room to optimise the accuracy of modulation parameters and realistic reproduction of outdoor stimuli. During the test procedure two rating methods were used. The first was an 11 point annoyance rating scale for easy comparison with previous studies. The second was a comparative rating method modelled on a procedure used in [5]: An un-modulated wind turbine sound (ABBS) was adjusted in $L_{Aeq}$ to be as annoying as a modulated AM test stimulus that was based on the same spectrum as the ABBS. The procedure is described in more detail in Sections 6.5.2, 6.6.2 and 18.3. This adjustment procedure was used to directly answer the question how much the $L_{Aeq}$ of the ABBS needs to increase to be equally annoying to a quieter AM stimulus. This type of information has previously been inferred from absolute annoyance ratings and has led to widely different results.

To keep the tests as simple as possible and within a manageable time for the participants the results focussed initially (Sections 9.1 - 9.3) on RFAM stimuli with WT masking noise and without the addition of garden noise (parameters in bold in Table 6.3). The results of Validation Test I are presented in Section 9.4 for comparison to show whether significantly different responses are expected from MFAM stimuli or for the addition of low levels of garden noise. Section 9.5 addresses the question whether the perception of stimuli is significantly different when the masking noise level is kept constant in comparison to the method of the Final Test where MN levels are dropped with increasing MD to achieve constant $L_{Aeq}$. The stimuli generation and detailed choice of parameters are described in Table 6.3 Section 6 and Appendix V.

## *9.1 Rating distributions*

Figure 9.1 shows an overview over the general rating behaviour of all participants and stimuli that were included in the Final Test. Part a) shows the rating distribution for the annoyance ratings on the 11 point rating scale. A polarised rating behaviour was observed where the ratings 0 (not at all annoyed) dominated and the number of ratings towards higher annoyance values generally decreased. This is expected as many stimuli were fairly quiet and the low modulation depths were included.

Part b) contains the $L_{Aeq}$ adjustments of the ABBS minus the $L_{Aeq}$ of the stimuli. In contrast Figure b) shows a near Gaussian distribution around the mean value of 2.3 dB(A). This is evidence that ABBS $L_{Aeq}$ were on average adjusted to higher values than the AM $L_{Aeq}$ and therefore AM stimuli (RFAM) were generally rated as being more annoying than the ABBS. Interestingly, there are negative values of up to -11 dB(A) among those ratings which might be an indication that participants found the adjustment task difficult.

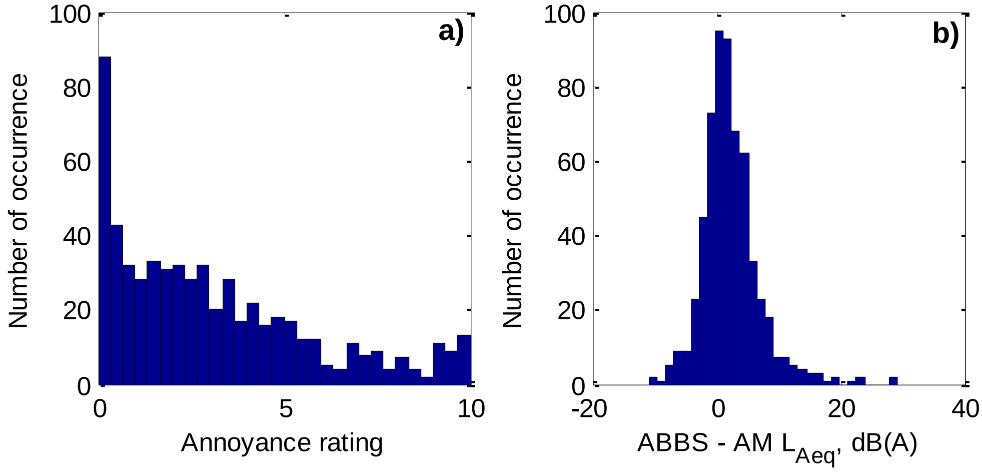



**Figure 9.1 Distribution of ratings a) for absolute annoyance scale, b) for the difference in $L_{Aeq}$ between the unmodulated ABBS and AM stimuli.**

The rating behaviour for the validation tests is plotted in Figure 20.1 and Figure 21.1. The distributions are very similar which is a first indicator that validation test stimuli were perceived to be similarly annoying to stimuli in the final test.

## *9.2 Annoyance ratings grouped by lines of constant MD*

Using the 11 point scale for absolute annoyance ratings in the final listening tests allowed plotting the annoyance ratings as a function of $L_{Aeq}$ in groups of MD (Figure 9.2, tabulated results in Table 23.1). This presentation is in a similar form to published results [15]. Figure 9.2 shows that mean annoyance ratings consistently increase with the $L_{Aeq}$ of the AM stimuli (RFAM). Un-modulated stimuli were clearly rated as less annoying than modulated stimuli. A systematic increase with modulation depth is also apparent although some of the ratings overlap especially at higher $L_{Aeq}$ values. This can be explained by results from [12] who found that it was difficult for listeners to correctly identify the change in modulation depth in a signal. Therefore when perception of these changes is difficult it is not surprising that annoyance ratings show similar inconsistency. The error bars denote 95% confidence intervals (CI). They are smaller than one point on the annoyance rating scale for low $L_{Aeq}$

and low MD. At high $L_{Aeq}$ and MD the error bars span up to 2.2 points (25%) of the rating scale which is a large value but not unexpected for an attitudinal parameter like annoyance. The statistical significance of the result in the presence of large error bars is further discussed in the context of Figure 9.3.

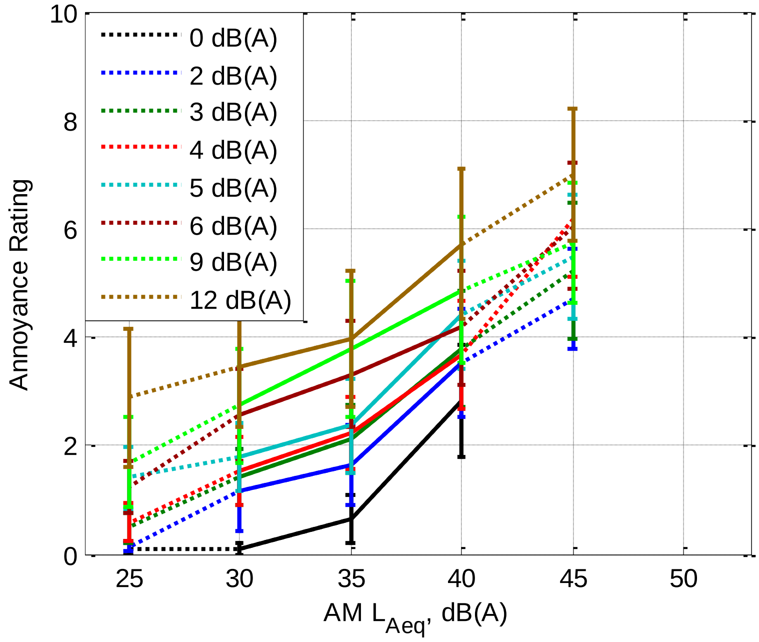



**Figure 9.2 Mean annoyance rating of AM test stimuli as a function of modulation depth. Solid lines are results from final test, dotted lines from the validation tests with reduced participant numbers. The legend specifies MD. See data in Table 22.1.**

A similar study [15] showed results with similar general features like the strong increase in annoyance ratings with $L_{Aeq}$ and less pronounced and sometimes overlapping ratings with increasing modulation factor.

Because Lee et al. used a different metric for modulation depth their results cannot be directly compared. The authors did not compare ratings to un-modulated signals. Lee et al. used the standard 11 point scale according to [24]. They found minimum annoyance ratings of 1.5/2.5 and maximum annoyance ratings around 7/8 for two different tests, respectively. The minimum values in the current study are lower because stimuli with lower $L_{Aeq}$ values were included. The maximum values are similar to results in Figure 9.2 which is surprising because Lee et al. included $L_{Aeq}$ values of up to 55 dB(A) in comparison to the 45 dB(A) used in the current study. This can possibly be explained by the descriptors used for the maximum annoyance rating was "very annoyed" in the current study and "extremely annoyed" for the study by Lee *et al.*

## *9.3 Ratings as a function of modulation depth*

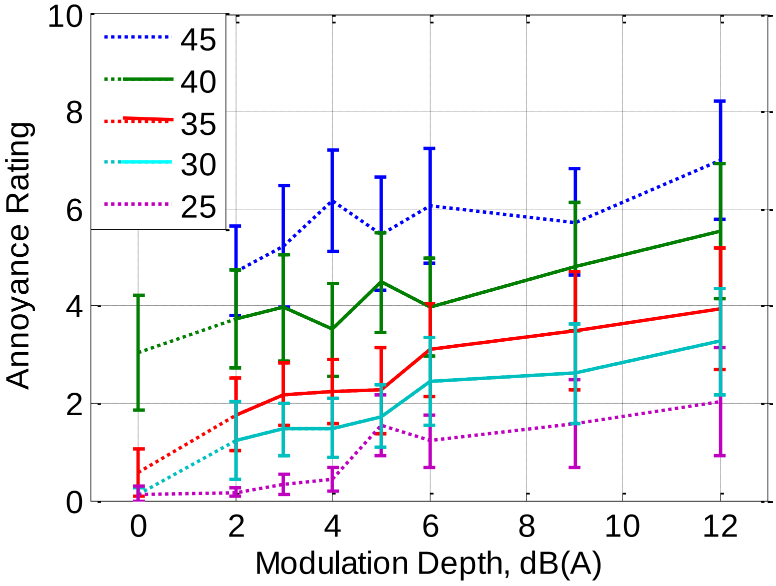



**Figure 9.3 Absolute annoyance ratings of AM stimuli as a function of modulation depth. Solid lines are results from final test, dotted lines from the validation tests with reduced participant numbers. The legend specifies the $L_{Aeq}$ of the test stimuli in dB(A). See data in Table 22.1.**

Figure 9.3 shows the mean annoyance ratings as a function of modulation depth with isolines of $L_{Aeq}$ to bring out any MD related trends more clearly. Like in Figure 9.2 it can be seen that $L_{Aeq}$ levels clearly change the average annoyance ratings and at the lowest $L_{Aeq}$ the stimuli are the least annoying. In comparison modulation depth increased the mean ratings only slightly which given the large error bars is statistically insignificant. A clear onset of annoyance with modulation depth is not apparent from the Figure. 95% CI are large as expected for an attitudinal parameter like annoyance.

To assess the significance of increased annoyance ratings, statistical analysis using a GLM ANOVA (e.g. [25] or [26]) has been performed using SPSS™. The results suggest that $L_{Aeq}$ increases annoyance significantly whereas the modulation depth does not with current numbers of participants. Given the consistent increase of annoyance ratings with MD it seems likely that a small but significant effect would be found with a larger number of participants. Beyond 6 dB(A) the curves seem to flatten off for $L_{Aeq}$ of 25, 30 and 35 dB(A). A similar decrease albeit with much less data has been observed [16] for low frequency AM broadband noise. [16] then go on to interpret results from absolute annoyance ratings to find equivalent levels of unmodulated sound. A more direct approach was taken in this study by posing the question:

"To which $L_{Aeq}$ would a typical broadband WTN have to be adjusted to be as annoying as an AM stimulus of a certain $L_{Aeq}$ and MD?" Therefore the participants were asked to adjust in volume an Adaptive BroadBand Stimulus (ABBS) until it was as annoying as the modulated sound. The ABBS was identical to the AM stimulus at 0 dB(A) modulation depth.

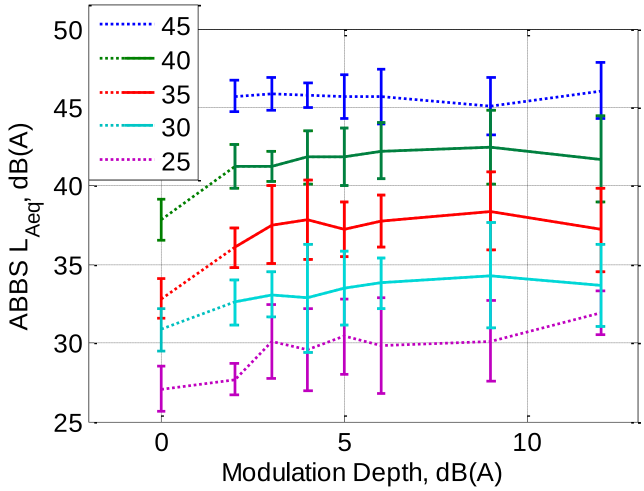



**Figure 9.4 ABBS level adjustments in comparison to AM stimuli as a function of modulation depth – Solid lines are results from final test, dotted lines from the validation tests with reduced participant numbers. The legend specifies the $L_{Aeq}$ of the test stimuli in dB(A).**

In that task the ABBS were adjusted to levels close to and generally slightly above the $L_{Aeq}$ of the AM stimulus (Figure 9.4, Table 23.2) indicating that $L_{Aeq}$ increases annoyance in accordance with Figure 9.2 and Figure 9.3. However, the ratings for increasing MD show a slightly different rating behaviour as the line of equal ABBS $L_{Aeq}$ appear rather flat. The adjusted levels therefore did not increase with statistical significance between 2 and 12 dB(A) modulation depths. They also increased only slightly between 0 and 2 dB(A) modulation depth. Like in Figure 9.3 a clear onset of annoyance with modulation depth is therefore not apparent.

It would be easy to think that the results from the two rating procedures plotted in Figure 9.3 (small but consistent increase of annoyance rating with MD) and Figure 9.4 (no increase of adjusted levels with MD for most $L_{Aeq}$) are contradictory. However, it is worth pointing out that the tasks were very different in that one rated annoyance directly and the other an equivalent level of an unmodulated WTN. And the results in both, Sensitivity Test I and the Final Test have shown that annoyance increases more strongly with $L_{Aeq}$ than with MD. So while annoyance might be consistently but slightly increasing with MD for an AM stimulus, an unmodulated sound at an ABBS $L_{Aeq}$ as shown in Figure 9.4 might be sufficient to account for the increased annoyance from AM.

The adjustments for quieter stimuli tended to be larger than for louder stimuli as seen in Figure 9.5 (Table 23.3). That figure shows the adjustments as the difference between the ABBS $L_{Aeq}$ and the $L_{Aeq}$ of the AM stimulus. On average the adjustments are about 2.3 dB(A) and mean maximum adjustments exceed 5 dB(A) only for the quietest stimuli. With a range from 0.8 – 3.2 dB(A) the 95% CI are the same size of as in Figure 9.4 and only appear to be larger because of the different scale.

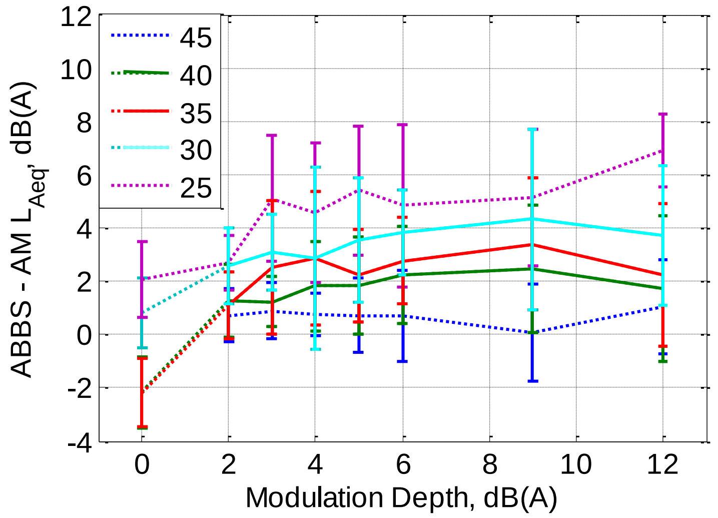



**Figure 9.5 Normalised relative annoyance ratings of AM stimuli as a function of modulation depth. Solid lines are results from final test, dotted lines from the validation tests with reduced participant numbers.**

At 0 dB(A) modulation depth the adjustments were expected to be around $L_{Aeq}$ = 0 dB(A) too as the two stimuli were identical. This is however not the case. Quieter stimuli were adjusted to higher levels such as a 2 dB(A) higher adjustment than the 25 dB(A) AM stimulus. And for the louder stimuli the level adjustment was on average 2 dB(A) lower than the AM stimulus $L_{Aeq}$. This confirms a participant observation from Section 19.1 stating that levels were hard to judge. Also participants might have felt the need to always make adjustments in the belief that the AM stimuli must be different from the ABBS.

## *9.4 Validation test I*

In Validation Test I the questions were addressed whether different rating behaviour was expected from stimuli that were also masked by garden noise and from MFAM stimuli. The results of rating distributions, absolute ratings and ABBS ratings are displayed in the Appendix, Section 20.

Figure 20.2 shows absolute annoyance ratings for stimuli without garden noise (top panels) and with garden noise (bottom panels) and for RFAM type stimuli (left panels) and MFAM type stimulus (right panels). For all four combinations the annoyance ratings were very similar. $L_{Aeq}$ levels clearly changed the average annoyance ratings whereas modulation depth increased the ratings only slightly. The increase of annoyance with modulation depth seemed steeper for the lowest $L_{Aeq}$ than for the highest when garden noise was added. It should be pointed out that GN was reproduced at $L_{Aeq}$ = 7.5 dB(A) below the $L_{Aeq}$ of the AM test stimulus. It was therefore loud enough to change the character of the stimulus but too quiet to make the AM noise less audible. In reality there are often situations with high levels of GN that mask the AM stimulus partially or completely.

In agreement with the results from in previous sections a clear onset of annoyance with modulation depth is not apparent and the size of the 95% confidence intervals is comparable to RFAM stimuli without GN. Because of the similarity of the results the types of AM and the presence/absence of garden noise were not thought to produce different responses of statistical significance and the possible effects seen in Sensitivity Test II were most likely the result of poor control of the $L_{Aeq}$ and modulation parameters in the headphone reproduction.

## *9.5 Validation Test II*

One concern with stimuli of constant $L_{Aeq}$ was that to keep the level constant with increasing MD the masking noise had to be reduced which might have led to changes in signal character. In Validation Test II (Section 21) a subgroup of nine participants therefore also rated a set of stimuli with the same MD levels and masking levels of 25, 30, 35 and 40 dB(A). They used the procedure of adjusting ABBS for this test because an equivalent result to Figure 9.5 could be produced using the total measured $L_{Aeq}$ of the stimuli to normalise the ABBS values. If these relative annoyance ratings changed significantly then this would be evidence of a perceptible change in signal character.

In general the rating behaviour of the participants shown in Figure 21.2 was very similar for the two sets of stimuli. Figure 21.2 a) and c) show the absolute ABBS $L_{Aeq}$ for the constant stimulus $L_{Aeq}$ and the constant masking noise $L_{Aeq}$ stimuli, respectively. The monotonous increase of the rating with modulation depth is evidence that the participants adjusted the louder stimuli to slightly higher ABBS $L_{Aeq}$ compared to a).

Figure 21.2 b) and d) show the difference between ABBS and RFAM $L_{Aeq}$. Importantly, the increase of $L_{Aeq}$ due to the increasing MD was measured and the correct total stimulus $L_{Aeq}$ was used to calculate this difference. Therefore if the nature of the stimuli changed by reducing masking noise then a difference between b) and d) should be seen. However, within the 95 % confidence intervals b) and d) are very similar which suggests that by the nature of the stimuli is not changed significantly in terms of annoyance when reducing the masking noise to retain the total $L_{Aeq}$.

## *9.6 $L_{A90}$ analysis*

All sound levels have so far been expressed as $L_{Aeq}$ values whereas another common measure is $L_{A90}$. In Section 22, Figure 9.2 and Figure 9.5 have been compared to the respective $L_{A90}$ equivalents. The increase in annoyance ratings with increasing MD is clearly visible in Figure 9.2 and in Figure 22.1 b). The difference between the two measures only becomes significant at MD ≥ 9 dB(A) when $L_{A90}$ suggests that the contribution of AM to annoyance might be larger than suggested by the $L_{Aeq}$ measure. This is because $L_{A90}$ is lower by up to 7 dB(A) at MD = 12 dB(A). It should be noted though that MD has rarely been observed to exceed 10 dB(A) (see WPC).

Figure 22.2 shows normalised ABBS levels a) as $L_{Aeq}$ and b) as $L_{A90}$. Note that the legend identifies the "design" $L_{Aeq}$ for each group of stimuli for both Figure 22.2 a) and b). When measured in $L_{A90}$ the normalised ABBS increase almost linearly with MD although the higher levels suggest a flattening of the curve from an MD of about 3dB(A). In summary $L_{A90}$ might be a suitable parameter to express annoyance ratings in the psychoacoustic context and should be investigated more closely in future studies.

## *9.7 Interpretation of results*

In its adjustment procedure the current study did not follow the common approach to quote the percentage of annoyed listeners. While this is often a very useful method when the occurrence of this type of noise is common enough to enable a large scale survey study in the affected neighbourhoods it is difficult to use in this context where the observations of AM are still too rare for such a study. The annoyance ratings derived from the laboratory study cannot be directly interpreted in these terms because the ratings are taken out of context so that absolute ratings can only be interpreted relative to each other.

## *9.8 Comparison of different metrics for modulation depths and fluctuation strength*

The MD metric, based on review of short-term $L_{Aeq}$ levels, was principally used to design a representative set of test stimuli, but it is not necessarily one that particularly relates to subjective response; however, WPB1 pointed out the complexities and pitfalls with these types of metrics, both given the uncertainties in reading the peaks and troughs values and in their application to general, realistic signals. The procedure described in Section 17.4 uses an averaging process which will work reasonably well given that the stimuli are artificially generated and are consistent in time, but inevitably some uncertainty will remain.

The results above were therefore analysed using other, more robust and generalised metrics, in addition to the MD as evaluated from $L_{Aeq100ms}$ signals:

- Method 1 described in WPB1: Fourier transform of signal envelope, with/without signal low-pass filtering below 500 Hz, and noise-referenced (normalised). The second method described in WPB1 gave similar results and is not considered further.
- The main metric analysis method described in WPBF, which is similar to method 1 of WPB1 but with alternative normalisation factors (peak modulation in modulation spectrum)
- The psychoacoustics Fluctuation Strength Metric, based on [9], as evaluated in an implementation in the 01dB dBSonic[12] software, either the calculated average or peak level of the calculated Fluctuation Strength over the 20 s stimuli recording.

The comparative values are detailed in Table 9.1 and Table 9.2.

[12] http://www.01db-metravib.com/nvh-instruments.477/dbsonic-psychoacoustic-software.558/?L=1

| MD ($L_{Aeq}$) | Main routine (WPBF) | WPB1 (full band, noise ref.) | WPB1, (0-500Hz, noise ref.) |
|---|---|---|---|
| 1 | 0.6 | 7.3 | 4.8 |
| 2 | 1.0 | 10.8 | 9.2 |
| 2.5 | 1.4 | 15.3 | 13.1 |
| 3 | 2.0 | 19.1 | 17.1 |
| 3.5 | 2.3 | 20.3 | 18.0 |
| 4 | 3.1 | 22.8 | 20.5 |
| 4.5 | 3.3 | 23.5 | 21.5 |
| 5 | 3.8 | 25.2 | 23.0 |
| 6 | 4.8 | 27.4 | 25.1 |
| 7 | 5.7 | 29.9 | 27.5 |
| 9 | 7.8 | 33.8 | 31.4 |
| 12 | 11.2 | 39.2 | 36.5 |

**Table 9.1 Comparison of modulation magnitude values (dB) resulting from the metrics based on physical signal properties.**

| MD(dB) | Fluctuation strength (cVacil) 40 dB(A) | | Fluctuation strength (cVacil) 30 dB(A) | |
|---|---|---|---|---|
| LAeq,100ms | max | Mean | max | Mean |
| 1 | 1.4 | 1.3 | 0.9 | 0.7 |
| 2 | 1.6 | 1.3 | 1.4 | 0.9 |
| 2.5 | 1.7 | 1.4 | 1.6 | 0.9 |
| 3 | 2.5 | 1.6 | 1.2 | 0.9 |
| 3.5 | 2.5 | 1.9 | 1.2 | 0.9 |
| 4 | 2.7 | 2 | 1.3 | 1 |
| 4.5 | 2.8 | 2.2 | 1.4 | 1.1 |
| 5 | 3.2 | 2.5 | 2 | 1.1 |
| 6 | 4.1 | 3 | 2.4 | 1.8 |
| 7 | *n/a* | 4.7 | 3.2 | 2.1 |
| 9 | 7.6 | 5.6 | 3.4 | 2.4 |
| 12 | 10.8 | 8.6 | 4.9 | 2.9 |

**Table 9.2 Comparison of fluctuation strength values.**

Figure 9.6 to Figure 9.8 show annoyance ratings for the sets of stimuli at $L_{Aeq}$ = 30 and 40 dB(A) for the parameter modulation depth in the different metrics. The main effect is that different metrics basically move and compress the annoyance curve along the x-axis. While Figure 9.6 a) and b) have a similar range of 2 – 12 dB(A) and 1 – 11 dB respectively, the metrics from WPB1 extend up to modulation depths of 32 -39 dB and start at modulation depths between 4 and 11 dB (Figure 9.7). Other metrics from WPB1 gave very similar results and are therefore not shown.

If the perception based measure fluctuation strength which includes loudness rather than $L_{Aeq}$ was able to explain the annoyance ratings completely the two curves in Figure 9.8 should merge. The implementation of different loudness standards can also lead to uncertainties as shown in [27]. Merging curves are also not expected because affective participant response is also influenced by contextual and attitudinal factors. From all metrics related figures results the important conclusion that the consideration of any potential threshold or correction must be consistent with the metric from which the data was analysed.

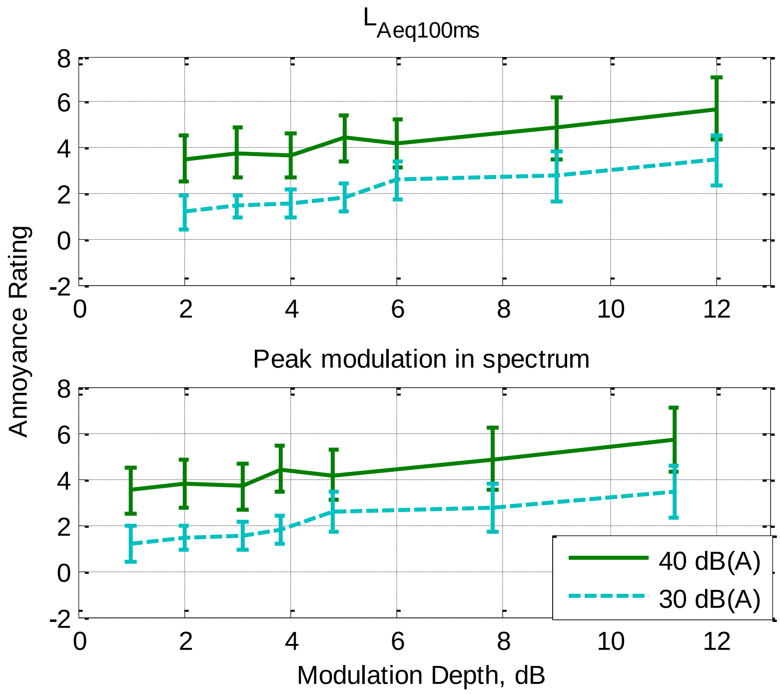



**Figure 9.6 Comparison of MD ($L_{Aeq}$) and main WPF routine metrics on absolute annoyance.**

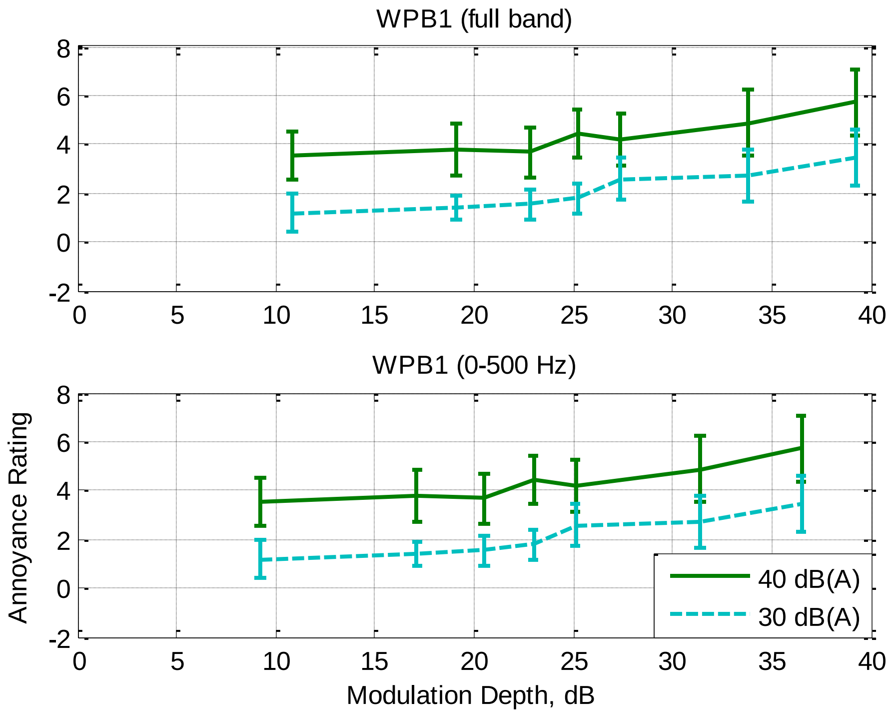



**Figure 9.7 Comparison of WPB1 metric (noise-referenced) on absolute annoyance.**

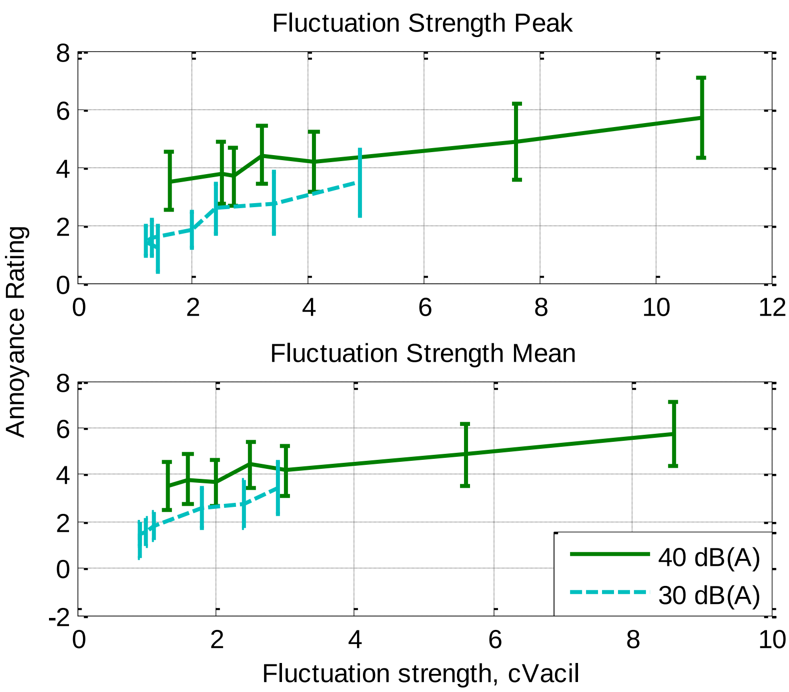



**Figure 9.8 Comparison of psychoacoustic metrics on absolute annoyance.**

# 10. Conclusions

Listening tests have been conducted to evaluate AM metrics in terms of their correlation with subjective listener response and to find a dose-response relation between the AM metric and the subjective annoyance response.

## *10.1 Sensitivity test results*

In a pilot phase test stimuli were synthesised for a characteristic range of AM modulation parameters and outdoor listening scenarios. The sensitivity tests showed in accordance with previous literature that annoyance crucially depended on $L_{Aeq}$ and to a lesser extent on MD.

The use of A-weighting both for the level and the modulation depth lead to consistent results. In contrast plotting the annoyance ratings as a function of $L_{A90}$ as an alternative to $L_{Aeq}$ produced similar annoyance rating results up to MD = 6 dB(A) where $L_{A90}$ and $L_{Aeq}$ values are similar. Whether normalised ABBS as a function $L_{A90}$ gives a consistently linear dose response relation should be subject of future work.

Annoyance response did not change significantly for the temporal parameters pulse shape and pulse width at constant modulation frequency as well as the spectral parameters frequency skew and bandwidth of the modulation pulse. These parameters were subsequently fixed at realistic values.

Annoyance ratings did vary with modulation frequency in agreement with changes in fluctuation strengths predicted by the model in [9]. For this reason and because the modulation frequency between large modern wind turbines does not change a lot modulation frequency was also eliminated as a parameter.

In the sensitivity tests conflicting results arose from the use of the two modulation types MFAM and RFAM as well as the role of a low level of garden noise in the broadband MN. This was thought to be mainly due to the insufficiently controlled $L_{Aeq}$ in this set of tests. For this reason, the decision to let the participants choose a sound reproduction level that they regarded as realistic turned out not to be successful. Another factor was possibly the quality of the sound reproduction using headphones.

## *10.2 Final Test and Validation Tests*

The final test which was conducted in the carefully controlled acoustic environment of the listening room there were therefore 3 sets of test sounds with the constant $L_{Aeq}$ of 30, 35, and 40 dB(A) for which the modulation depth was systematically varied from 0 to 12 dB(A) in increasing steps. After taking the effect of $L_{Aeq}$ which always dominated the annoyance rating into account, increases in modulation depth seemed to increase the annoyance rating slightly and consistently (monotonically), in agreement with previous research. However, the effect was not statistically significant because there was a large spread of ratings. The consistency of the increase for all $L_{Aeq}$ suggests that given a large enough group of participants it can possibly be shown that average annoyance rating increase slightly but consistently (monotonically) with modulation depth. The 95% CI are however expected to remain large because affective response varies between listeners. In contrast average ABBS $L_{Aeq}$ was constant from MD of about 3 dB(A). This answered the question how much louder would an equivalent unmodulated sound have to be to be equally annoying to a modulated sound. The adjustments were on average 1.7 dB(A) for a 40 dB(A) test sound and 3.5 dB(A) at 30 dB(A). Validation tests at two additional levels of 45 dB(A) and 25 dB(A) confirmed this trend. A clear onset of annoyance at a particular modulation depth could not be found for either of the two rating methods. The average ABBS adjustment for all stimuli was found to be 2.3 dB(A) higher than the test stimulus $L_{Aeq}$. When levels were measured as $L_{A90}$, results suggest that annoyance ratings were similar for MD of up to 6 dB(A) and generally increased with both, MD and $L_{A90}$. Because results for sets of stimuli with constant $L_{A90}$ and changing MD are not available simple average adjustments cannot be identified and further work would be necessary.

The comparison of the RFAM results without GN with MFAM ratings and the addition of GN for both modulation types in Validation Test I with a subgroup of 11 participants both the absolute annoyance ratings and ABBS $L_{Aeq}$ did not show significant different annoyance ratings for the different sets of stimuli. This suggests that the results of the Final Test can be generalised to a wider range of AM sounds than just the two chosen examples.

A similar result for Validation Test II when the masking noise (MN) was kept constant to avoid a possible change in stimulus character implies that the same is probably true for GN/MN as long as the level is low enough not to affect AM audibility.

In a last step the annoyance ratings were compared for 6 different metrics, four of them based on different physical definitions of modulation depth and 2 using the perceptive measure fluctuation strength. The comparison showed that the main effect of the physical metric is to change the range of modulation depths. The same stimuli would have a range of

0 – 12 dB(A) in the preliminary MD metric but 4 – 32 dB(A) in another metric. Fluctuation strength results showed a further step towards a metric that correlates with listener response but it was evident from the results that not even a perception based metric can ever account for contextual and attitudinal aspects of annoyance rating.

## *10.3 Scope and future work*

This study has focussed on steady AM sounds with constant AM amplitude. In reality both the modulation amplitude and spectral characteristics can vary widely on time scales as short as a few seconds. The occurrence of AM has also been observed to be intermittent at times. While both phenomena will certainly affect annoyance it is with the current knowledge on AM not possible to define a representative set of stimuli to study listener perception of this phenomenon.

The listening tests were designed to allow relative comparisons between ratings both for the absolute ratings and for the ABBS adjusted ratings. It is therefore not advisable to compare the results directly to survey studies that were conducted in participants homes because of the contextual differences between the different types of study.

# 11. References

## *11.1 Reports from other work packages within the project*


WPA2: Smith, M. G., (2012) Wind Turbine Amplitude Modulation: Research to Improve Understanding as to its Cause & Effect: WP-A report: Mechanisms of amplitude modulation in wind turbine noise, ISVR, Southampton, UK

WPB1: White, P.R., (2012) Wind Turbine Amplitude Modulation: Research to Improve Understanding as to its Cause & Effect: WP-B1 report: The Measurement and Definition of Amplitude Modulations, ISVR, Southampton, UK.

WPC: Cand, M.M., Bullmore A.J, (2012) Wind Turbine Amplitude Modulation: Research to Improve Understanding as to its Cause & Effect: WPC - Collation and Analysis of Existing Acoustic Recordings, Hoare Lea, Bristol, UK

WPF: Cand, M.M., Bullmore A.J., (2012) Wind Turbine Amplitude Modulation: Research to Improve Understanding as to its Cause & Effect: Collation of Work Package Reports and Final Reporting, Hoare Lea, Bristol, UK


## *11.2 General literature*


[1] Moorhouse, A., von Hünerbein, S., Hayes, M., Adams, M., and Piper, B., (2007), Research into aerodynamic modulation of wind turbine noise, DEFRA, BERR, DCLG

[2] ETSU-R-97, (1997) The assessment and rating of noise from wind farms. http://webarchive.nationalarchives.gov.uk/+/http://www.berr.gov.uk/energy/sources/renewables/explained/wind/onshore-offshore/page21743.html, [29 June 2011]

[3] Pedersen, T. H. (2008) The Semantic Space of Sounds – Lexicon of Sound-Describing Words – Version 1, Technical Report AV 11/05, Delta, Denmark http://www.madebydelta.com/imported/senselab/The_Semantic_Spaces_of_Sounds.pdf [2/12/2011]

[4] Pedersen, E. (2009) Effects of wind turbine noise on humans, Wind Turbine Noise Conference, Aalborg, Denmark, 17-19 June 2009

[5] von Hünerbein, S.; King, A.; Hargreaves, J.; Moorhouse, A. and Plack, C., Perception of Noise from Large Wind Turbines (EFP-06 Project), The University of Salford, Greater Manchester UK, November 2010.

[6] Poulson, T. (1991) Influence of session length on judged annoyance. *Journal of Sound and Vibration*, 145 (2), 8 March 1991, Pages 217–224

[7] Møller, H., Annoyance of Audible Infrasound, *Journal of Low Frequency Noise and Vibration*, 6 (1) pp. 1-17, 1987

[8] Schultz, T. J. (1978) Synthesis of social surveys on noise annoyance. *Journal of the Acoustical Society of America, 64 (2),* 377-405

[9] Fastl, H. and Zwicker, E., (2007) Psychoacoustics, Facts and Models. Springer, Berlin, Germany,

[10] Van den Berg, G. (2006). The sound of high winds: the effect of atmospheric stability on wind turbine sound and microphone noise: PhD Thesis, University of Groningen, Netherlands, pp 210,
http://dissertations.ub.rug.nl/faculties/science/2006/g.p.van.den.berg/ (27/4/2010)

[11] Lenchine, V.V. (2009), Amplitude modulation in wind turbine noise, Proceedings of ACOUSTICS, Adelaide, Australia, November 2009.

[12] Legarth, S. V. (2007) Auralization and Assessments of Annoyance from Wind Turbines. Second International Meeting on Wind Turbine Noise. Lyon, France, 20-21 September 2007, pp 14

[13] Jiggins, M. (1997) Subjective Perception of Wind Turbine Blade Noise, thesis, Edinburgh, UK, October 1997

[14] Persson, K., Waye, K. P. & Öhrström, E. (2002) Psycho-acoustic characters of relevance for annoyance of wind turbine noise. *Journal of Sound and Vibration*, 250, pp. 65-73.

[15] Lee, S., Kim, K., Lee, S., Kim, H., and Lee, S., (2009) An estimation method of the amplitude modulation in wind turbine noise for community response assessment, Third International Meeting on Wind Turbine Noise, Aalborg, DK, June 2009.

[16] Vos, J.; M., Houben; M. J., van der Ploeg, F., Buikema, E. (2010) Annoyance caused by low frequency sounds: spectral and temporal effects, Internoise, Lisbon, Portugal, June 2010

[17] Moorhouse, A., Waddington, D. and Adams, M., (2005), 'Procedure for the assessment of low frequency noise complaints', DEFRA Contract No NANR45, null, pp.24.
http://archive.defra.gov.uk/environment/quality/noise/research/lowfrequency/documents/nanr45-criteria.pdf [21/02/2012]

[18] Bass, J. (2011) Investigation of 'Den Brook' Ampitude Modulation methodology for wind turbine noise. Acoustics Bulletin, 36(6), pp. 18

[19] ISO 1996-2 (2007) Acoustics – Description, assessment and measurement of environmental noise – Part 2: Determination of environmental noise levels. Annex C: Objective method for assessing the audibility of tones in noise. Second edition 15-3-2007.

[20] IEC 61400-11 (2006) Acoustic noise measurement techniques IEC. [3/4/2012]

[21] Plovsing, B. & Kragh, J. (2006) Nord2000. Comprehensive Outdoor Sound Propagation Model. Part 1: Propagation in an Atmosphere without Significant Refraction. Revised Report AV 1849/00, DELTA Acoustic & Vibration.

[22] Celestinos, A. and Nielsen, S. B. (2008) Controlled Acoustic Bass System (CABS): A Method to Achieve Uniform Sound Field Distribution at Low Frequencies in Rectangular Rooms., *J. Audio Eng. Soc.*, 56(11), 915 – 931

[23] di Napoli C. (2011) Long Distance Amplitude Modulation, Fourth International Meeting on Wind Turbine Noise, Rome, Italy, 11-14 April 2011

[24] ISO 15666, (2003) Acoustics - Assessment of noise annoyance by means of social and socio-acoustic surveys

[25] Howitt D. and Cramer, D (2011), Introduction to SPSS Statistics in Psychology: For Version 19 and Earlier, Prentice Hall; 5 ed., pp. 568

[26] Rasch, D. and Kubinger K. (2011) Statistics in Psychology Using R and SPSS, Wiley-Blackwell, pp. 568

[27] King, A. (2012) Perception of Low Frequency Noise From Large Wind Turbines, MSc thesis, University of Salford.

[28] Zimmer, K. & Ellermeier, W., (1998) Ein Kurzfragebogen zur Erfassung der Lärmempfindlichkeit. *Umweltpsychologie*, 2 (2), 54-63.

# 12. Glossary

| | |
|---|---|
| ABBS | Adaptive BroadBand Stimulus: un-modulated broadband WTN noise with the same spectral shape as the un-modulated (0 dB) stimulus (masking noise without garden noise) that was used in the final test. The participants adjusted the level of the ABBS until it was equally annoying as the modulated wind turbine sound (*paired comparison method*). |
| AM | Amplitude modulation |
| Annoyance | An unpleasant mental state that is characterised by effects as irritation. It can be distracting and lead to frustration and anger. |
| Background Noise (BN) | Any noise present at the listener position that does not originate from the wind turbines |
| CABS | Controlled Acoustic Bass System (CABS), a special arrangement of subwoofers to minimise room modes. |
| 95 % CI | A 95 % Confidence Interval (CI) is a statistical measure that provides an estimated range in which the rating is expected to fall in 95 % of all cases. It is calculated as the 1.96*standard deviation/√(number of ratings) |
| Frequency skew | If the frequency distribution of a signal has unequal energy in either the higher or the lower frequency range it is said to be skewed. This is visualised in Figure 17.6 where the red curve "leans" to the left in contrast to Figure 17.7 where the red curve is symmetrical around its maximum value. |
| Garden Noise (GN) | Type of noise that is heard from outside a residence which in the context of this study is regarded to be the sum of vegetation and other outdoors background noise. |
| GLM ANOVA | Abbreviation for General Linear Model ANalysis Of VAriance. Standard statistical method used here to decide whether mean values are different or the same. |
| Masking Noise (MN) | Any noise that reduces or eliminates the audibility of a particular noise source. In this study the term is used for either WTN on its own or the combination of WTN and GN all or some of wind turbine noise, vegetation noise and background noise. |

| | |
|---|---|
| Modulation depth parameter α | A measure of the strength of amplitude modulation as defined in the annex to WPB1. |
| Modulation Depth MD | In this report, MD is derived from 100 ms averages of $L_{Aeq}$: The modulation depth is defined as the difference between the mean peak level and the mean trough level in the A-weighted RMS time series for any consecutive group of all pulses over the length of the test stimuli (Figure 17.8 and Figure 17.9). |
| MPL | Peak level of the modulated part of the signal (highest value in blue curves in Figure 6.1) (MPL) |
| Medium Frequency Amplitude Modulation (MFAM) | Wind turbine amplitude modulated sound with a frequency content centred between 500 and 1000 Hz. It can be described as a “swish”. |
| Reduced Frequency Modulation (RFAM) | Wind turbine amplitude modulated sound with a frequency content centred between 200 and 600 Hz. It can be described as a “swoosh” or “whoomp”. |
| Rise time | Property of the triangular shaped envelope of a modulated signal that describes for how long the signal amplitude increases. |
| Saw tooth | Triangular shaped envelope of a modulated signal |
| Spectral shape | The range of frequencies contained in a sound |
| Time envelope | The shape of the modulation in a graph showing the modulated signal as a function of time. Red line in Figure 6.1. |
| Vegetation noise (VN) | Sound originating from vegetation and often masking wind turbine noise |

# 13. Appendix I: Participant recruitment documents

## *13.1 Advertisement Final Test I*

THE UNIVERSITY OF SALFORD
University of Salford
A Greater Manchester University

### UNIVERSITY OF SALFORD NEWS RELEASE

### Salford seeks paid volunteers for sound study

Would you like to be involved in research that paves the way for reducing noise problems from the adoption of wind energy?

Are you interested in issues of noise control?
Do you get frustrated by noisy environments?
Do you find fluctuating sounds annoying or bothersome?

The University of Salford's Acoustics Research Centre is seeking volunteers for a study into the preference for and against different amounts of Amplitude Modulation (how much a sound fluctuates, or 'comes and goes', periodically) for a range of sounds/wind turbine sounds.

Participants will have their hearing tested, and then they will rate how annoying they find a series of sound samples in a living room environment. In total, participation will take approximately 2 hours.

**Eligible volunteers will be paid for their time.**

People who are over 18 years old and without hearing problems are asked to apply. Applicants will be asked some questions for screening purposes.

Testing will take place between ##/##/## and ##/##/##.

For more information or to apply contact Andrew King by emailing a.king@edu.salford.ac.uk or call 0161 295 4669.

*13.2 Advertisement Final Test II*

University of Salford MANCHESTER

Dr Benjamin Piper
Acoustic Research Centre
G11 Newton Building
University of Salford
M5 4WT
07538277816

# LEND US YOUR EARS

Next Week 6th-9th of December 2011, the Acoustics Research Centre, based in the Newton Building, would like to invite you to take part in a subjective listening test, can you help?

The project aims to better our understanding of what makes noise from wind turbines a problem for residents near them.

The experiment starts with a basic hearing test. Then you to listen to a collection of wind turbine noises in a specially designed, surround-sound listening room and rate how annoying they would be to hear in your garden.

You will receive payment for participating.

In total the testing should take 90 minutes, however this can be broken up into shorter sessions if that is more convenient. Your participation would make a significant contribution to cutting edge research and it would be a great help to the researchers.

If you have a preference for living or spending time in the countryside then we would be particularly eager to hear from you.

If you have any questions, or would like to take part, please get in touch with Ben Piper at **b.j.piper1@salford.ac.uk**

### *13.3 Consent Form*

# Consent Form

Project : Comparative Annoyance from Amplitude Modulated Noise

Researcher : Benjamin Piper

Contact Details: researcher @salford.ac.uk

Supervisor : Sabine von Hünerbein

Contact Details: supervisor @salford.ac.uk

Thank you for agreeing to participate in this study, taking place on ......................................

This form outlines the objectives of the study and your involvement.

The objectives are:

- To measure how sensitive your hearing is.
- To compare how annoying you find different types of noise.

First we will test your hearing in an audiometric booth in which you will be asked to indicate when you can hear a tone by holding down a button. The testing is done on each ear individually over headphones. If your hearing is not sensitive enough, you may not continue the experiment.

In the second test you will sit in a listening room (designed similar to a living room) and will be asked to imagine you are at home, in the garden or living room whilst you listen to sounds and rate them on a scale of annoyance and by adjusting one until it is equally as annoying as another.

The levels of sound are quite low, so there is no risk to your hearing or your health. There will be regular breaks approximately every 30 to 40 minutes. However, if you feel tired or uncomfortable at any time or would like a break please press the 'help' button to pause the test and alert the researcher. The experiment should take no more than 2 hours in total.

The information gathered from this study will be used for no other purpose except the completion of this study and the publication of its results. The results of this test will be stored anonymously. **Your participation is voluntary – you have the right to withdraw at any time without giving any reason and your data will not be used.**

Please feel free to ask any questions at any time about the nature of the study and methods being used – the contact details are listed above.

☐ Please tick this box if you would like to be de-briefed after the current study.

☐ Please tick this box if you are happy to be contacted about participating in the future.

Participant : I agree to the terms

Name .................................................. Signature ............................................... Date ....................

Researcher : I agree to the terms

Name .................................................. Signature ............................................... Date ....................

# 14. Appendix II: Participant Screening for Final Tests

This section details the exact wording and layout of the screening form sent to prospective participants. The Noise Sensitivity Scale at the end is the short version of the Zimmer and Ellermeier Noise sensitivity scale [28]:

All information provided here will be kept confidential (only available in its raw form to project members) and shall not be published in any way that identifies the participant. Information shall only be kept if applicant participates.

| | | | |
|---|---|---|---|
| Forename: | | Surname: | |
| Age: | | Sex: | Male / Female<br>(Delete as appropriate) |
| Occupation: | | Nationality: | |
| Previous listening test experience | **A lot** (participated in more than 5 tests)<br>**Some** (participated in between 2 and 5 tests before)<br>**A Little** (participated in 1 test before)<br>**None** (never participated in a test before)<br>(Delete as appropriate) | | |

For the following questions, please give one answer by deleting the answers that do not apply to you.

**Q1. What best describes the area surrounding your home?**

Inner city
Suburb (eg. City outskirts)
In the countryside
Other (please specify) ………………………………………………..

**Q2. How content are you with the area surrounding your home?**

Very unhappy

Unhappy

Neither unhappy or happy

Happy

Very Happy

If you wish you lived in a different area type, please answer Qs 3 and 4. If not, please go to Q5.

**Q3. Which of the following area types do you wish you lived in?**

Inner city

Suburb (eg. City outskirts)

In the countryside

Other (please specify) ……………………………………………………

**Q4. How strong is your desire to live in the area selected in q.3?**

Strong

Moderately Strong

Moderately Mild

Mild

**Q5. How good is your hearing, in general?**

Very good

Good

Moderate

Poor

Very Poor

**Q6. Do you have any specific problems with your hearing?**

Yes

No

If yes, please provide details in the space below

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

Now please complete the Noise Sensitivity Scale below. Show whether you agree fully, rather agree, rather disagree or fully disagree with each statement by putting a tick in the relevant box.

| **Noise Sensitivity Scale** | **Agree fully** | **Rather agree** | **Rather disagree** | **Disagree fully** |
|---|---|---|---|---|
| 1. It is no fun keeping up a conversation while the radio is on. | | | | |
| 2. I tend to notice disturbing sounds later than do other persons. | | | | |
| 3. I avoid noisy pastimes such as going to soccer matches or fairs. | | | | |
| 4. I wake up at the slightest sound. | | | | |
| 5. Even in noisy surroundings. I am able to work quickly and with concentration. | | | | |
| 6. On doing my shopping in the city. I hardly hear the street noise. | | | | |

| | | | | |
|---|---|---|---|---|
| 7. After having passed an evening in a noisy pub<br>I feel drained. | | | | |
| 8. When I want to fall asleep, hardly any sound can disturb me. | | | | |
| 9. On weekends I like to be in quiet places. | | | | |

**Thank you very much. We will contact you soon about participation, which is scheduled to take place in April.**

**Figure 14.1 Zimmer-Ellermeier Questionnaire for the assessment of noise sensitivity**

The items within the scale are scored: 1: 3-0, 2: 0-3, 3: 3-0, 4: 3-0, 5: 0-3, 6: 0-3, 7: 3-0, 8: 0-3, 9: 3-0 (scored left to right across response boxes).

# 15. Appendix III: Check lists for listening test procedures

## *15.1 Participant screening*

- When an individual emails to show interest in taking part, email them back with thanks, and ask them to fill in the screening document you attach to your reply.
- If they do not reply within a couple of days, send reminder asking if they still want to take part.
- When they return the Screening document, input the results into password protected participant file:
- To calculate the Noise Sensitivity Score, use Zimmer-Ellermeier scoresheet. This gives a list for the scores for each item of the NSS questionnaire. For each item, note the score in the corresponding box. Notice they either go 0-3 or 3-0 left to right. Add up the scores and multiply by 3.7 for the participant's NSS score.

**Selection:**
We are looking for people who either say they want to live in the countryside or already do live in the countryside. Obviously the preference is for people who have a stronger desire to live there, or are happy there.

Last time we found countryside dwellers and those wanting to had higher scores on the NSS than the other demographics, so I would say to aim for participants will scores between 40-80.

## *Pre-test setup routine*

- Send reminder email to next day's participants
- Update the respective password protected subject data files on the Listening room PC if any new participants are in that day, then run Compile_Participant_List.m in main folder.
- Go around outside of curtain, turning on all the mains sockets with plugs inserted (pretty much every socket!), checking that all the subs are on (little green indicator light on the side near the base. Check sound card is on too (and set to input -10 dBV.
- Turn on desktop PC & laptop for webcam.
    - On desktop, open DigiCheck twice, on each instance, press F3 and from the drop down menu labelled 'Source' select 'Input' for one and 'Playback' for the other. This tells you what is going in and what is going out of the sound card. Now open, MATLAB. Set the directory to 'D:\ReUK_AM'
- On laptop, open Creative… This should automatically find the webcam and you can resize the stream to your liking
- Best to have just centre lights on in listening room.
- Turn off ventilation
- To check all 8030 As (Ambisonic ring) are on and set to correct levels:
    - turn on Norsonic, check battery.
    - Place Microphone in listener position (using bolt on string hanging from ceiling) and point at a speaker
    - run daily_connection_check.m
    - select speaker mic is facing, get 'Volume' bar to turn green, make sure chair is not in sound path
- Measure white noise or GN with SLM compare with intended.

## *15.2 Test procedure for each participant*

- Meet participant
- Explain procedure, context, what annoyance means and how long each bit will be for that session
- Ask them to sign the consent form
- Take them to the audiometric booth,
- Perform audiometric test
- Present participant with paper instructions
- Take them to listening room.
- Ask participant whether they would like to have screen on right or left, and angle the seat and coffee table accordingly, make sure they are comfortable (ask if they would like a drink).
- Mention the ‘help’ button and that it allows them to break from testing if they wish.
- Select test m file and run, select scenario.
- Guide participant through one practice trial. Indicate the ‘Okay’ button to end practice when they are ready.
- Ask participant if they would like a break or a drink after finishing each block.
- If they finish a session, but have more to come back and do, make sure you have booked a $2^{nd}$ session.
- After finishing all testing, debrief them on what the project aims are and how their data will be used, (i.e. mean levels of all participants will be evaluated to find the point where levels start to be influenced by modulation and above this point, how it influences it).
- Have them sign the payment receipt form and pay them.

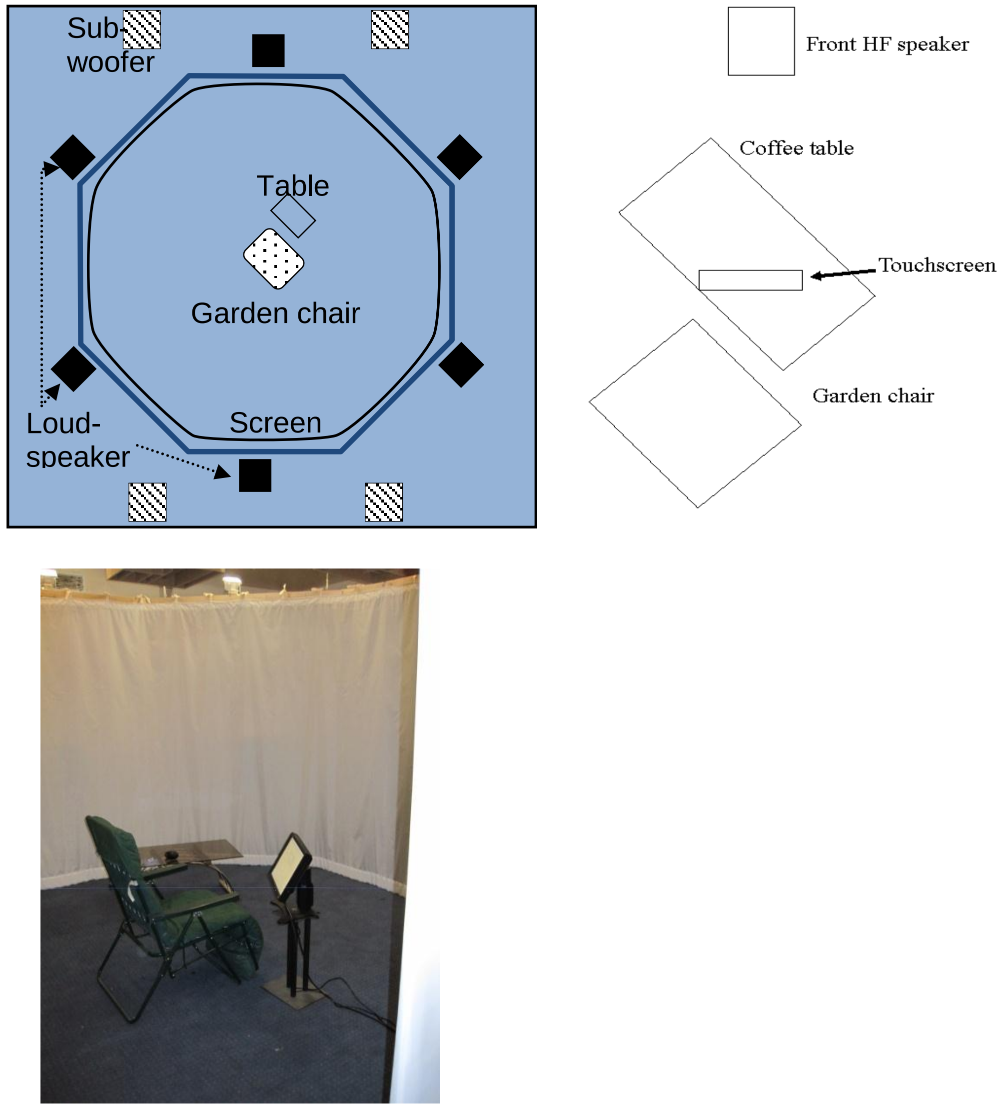



**Figure 15.1 Loudspeaker and seating arrangement in listening room**

## *15.3 Test protocol (quality assurance)*

Test stimuli have been calibrated and calibrations evaluated during set-up periods. A daily routine ensured full functionality of the reproduction system. Stimuli were presented in random order or manually counter balanced to avoid fatigue bias. Reproducible communication was ensured through written instructions and a checklist for oral communication.

## *15.4 Participant selection criteria, quantity*

Participants were mainly recruited from staff and students via the internal University communication channels. To widen the age range and background we also attempted to recruit participants from a rural area away from existing wind farms to avoid pre-sensitisation bias. Selection criteria were normal hearing tested by a standard audiometric procedure and that the participants are either living in the country side or want to live in the countryside. This is to take noise sensitivity into account. Noise sensitivity was also established by asking the volunteers to fill in a Zimmer-Ellermeier questionnaire.

# 16. Appendix IV: Sound Production System

## *16.1 Headphone calibration for Sensitivity Tests*

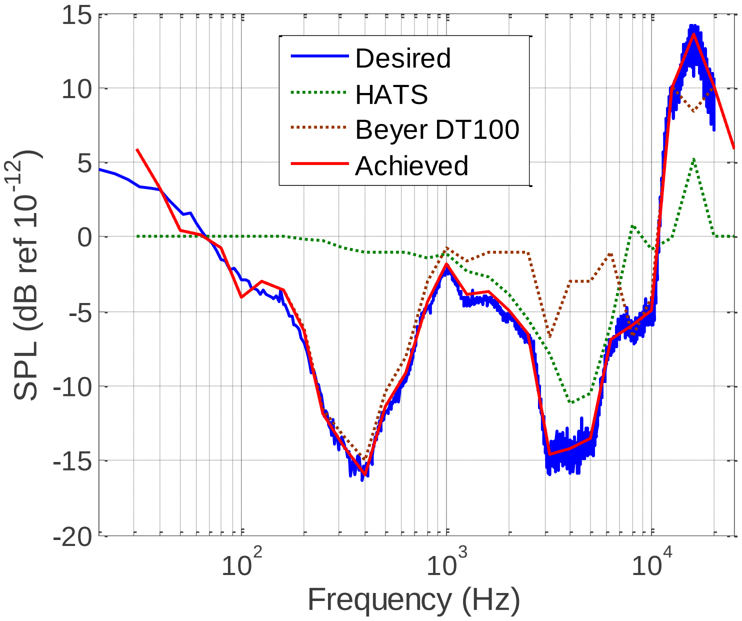



**Figure 16.1 Filters used to correct for headphone (dotted green) and HATS (dotted red) measurement system response. The blue line is the intended response, the red the smoothed achieved response.**

## *16.2 Final test*

To minimise the influence of the room on the stimuli reproduction the listening room was calibrated following the procedure laid out in [5]. For the current tests the number of loudspeakers in the ambisonic ring was reduced to six instead of eight. The calibration results for each loudspeaker are plotted in Figure 16.2 - Figure 16.15. The graphs represent a narrow-band, detailed frequency analysis which represents the effective audio reproduction of the system, for each loudspeaker. H(f) is the measured response and G(f) is the correction applied, and the difference between the two is then shown. Two sets of calibration data are shown, the first was for the first participant subgroup of final test and the second were used at the final stage of the final test.

### 16.2.1 Final Test, participant subgroup I

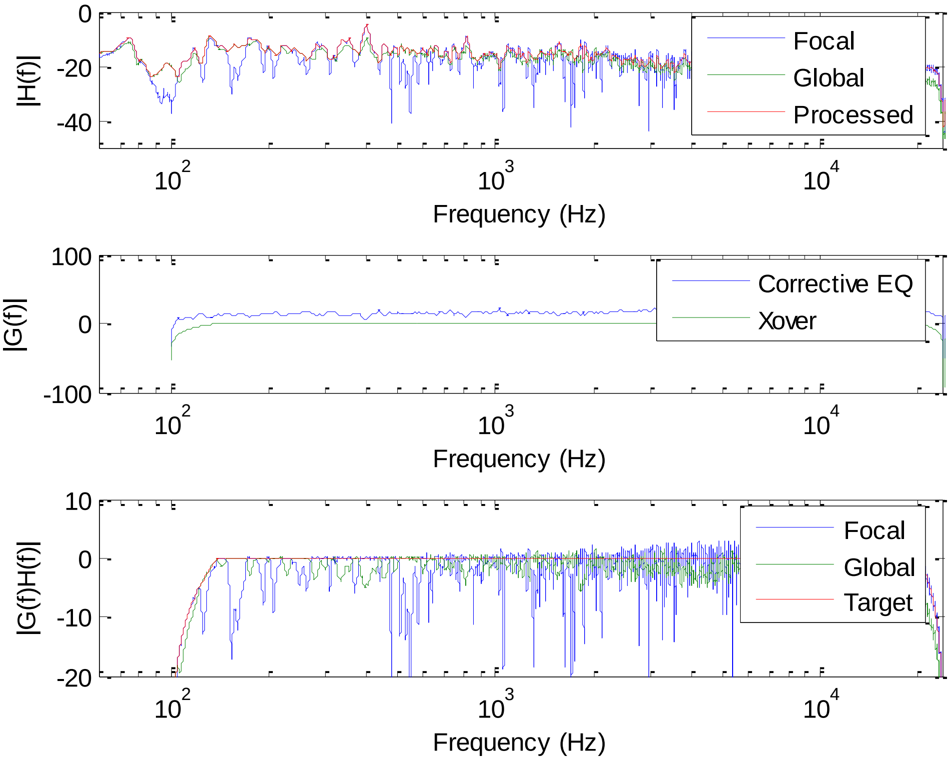



**Figure 16.2 Left Loudspeaker Response and Correction**

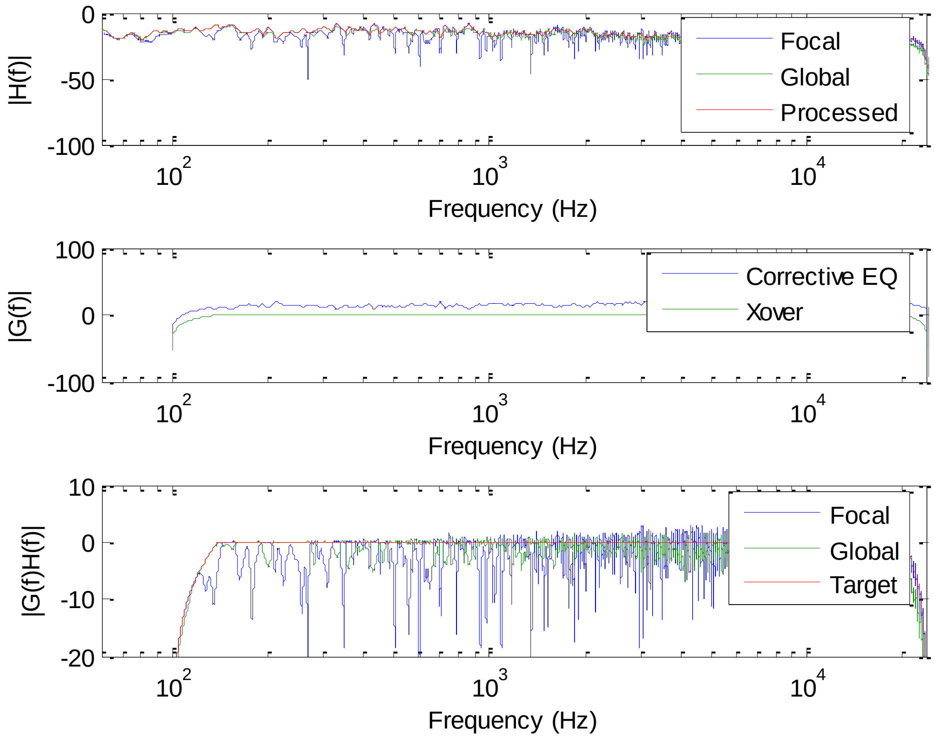



**Figure 16.3 Front Left Loudspeaker Response and Correction**

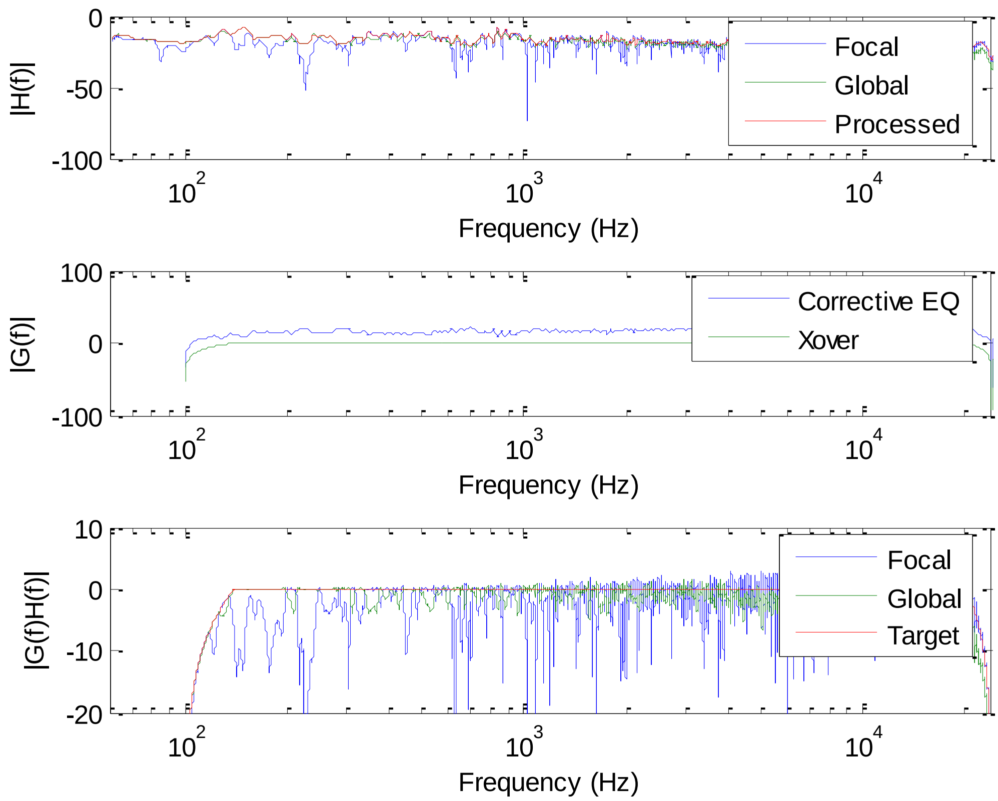



**Figure 16.4 Front Right Loudspeaker Response and Correction**

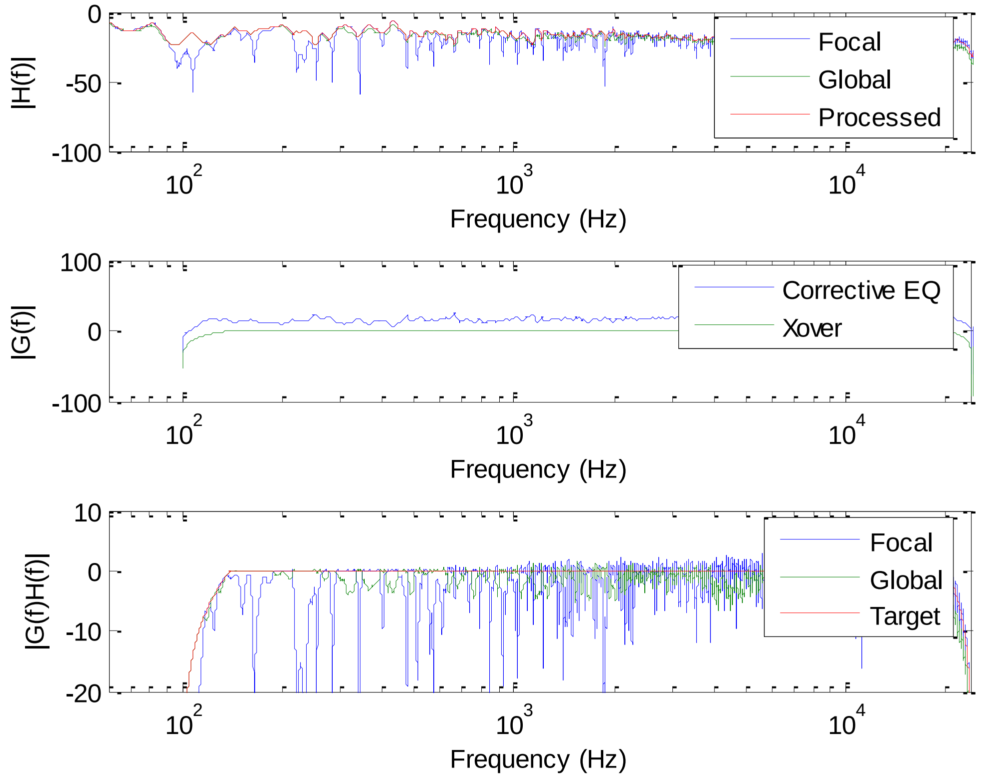



**Figure 16.5 – Right Loudspeaker Response and Correction**

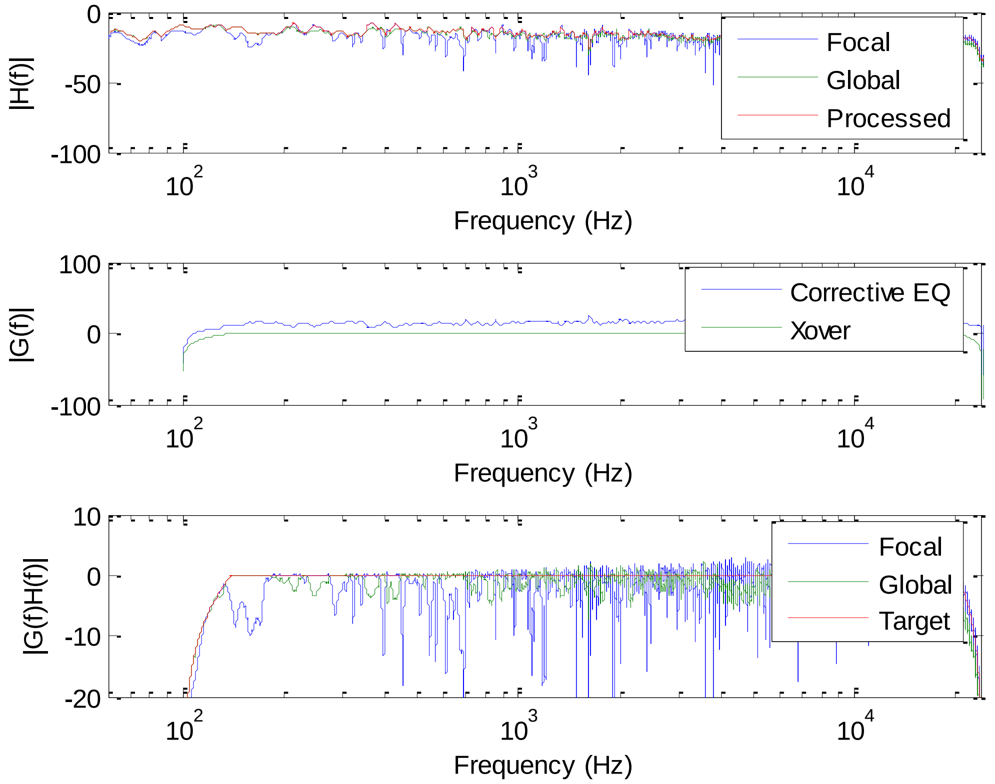



**Figure 16.6 Back Right Loudspeaker Response and Correction**

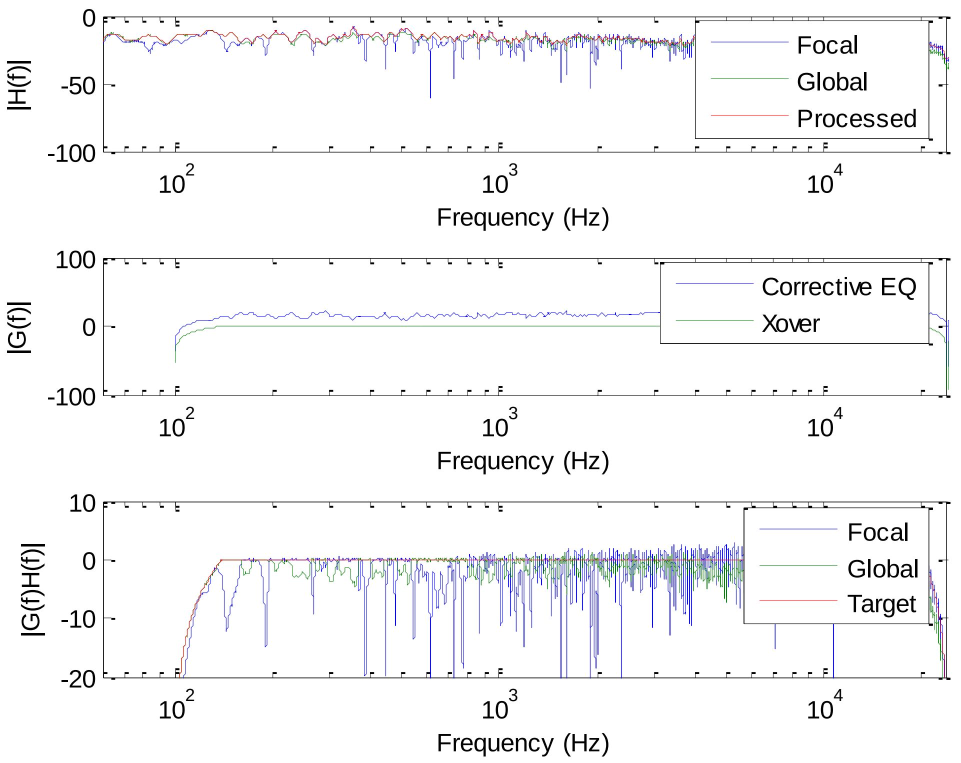



**Figure 16.7 Back Left Loudspeaker Response and Correction**

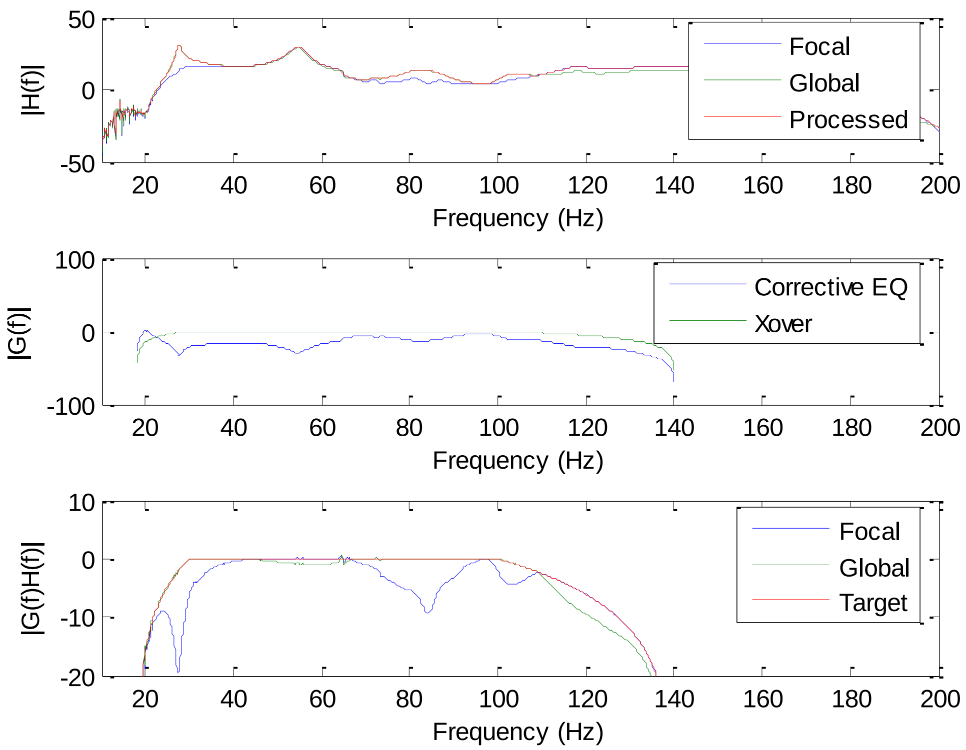



**Figure 16.8 CABS Loudspeaker Response and Correction**

## 16.2.2 Final Test, Subgroup II

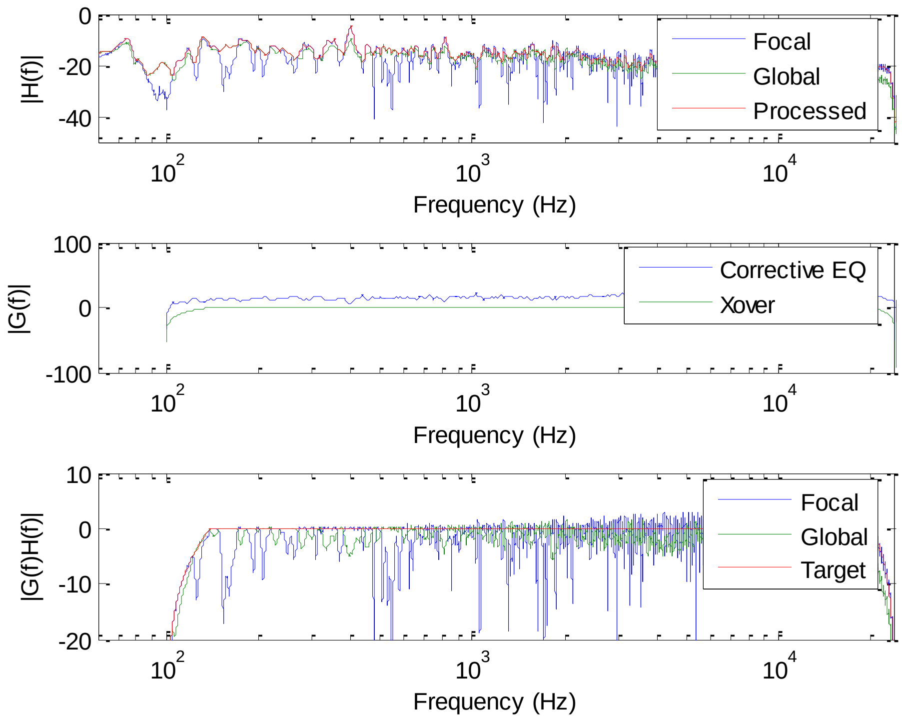



**Figure 16.9 Left Loudspeaker Response and Correction**

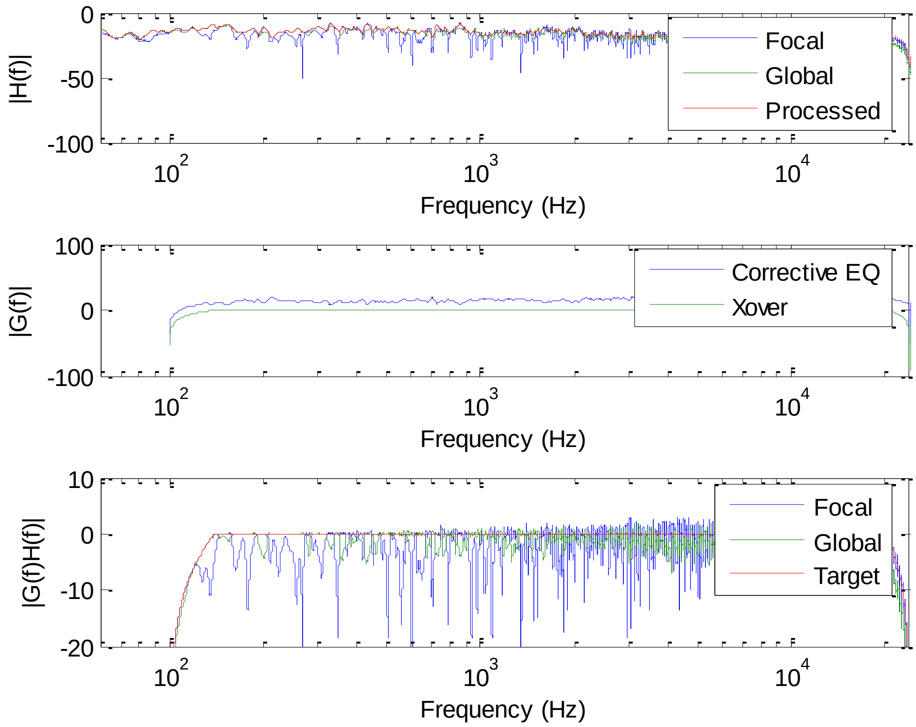



**Figure 16.10 Front Left Loudspeaker Response and Correction**

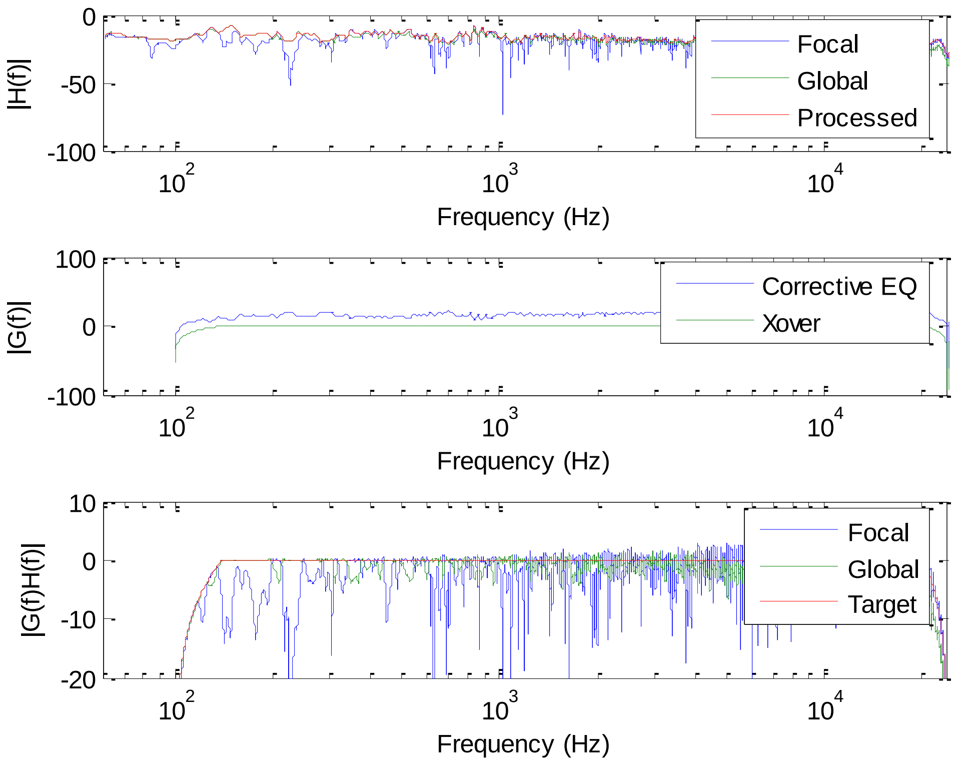



**Figure 16.11 Front Right Loudspeaker Response and Correction**

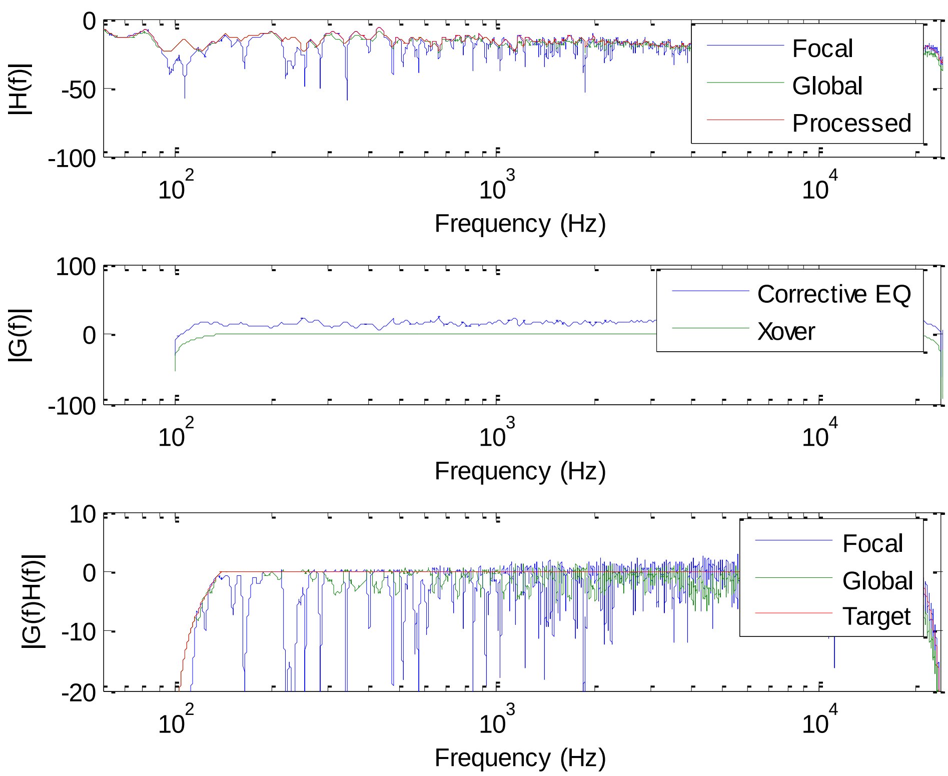



**Figure 16.12 Right Loudspeaker Response and Correction**

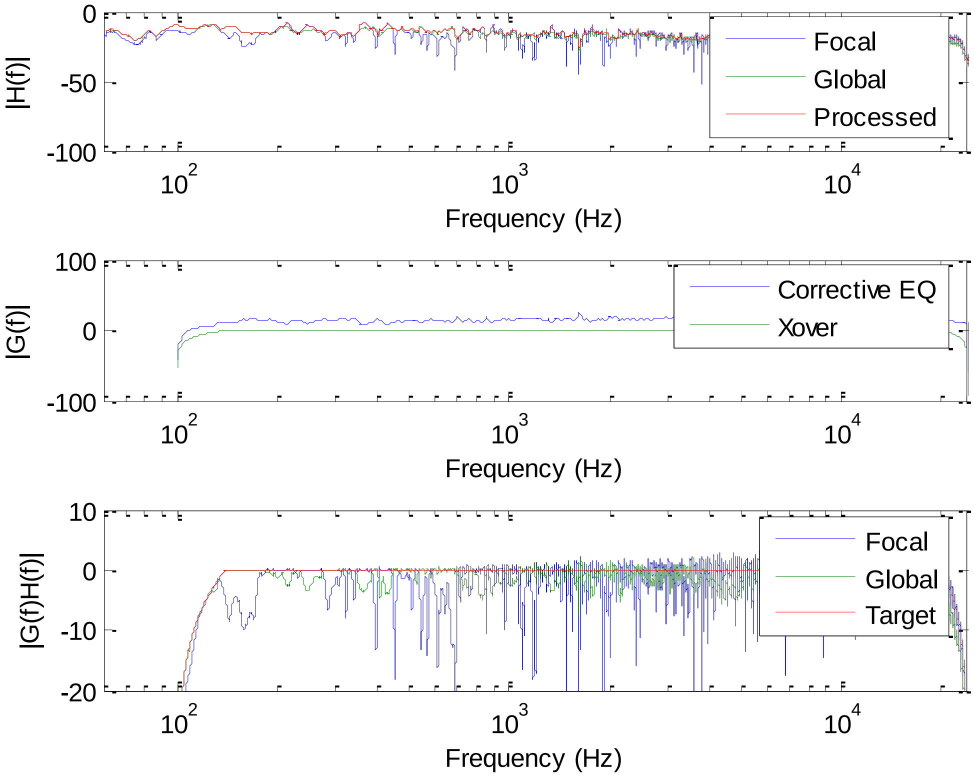



**Figure 16.13 Back Right Loudspeaker Response and Correction**

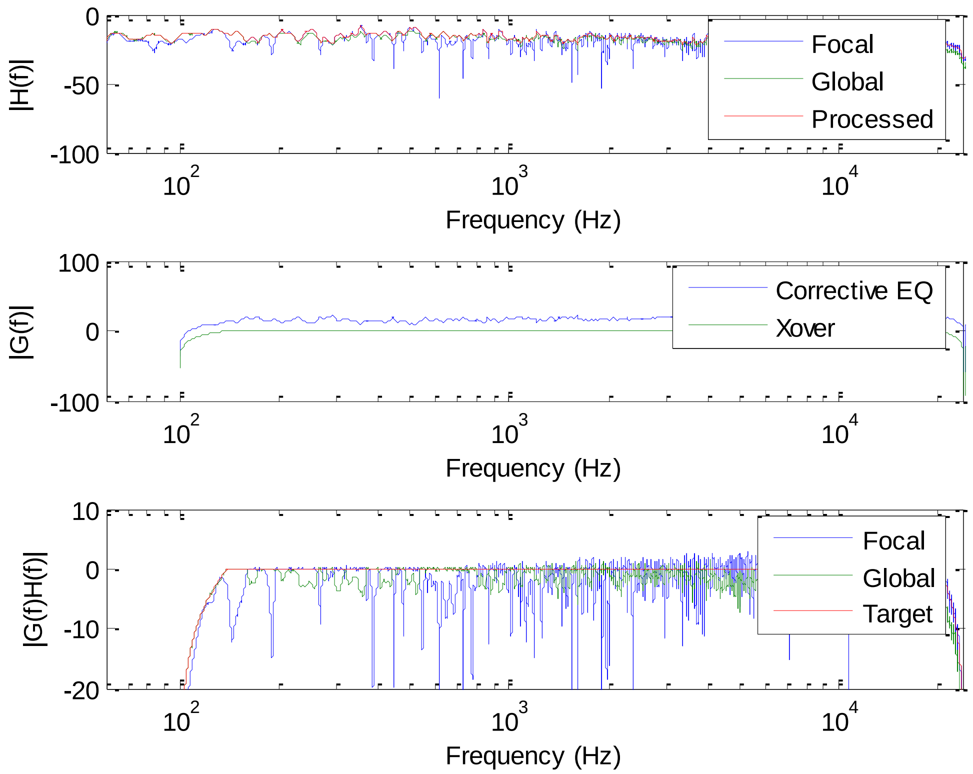



**Figure 16.14 Back Left Loudspeaker Response and Correction**

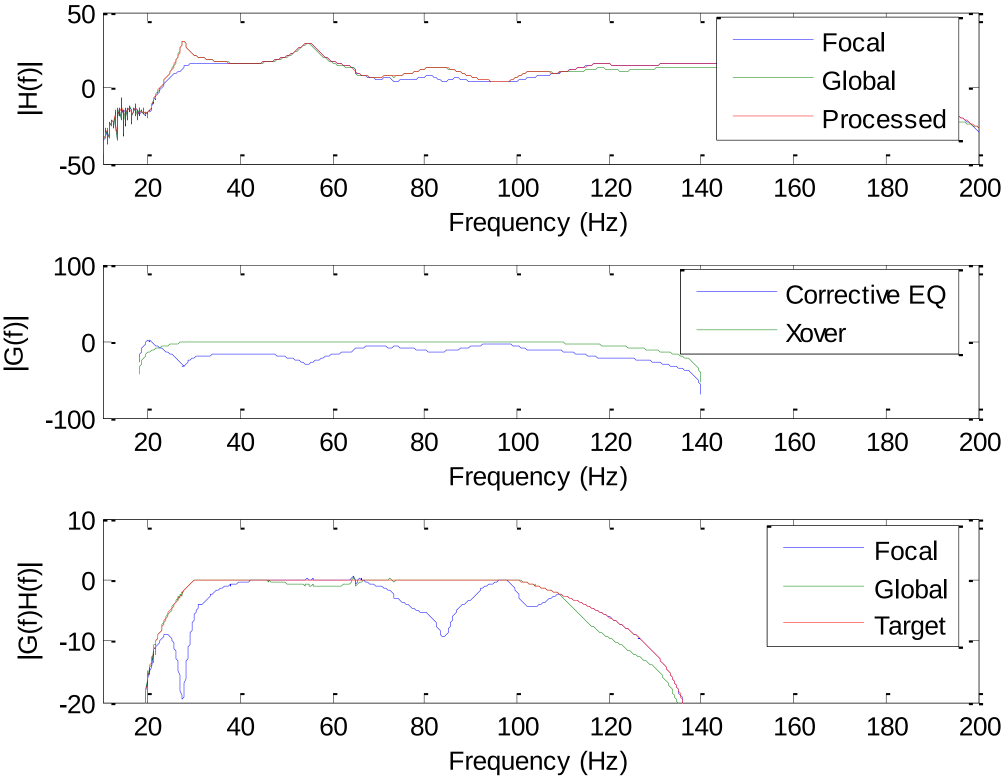



**Figure 16.15 CABS Loudspeaker Response and Correction**

# 17. Appendix V: Stimuli Design Method

This appendix contains information on how the stimuli were created from different component parts and measurements of the final stimuli.

## *17.1 Stimuli Components Step I*

The wind turbine noise signal was composed in three different steps as detailed in Sections 6.3 and 17. The MN consisted of un-modulated wind turbine noise (WTN) with and without garden noise (GN). These two noise types were created from third octave band data as detailed in [5]. Figure 17.1 shows the measured third octave spectra of these signals when played through the listening room reproduction system. All measurements were made with a Svantek 927 sound level meter (SLM) giving A-weighted RMS time history data based on a 100 ms integration time and a monophonic .wav file.

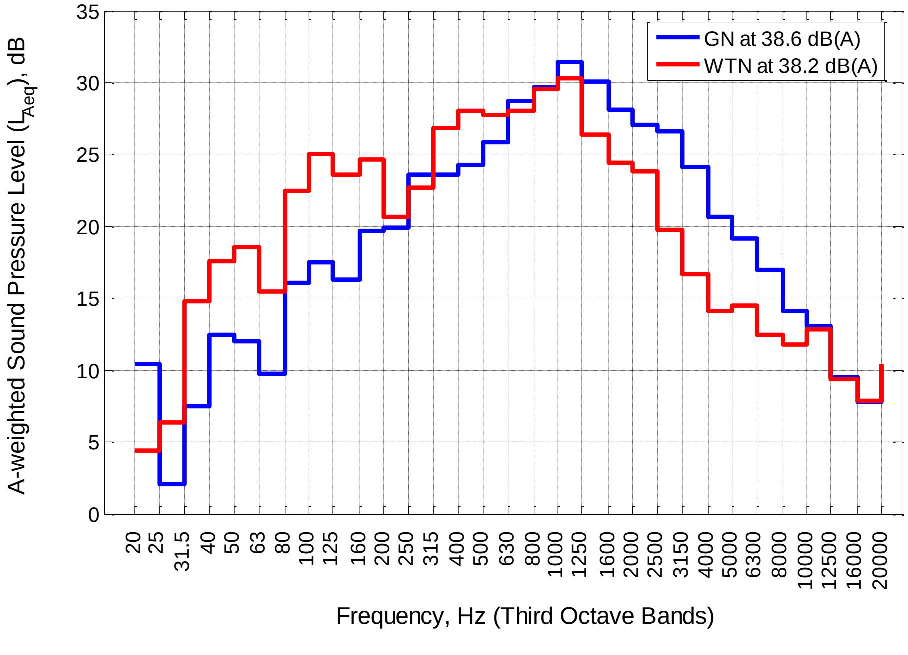



**Figure 17.1 Measured third-octave data for wind turbine and garden masking noises.**

In the Sensitivity Tests the $L_{Aeq}$ of WTN corresponded to 44 dB(A) alone, GN to 46.2 dB(A) alone, and when combined at a ratio of 5:1 between WTN and GN respectively the combined nominal level was 49.8 dB(A). The ratio between the two levels was chosen empirically in consultation with the members of the project team, which included some experienced “expert” listeners, when everybody listened to various stimuli in comparison to recordings from WPC. Absolute levels varied between participants as described in Section 6.6.1.

## *17.2 Step 2: Modulation for Sensitivity Tests I and II*

The modulated noise signals were created using the model developed and described in WPB1.

Two types of signal were created for Sensitivity Tests I and II, one to represent RFAM and one to represent MFAM. Where not otherwise specified the basic parameters are shown in Table 17.1

| | Centre Frequency (Hz) | Bandwidth – 3 dB (Hz) | Frequency Skew (%) | Pulse Width (s) | Envelope Rise time (%) |
|---|---|---|---|---|---|
| RFAM | 350 | 400 | 33 | 0.2 | 70 |
| MFAM | 600 | 350 | 50 | 0.2 | 70 |

**Table 17.1 MFAM and RFAM parameters used for stimuli in Sensitivity Tests.**

Figure 17.2 and Figure 17.3 show the spectra and time-envelope of the ‘MFAM’ and ‘RFAM’ stimuli, without the background masking noise.

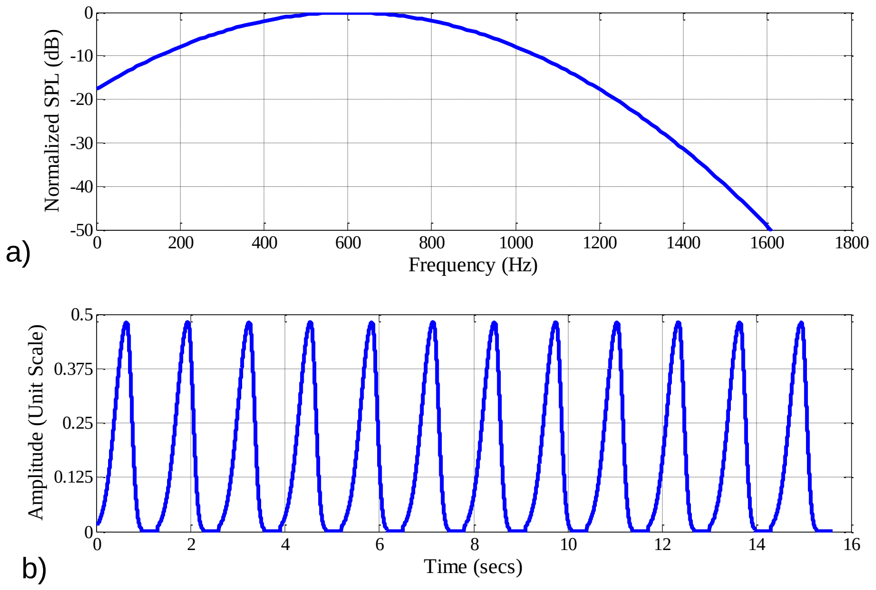



**Figure 17.2: MFAM stimulus properties a) frequency envelope with centre frequency of 600 Hz and bandwidth of 350 Hz. b) time envelope with a rise and drop time of 70%.**

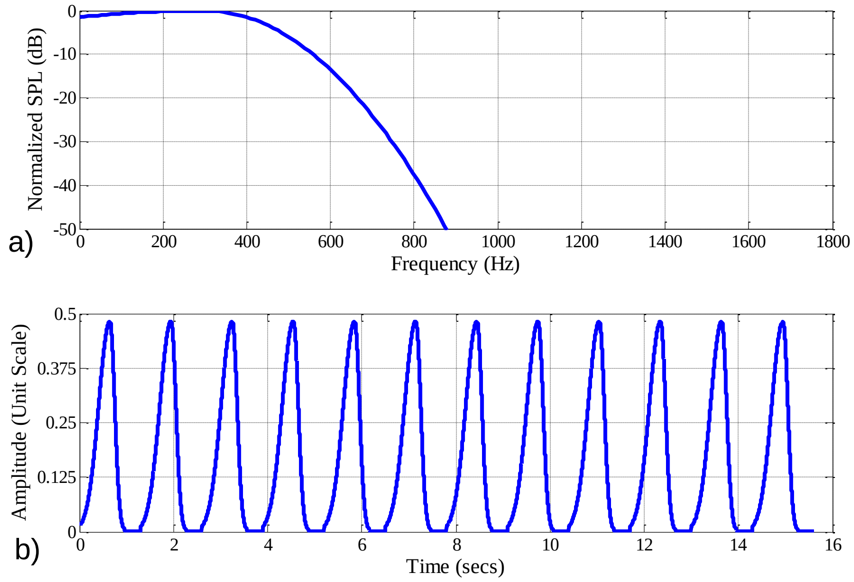



**Figure 17.3: RFAM stimulus properties a) frequency envelope with peak at 350 Hz, and bandwidth of 400 Hz. b) time envelope with a rise time of 70% and a drop time of 30%.**

Representations of two examples of finalised stimuli including the masking noise have been produced by White and are shown in Figure 17.4 and Figure 17.5.

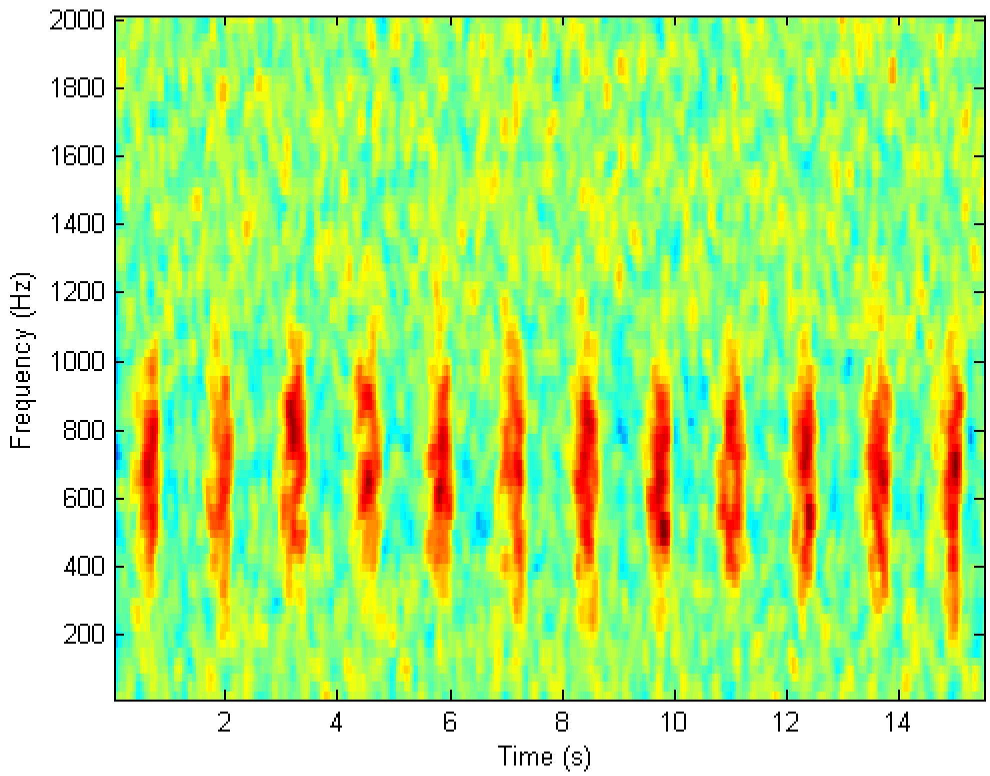



**Figure 17.4 Spectrogram of MFAM stimulus with Code Identifier AC. For details on parameter specification see Section 17.5.2.**

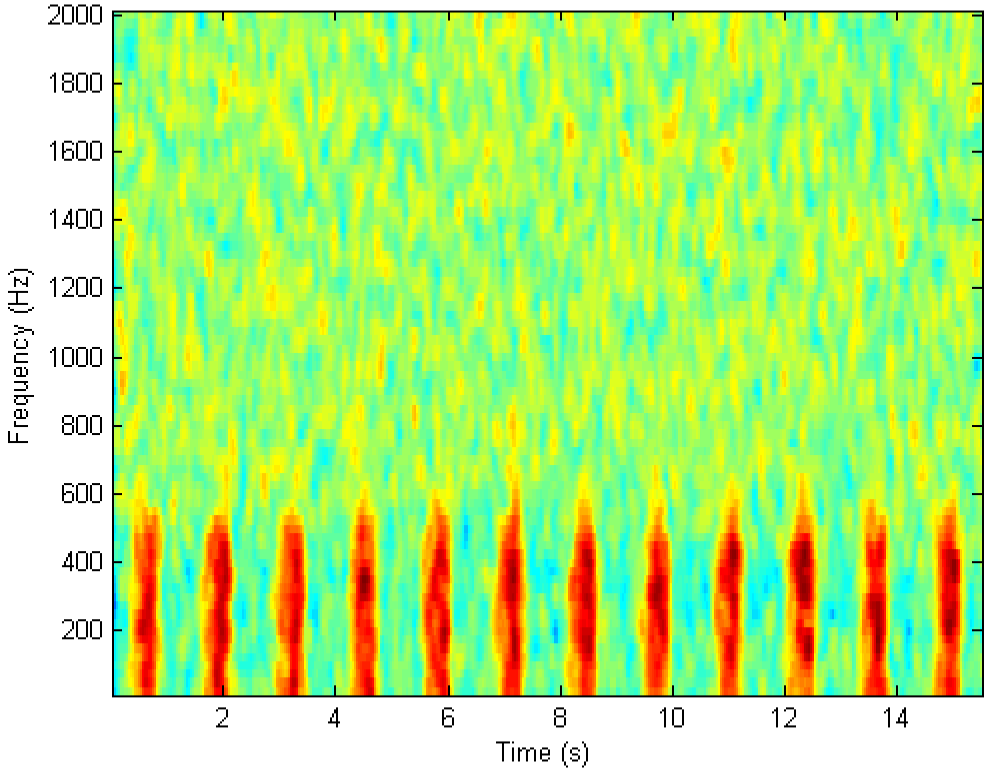



**Figure 17.5 Spectrogram of RFAM stimulus with Code Identifier MI. For details on parameter specification see Section 17.5.2.**

## *17.3 Step 2: Modulation for Final Tests I and II*

Two types of signal were created for Final Tests I and II, one to represent RFAM and one to represent MFAM. The parameters are shown in Table 17.2.

| | Centre Frequency (Hz) | Bandwidth – 3 dB (Hz) | Frequency Skew (%) | Pulse Width (s) | Envelope Skew (%) |
|---|---|---|---|---|---|
| RFAM | 300 | 180 | 33 | 0.2 | 70 |
| MFAM | 600 | 350 | 50 | 0.2 | 70 |

**Table 17.2 Parameters for RFAM and MFAM signals for Final Test and validations.**

The time and frequency content of these signals are shown in Figures 17.2 and 17.3.

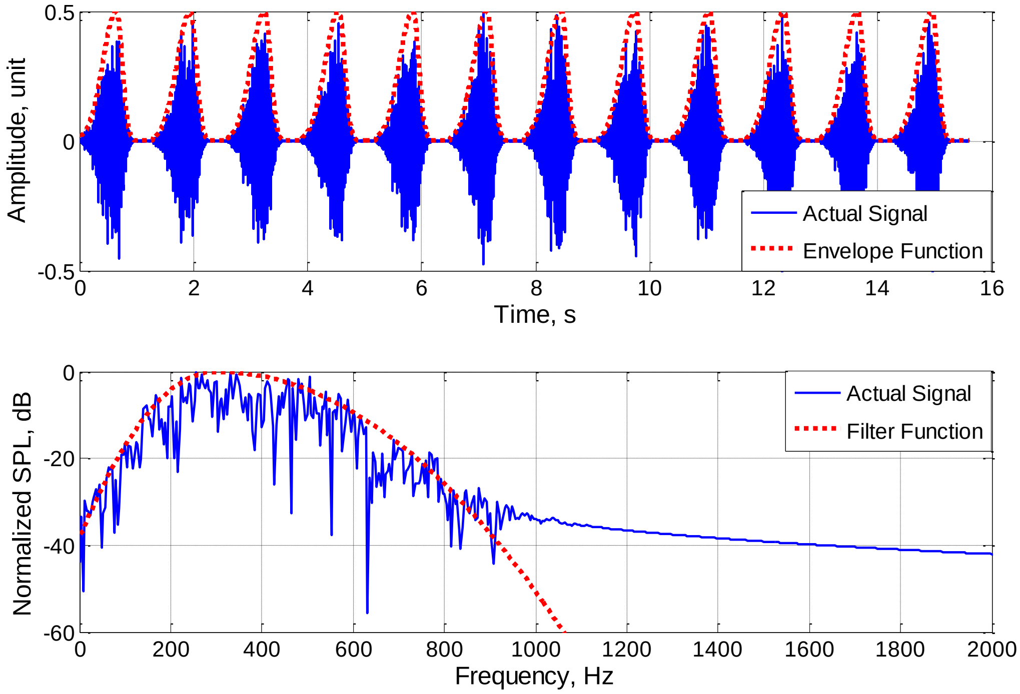



**Figure 17.6 Time and frequency content of RFAM signal used to create test stimuli (modulation pulses shown without masking noise).**

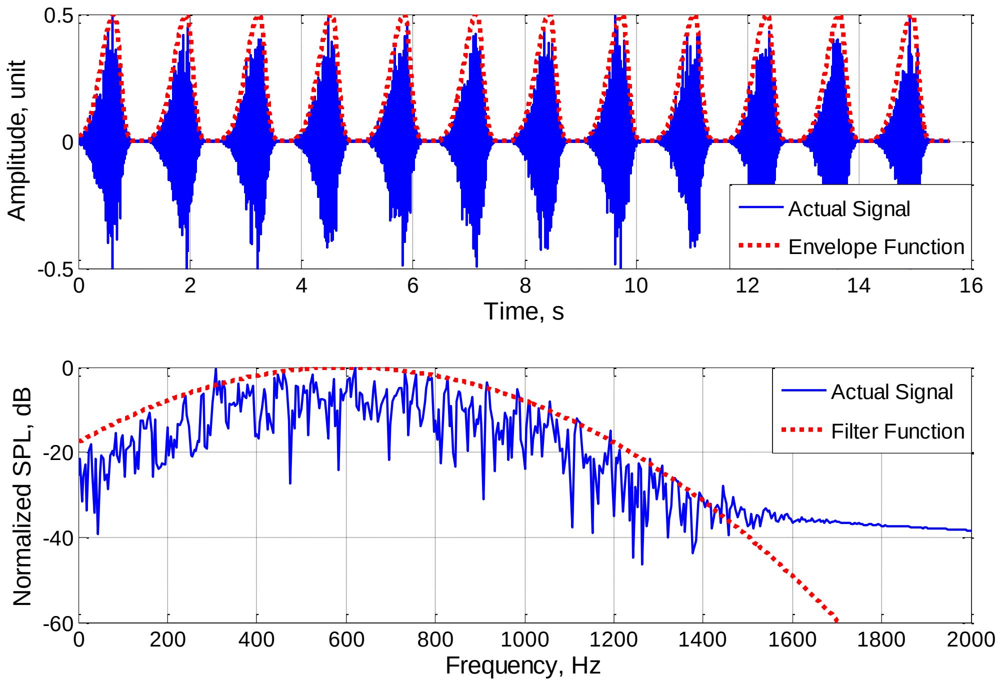



**Figure 17.7 Time and frequency content of MFAM signal used to create test stimuli (modulation pulses shown without masking noise).**

## *17.4 Approach to Combining Stimuli Components to Achieve Target $L_{Aeq}$ and Modulation Depth*

In an initial step, input levels for AM and MN components were estimated to achieve the desired representative range of stimuli. To guide this process, the modulation depth was calculated as the difference between the mean peak and the mean trough in the A-weighted RMS time series for any consecutive group of 12 pulses as 12 pulses occur in each 20 s loop. It is difficult to define the exact modulation depth below 3 dB(A) due to the variation in the masking noise therefore initially signals were created and measured for modulation depths of 3 dB(A) and above for an $L_{Aeq}$ of 40 dB(A). These were then measured at the listening position in the listening room and the ratios of the WTN and AM pulses where altered until the $L_{Aeq}$ and modulation depth was within ±0.25 dB(A) of the target. Figure 17.8 - Figure 17.10 show the measured time series for 12, 6 and 3 dB modulation depths, in that order. Note that the uncertainty of MD below 3 dB(A) is therefore larger.

Although the stimuli were generated artificially based on a fixed ratio of MN to AM pulses, the effects of random masking from the broadband nature of both components of the signal meant that, subjectively, the different pulses sounded subtly different from each other, and therefore the signal did not sound too artificial. This is reflected in the variations in short-term $L_{Aeq}$.

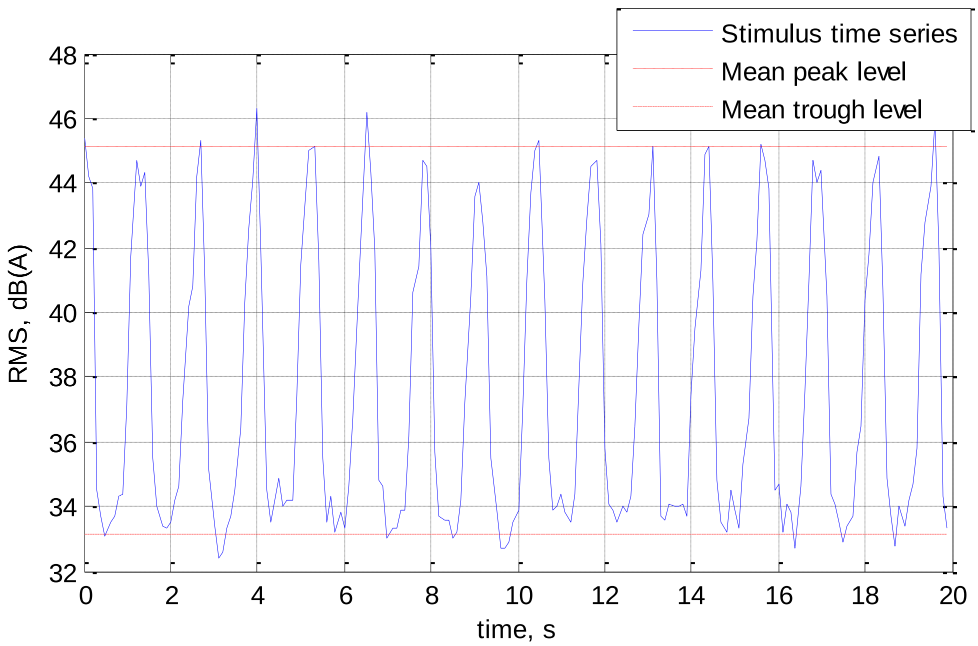



**Figure 17.8 Measured time series for stimulus (24) containing a 12±0.25 dB(A) modulation depth and with an $L_{Aeq}$ of 40±0.15 dB(A).**

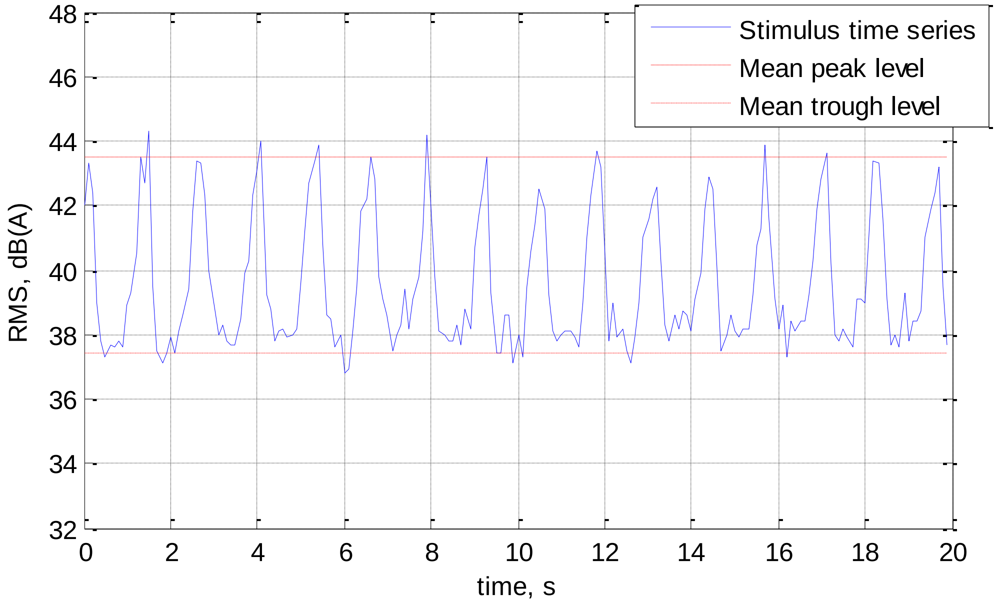



**Figure 17.9 Measured time series for stimulus (21) containing a 6±0.25 dB(A) modulation depth and with an $L_{Aeq}$ of 40±0.15 dB(A).**

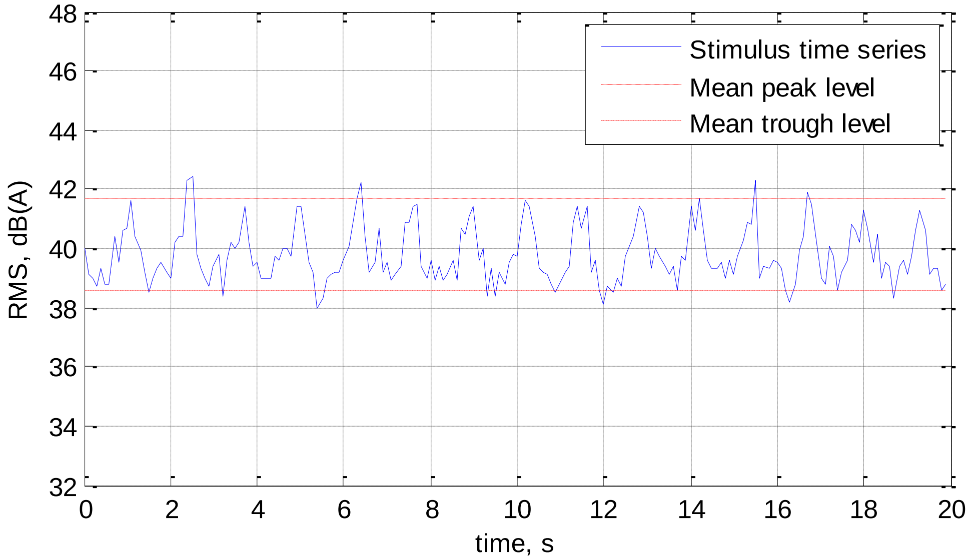



**Figure 17.10 Measured time series for stimulus (16) containing a 3±0.25 dB(A) modulation depth and with an $L_{Aeq}$ of 40±0.15 dB(A).**

Based on values used to create the stimuli with modulation depths of 3 dB(A) or above the stimuli with smaller modulation depths were created using a cubic fit to find the input levels, shown in Figure 17.11 for RFAM.

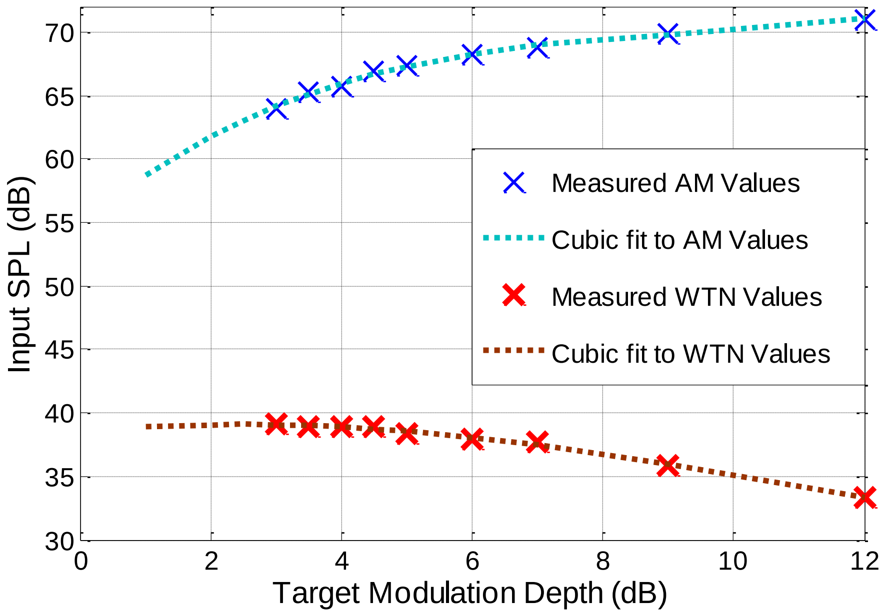



**Figure 17.11 Cubic fit to Measured AM and un-modulated WTN values to create stimuli with an $L_{Aeq}$ of 40 dB(A).**

These stimuli were measured to ensure the $L_{Aeq}$ was correct. It was then assumed that scaling all the input values up by 5 dB and down by 5 and 10 dB would give stimuli sets with $L_{Aeq}$ levels at 45, 35 and 30 dB(A). The stimuli were all measured to confirm this. The SLM had a noise floor of 20 dB(A) and showed non-linear behaviour below 25 dB(A) and therefore it is impossible from these measurements to confirm the accuracy of the stimuli with high modulation depths at 30 dB(A) and 35 dB(A). It is assumed they are correct because all the other stimuli scale in the expected manner.

This process was then repeated with added GN. For this set of stimuli continuous GN was independently played at $L_{Aeq}$ = 7.5 dB below the $L_{Aeq}$ of the AM test stimulus in order not to significantly mask modulation depths of up to 12 dB(A).

## *17.5 Detailed list of stimuli parameters*

### 17.5.1 Sensitivity Test I

Three aspects of the stimuli were varied; the modulation depth and the time envelope properties rise time percentage and pulse width based on the full width half maximum (FWHM). In the Sensitivity Tests modulation depths were different for RFAM and MFAM stimuli as initially a measure for modulation depth that was based on unweighted spectra was used to design the stimuli. This **depth parameter α** is defined in WPB1. The equivalent values of MD are specified together with all other variable modulation parameters in Table 17.3 for Sensitivity Test I and in Table 17.4.

| ID | α (dB) | MD, dB(A) | Rise Time (%) | Pulse Width (s) |
|---|---|---|---|---|
| **Garden** | n/a | 1.2 | n/a | n/a |
| **A** | 0 | 1.0 | n/a | n/a |
| **B** | 10.5 | 5.6 | 70 | 0.2 |
| **C** | 6 | 1.8 | 10 | 0.2 |
| **D** | 12 | 8.0 | 70 | 0.2 |
| **E** | 1.5 | 1.2 | 70 | 0.2 |
| **F** | 6 | 1.7 | 20 | 0.2 |
| **G** | 6 | 1.8 | 30 | 0.2 |
| **H** | 3 | 1.2 | 70 | 0.2 |
| **I** | 6 | 1.6 | 40 | 0.2 |
| **J** | 4.5 | 1.7 | 70 | 0.2 |
| **K** | 6 | 1.6 | 50 | 0.2 |
| **L** | 6 | 1.5 | 60 | 0.2 |
| **M** | 6 | 1.7 | 70 | 0.2 |
| **N** | 6 | 1.6 | 70 | 0.2 |
| **O** | 7.5 | 2.6 | 70 | 0.2 |
| **P** | 6 | 1.6 | 80 | 0.2 |
| **Q** | 6 | 1.6 | 90 | 0.2 |
| **R** | 9 | 4.1 | 70 | 0.2 |
| **S** | 6 | 1.7 | 70 | 0.15 |
| **T** | 6 | 1.6 | 70 | 0.1 |
| **U** | 6 | 1.7 | 70 | 0.25 |
| **V** | 6 | 1.7 | 70 | 0.2 |
| **W** | 6 | 1.9 | 70 | 0.35 |
| **X** | 6 | 1.3 | 70 | 0.3 |
| **Y** | 6 | 1.9 | 70 | 0.45 |
| **Z** | 6 | 2.0 | 70 | 0.4 |

**Table 17.3 Sensitivity Test I:Stimuli identifier and details on modulation parameters.**

The order of stimuli for each varied aspect is as follows:

Modulation Order from 1.5dB : 12dB - E H J M O R B D

Rise Time Order from 10% : 90% - C F G I K L N P Q

Width Order from 0.1:0.45 - T S V U X W Z Y

The samples where generated using the first edition of the ISVR code four_G_model.m

### 17.5.2 Sensitivity Test II

AM time envelope was 0.2 s

| Stimulus ID | α, dB | MD dB(A) | AM Period (s) | Bandwidth (Hz) | Peak Frequency (Hz) | Freq. Skew (%) | Rise (%) | BNL Fixed | Mod Type |
|---|---|---|---|---|---|---|---|---|---|
| **SC** | 0 | 1 | | | | | | no | |
| **VT** | 0 | 1 | | | | | | no | |
| **AC** | 3 | 5 | 1.3 | 350 | 600 | 50 | 70 | yes | MFAM |
| **DN** | 3 | 5 | 0.65 | 350 | 600 | 50 | 70 | no | MFAM |
| **EF** | 3 | 5 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **EH** | 3 | 1 | 0.65 | 400 | 300 | 70 | 70 | no | RFAM |
| **GN** | 3 | 5 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **JM** | 3 | 5 | 1.3 | 350 | 600 | 50 | 70 | yes | MFAM |
| **QD** | 3 | 1 | 1.3 | 400 | 300 | 70 | 70 | yes | RFAM |
| **TH** | 3 | 1 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **UK** | 3 | 1 | 1.3 | 400 | 300 | 70 | 70 | yes | RFAM |
| **XE** | 3 | 1 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **AV** | 6 | 8 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **AX** | 6 | 8 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **BU** | 6 | 2 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **CT** | 6 | 2 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **CY** | 6 | 2 | 0.65 | 400 | 300 | 70 | 70 | no | RFAM |
| **DB** | 6 | 8 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **EG** | 6 | 2 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **FN** | 6 | 2 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **GM** | 6 | 8 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **JO** | 6 | 2 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **KP** | 6 | 8 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **NI** | 6 | 2 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **QF** | 6 | 2 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **QS** | 6 | 2 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **RH** | 6 | 8 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **RJ** | 6 | 9 | 0.65 | 350 | 600 | 50 | 70 | no | MFAM |
| **SB** | 6 | 10 | 1.3 | 250 | 600 | 30 | 70 | no | MFAM |
| **TJ** | 6 | 2 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **UL** | 6 | 8 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **VY** | 6 | 8 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **WD** | 6 | 8 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **WG** | 6 | 2 | 1.3 | 500 | 300 | 80 | 70 | no | RFAM |
| **XT** | 6 | 8 | 1.3 | 450 | 600 | 61 | 70 | no | MFAM |
| **YK** | 6 | 8 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **ZR** | 6 | 2 | 1.3 | 300 | 300 | 67 | 70 | no | RFAM |
| **ZW** | 6 | 2 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **BY** | 9 | 3 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **CR** | 9 | 12 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **DV** | 9 | 4 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **FM** | 9 | 4 | 0.65 | 400 | 300 | 70 | 70 | no | RFAM |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **IL** | 9 | 12 | 1.3 | 350 | 600 | 50 | 70 | yes | MFAM |
| **MI** | 9 | 4 | 1.3 | 400 | 300 | 70 | 70 | yes | RFAM |
| **VU** | 9 | 11 | 0.65 | 350 | 600 | 50 | 70 | no | MFAM |
| **WF** | 9 | 11 | 1.3 | 350 | 600 | 50 | 70 | yes | MFAM |
| **YJ** | 9 | 11 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **ZE** | 9 | 3 | 1.3 | 400 | 300 | 70 | 70 | yes | RFAM |
| **FB** | 12 | 15 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **HM** | 12 | 5 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **KO** | 12 | 6 | 1.3 | 400 | 300 | 70 | 70 | yes | RFAM |
| **LP** | 12 | 15 | 1.3 | 350 | 600 | 50 | 70 | yes | MFAM |
| **NU** | 12 | 14 | 1.3 | 350 | 600 | 50 | 70 | no | MFAM |
| **RG** | 12 | 6 | 1.3 | 400 | 300 | 70 | 70 | no | RFAM |
| **SV** | 12 | 5 | 1.3 | 400 | 300 | 70 | 70 | yes | RFAM |
| **XR** | 12 | 14 | 1.3 | 350 | 600 | 50 | 70 | yes | MFAM |

**Table 17.4 Sensitivity Test II: Stimuli identifier and details on modulation parameters.**

| Stimulus ID | α, dB | MD dB(A) | Mean LAeq dB(A) | 95% CI Laeq | Mean Annoyance rating | 95% CI Annoyance |
|---|---|---|---|---|---|---|
| **AC** | 3 | 5 | 37 | 1 | 4.7 | 0.6 |
| **AV** | 6 | 8 | 44 | 1 | 7.5 | 0.6 |
| **AX** | 6 | 8 | 38 | 1 | 5.6 | 0.9 |
| **BU** | 6 | 2 | 30 | 1 | 3.6 | 0.6 |
| **BY** | 9 | 3 | 34 | 1 | 5.9 | 0.6 |
| **CR** | 9 | 12 | 37 | 1 | 6.5 | 0.7 |
| **CT** | 6 | 2 | 31 | 1 | 3.7 | 0.8 |
| **CY** | 6 | 2 | 35 | 1 | 6.3 | 0.8 |
| **DB** | 6 | 8 | 32 | 1 | 4.8 | 0.7 |
| **DN** | 3 | 5 | 40 | 1 | 6.8 | 0.9 |
| **DV** | 9 | 4 | 33 | 1 | 5.1 | 0.9 |
| **EF** | 3 | 5 | 37 | 1 | 5.3 | 0.8 |
| **EG** | 6 | 2 | 28 | 1 | 3.0 | 0.9 |
| **EH** | 3 | 1 | 38 | 1 | 5.6 | 1.1 |
| **FB** | 12 | 15 | 38 | 1 | 6.7 | 1.1 |
| **FM** | 9 | 4 | 34 | 1 | 6.3 | 1.3 |
| **FN** | 6 | 2 | 37 | 1 | 5.6 | 0.5 |
| **GM** | 6 | 8 | 44 | 1 | 7.3 | 0.8 |
| **GN** | 3 | 5 | 40 | 1 | 5.6 | 1.0 |
| **HM** | 12 | 5 | 32 | 1 | 6.0 | 0.7 |
| **IL** | 9 | 12 | 40 | 1 | 7.0 | 0.5 |
| **JM** | 3 | 5 | 34 | 1 | 4.6 | 0.9 |
| **JO** | 6 | 2 | 39 | 1 | 5.6 | 0.7 |
| **KO** | 12 | 6 | 36 | 1 | 7.1 | 0.7 |
| **KP** | 6 | 8 | 41 | 1 | 7.4 | 0.6 |
| **LP** | 12 | 15 | 44 | 1 | 8.1 | 0.6 |
| **MI** | 9 | 4 | 36 | 1 | 6.2 | 0.6 |
| **NI** | 6 | 2 | 33 | 1 | 4.4 | 0.7 |
| **NU** | 12 | 14 | 37 | 1 | 6.9 | 0.7 |
| **QD** | 3 | 1 | 35 | 1 | 3.9 | 1.0 |
| **QF** | 6 | 2 | 42 | 1 | 6.8 | 1.0 |
| **QS** | 6 | 2 | 34 | 1 | 4.9 | 0.7 |
| **QS** | 6 | 2 | 30 | 1 | 5.2 | 0.7 |
| **RG** | 12 | 6 | 35 | 1 | 5.6 | 0.7 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **RH** | 6 | 8 | 39 | 1 | 7.0 | 0.7 |
| **RJ** | 6 | 9 | 39 | 1 | 6.3 | 0.6 |
| **SB** | 6 | 10 | 39 | 1 | 3.6 | 1.4 |
| **SC** | 0 | 1 | 38 | 1 | 7.4 | 0.3 |
| **SV** | 12 | 5 | 38 | 1 | 4.6 | 0.9 |
| **TH** | 3 | 1 | 40 | 1 | 6.4 | 0.6 |
| **TJ** | 6 | 2 | 34 | 1 | 2.6 | 0.7 |
| **UK** | 3 | 1 | 35 | 1 | 5.8 | 0.8 |
| **UL** | 6 | 8 | 41 | 1 | 4.6 | 2.1 |
| **VU** | 9 | 11 | 38 | 1 | 7.7 | 0.9 |
| **VY** | 6 | 8 | 41 | 1 | 6.8 | 0.5 |
| **WD** | 6 | 8 | 38 | 1 | 6.5 | 0.5 |
| **WF** | 9 | 11 | 41 | 1 | 7.4 | 0.7 |
| **WG** | 6 | 2 | 36 | 1 | 5.2 | 0.9 |
| **XE** | 3 | 1 | 37 | 1 | 4.2 | 0.9 |
| **XR** | 12 | 14 | 43 | 1 | 8.4 | 0.7 |
| **XT** | 6 | 8 | 38 | 1 | 5.7 | 0.8 |
| **YJ** | 9 | 11 | 38 | 1 | 6.3 | 0.6 |
| **YK** | 6 | 8 | 32 | 1 | 5.0 | 0.8 |
| **ZE** | 9 | 3 | 37 | 1 | 6.1 | 0.7 |
| **ZR** | 6 | 2 | 35 | 1 | 5.6 | 0.5 |
| **ZW** | 6 | 2 | 36 | 1 | 5.0 | 0.8 |

**Table 17.5 Sensitivity test stimuli average $L_{Aeq}$ and annoyance ratings including 95% CI.**

### 17.5.3 Final Test, participant subgroup 1

The target stimuli parameters were 12 modulation depths (1, 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, 7, 9 and 12 dB(A)) at 4 $L_{Aeq}$ levels (30, 35, 40 and 45 dB(A)) for RFAM and MFAM with and without the presence of an unchanging level of garden noise within the masking signal. The resulting 192 stimuli were split into groups of 12 where modulation depth increased with stimuli number. The details of each group are shown in Table 17.6.

| Stimulus Number | Modulation Type | $L_{Aeq}$ , dB(A) | GN Present? |
|---|---|---|---|
| **1-12** | RFAM | 45 | No |
| **13-24** | RFAM | 40 | No |
| **25-36** | RFAM | 35 | No |
| **37-48** | RFAM | 30 | No |
| **49-60** | MFAM | 45 | No |
| **61-72** | MFAM | 40 | No |
| **73-84** | MFAM | 35 | No |
| **85-96** | MFAM | 30 | No |
| **97-108** | RFAM | 45 | Yes |
| **109-120** | RFAM | 40 | Yes |
| **121-132** | RFAM | 35 | Yes |

| | | | |
|---|---|---|---|
| **133-144** | RFAM | 30 | Yes |
| **145-156** | MFAM | 45 | Yes |
| **157-168** | MFAM | 40 | Yes |
| **169-180** | MFAM | 35 | Yes |
| **181-192** | MFAM | 30 | Yes |

**Table 17.6 Final Test, participant subgroup 1:Stimuli order and details.**

### 17.5.4 Final Test, participant subgroup 2

The stimuli consist of 8 modulation depths at 4 $L_{Aeq}$ levels for RFAM without the presence of garden noise within the masking signal. This gives 32 stimuli split into the following groups of 8 where modulation depth increases with stimuli number.

Two sets of these stimuli were produced. The first set replicates stimuli from Final Test I with a fixed overall $L_{Aeq.}$ The second set has a fixed masking level with the modulation signal being gradually increased giving an increase to the overall level. This was done for validation purposes so that it could be ensured that the nature of the stimuli did not change sufficiently to affect perception when MN levels were reduced. The results are presented in Section 20.

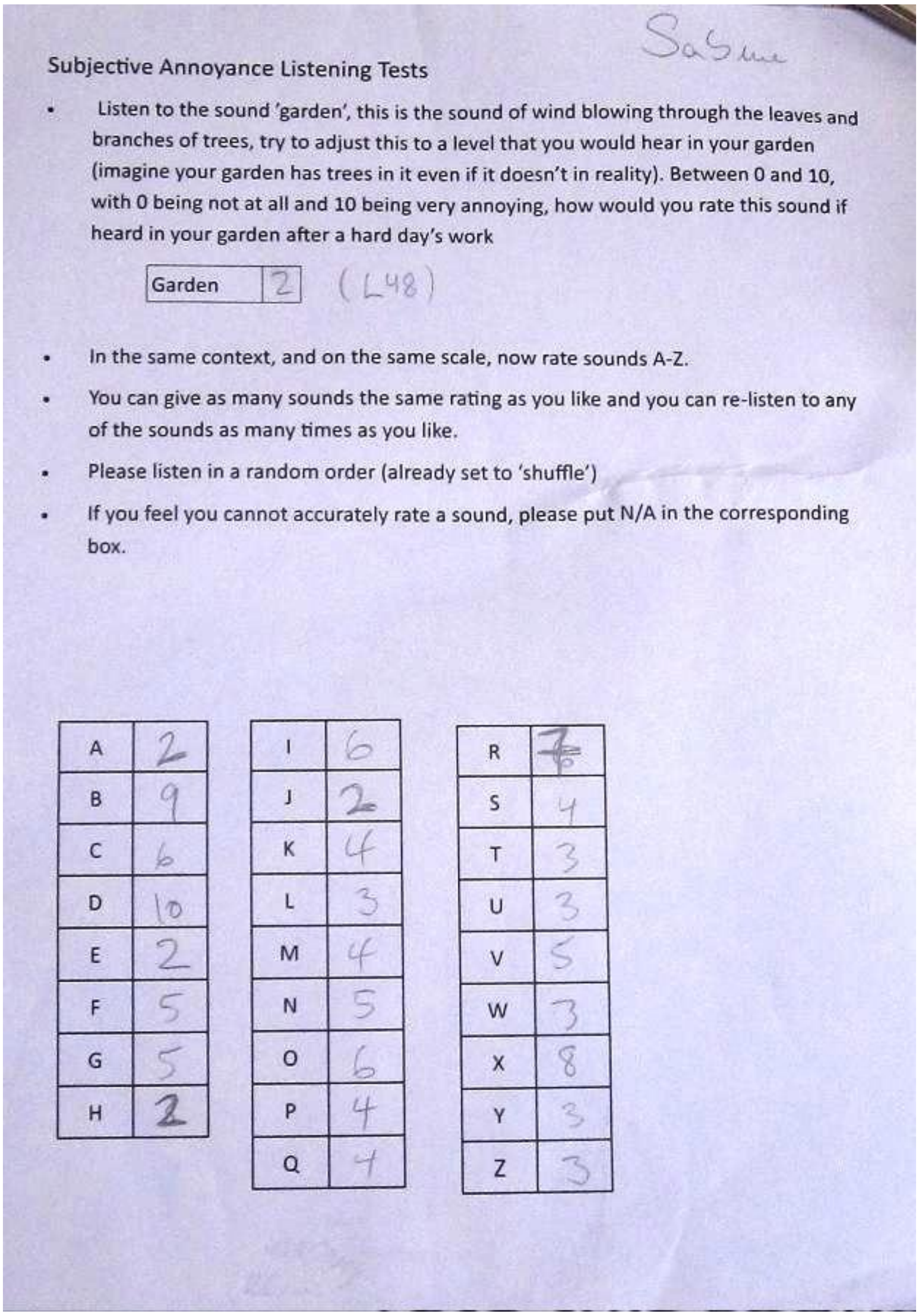
| Stimulus Number | Modulation Type | $L_{Aeq}$ dB(A) | GN Present? |
|---|---|---|---|
| **1-8** | RFAM | 40 | No |
| **9-16** | RFAM | 35 | No |
| **17-24** | RFAM | 30 | No |
| **25-32** | RFAM | 25 | No |

**Table 17.7 Final Test, participant subgroup 2:Stimuli order and details**

The target modulation depths are 0, 2, 3, 4, 5, 6, 9, 12 dB(A) and are derived from the average difference between the peak and trough in the A-weighted RMS of the stimuli.

# 18. Appendix VI: Participant instruction documents

## *18.1 Sensitivity Test I*



Subjective Annoyance Listening Tests

- Listen to the sound 'garden', this is the sound of wind blowing through the leaves and branches of trees, try to adjust this to a level that you would hear in your garden (imagine your garden has trees in it even if it doesn't in reality). Between 0 and 10, with 0 being not at all and 10 being very annoying, how would you rate this sound if heard in your garden after a hard day's work

| | |
|---|---|
| Garden | 2 |

(L48)

- In the same context, and on the same scale, now rate sounds A-Z.
- You can give as many sounds the same rating as you like and you can re-listen to any of the sounds as many times as you like.
- Please listen in a random order (already set to 'shuffle')
- If you feel you cannot accurately rate a sound, please put N/A in the corresponding box.

| | |
|---|---|
| A | 2 |
| B | 9 |
| C | 6 |
| D | 10 |
| E | 2 |
| F | 5 |
| G | 5 |
| H | 2 |

| | |
|---|---|
| I | 6 |
| J | 2 |
| K | 4 |
| L | 3 |
| M | 4 |
| N | 5 |
| O | 6 |
| P | 4 |
| Q | 4 |

| | |
|---|---|
| R | 7 |
| S | 4 |
| T | 3 |
| U | 3 |
| V | 5 |
| W | 3 |
| X | 8 |
| Y | 3 |
| Z | 3 |

**Figure 18.1 Test sheet for Sensitivity Test I**

## *18.2 Sensitivity Test II*

PS1 19/9/2011

**Subjective Annoyance Listening Tests**

- Listen to the sound 'garden', this is the sound of wind blowing through the leaves and branches of trees, try to adjust this to a level that you would hear in your garden (imagine your garden has trees and/or bushes in it, even if it doesn't in reality).
- Record the number that appears on the volume slider at your chosen level.

| | |
|---|---|
| Adjustment | 60% |

- Between 0 and 10, with 0 being not at all and 10 being very annoying, how would you rate this sound if heard in your garden after a hard day's work

| | |
|---|---|
| Garden | 4 |

- In the same context, and on the same scale, now rate sounds AC-ZW, press |<< and >>| to move through the playlist randomly.
- You can give as many sounds the same rating as you like and you can re-listen to any of the sounds as many times as you like. If you feel you cannot accurately rate a sound, please put N/A in the corresponding box.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AC | 8 | FB | 9 | NU | 5 | VT | ~~7~~ 9 |
| AV | 9 | FM | 9 | QD | 7 | VU | 10 |
| AX | 8 | FN | ~~8~~ 6 | QF | ~~8~~ 9 | VY | 7 |
| BU | 5 | GM | 7 | QS | 5 | WD | 8 |
| BY | 6 | GN | 7 | RG | 5 | WF | 7 |
| CR | 7 | HM | 5 | RH | 5 | WG | ~~8~~ 7 |
| CT | 6 | IL | 9 | RJ | 9 | XE | 4 |
| CY | 8 | JM | 4 | SB | ~~8~~ 5 | XR | 8 |
| DB | ~~8~~ 4 | JO | 8 | SC | 5 | XT | 7 |
| DN | 6 | KO | 9 | SV | 10 | YJ | 6 |
| DV | 7 | KP | 8 | TH | 6 | YK | 4 |
| EF | 6 | LP | 10 | TJ | 8 | ZE | 8 |
| EG | 3 | MI | ~~7~~ 8 | UK | 4 | ZR | 7 |
| EH | 6 | NI | 6 | UL | 7 | ZW | 6 |

**Figure 18.2 Test sheet Sensitivity Test II**

## 18.3 Final Test I and II

**Instructions**

Thank you for participating in this study. It is designed to investigate if fluctuations in Wind Turbine noise have an effect on how annoying the noise is.

Please sit facing the touch-screen, try not to move your head around too much, as the sounds you will hear are tailored to the centre point of the room.

Annoyance is an attribute that is heavily dependent on context. Therefore, we will be asking you to imagine you are hearing the sounds in specific situations. Try to imagine you are in your garden, relaxing and trying to enjoy your free time. The first sound you hear is the sound of wind in trees and bushes. This sound reflects how windy it is in your garden, so imagine there is moderate breeze.

Now, whilst you maintain this mental image, press 'Begin' and a control panel will appear. The 'Test Sound' button will be initially selected, this plays the reference sound for a minimum of 7 seconds. Listen to it until you can confidently rate how annoying this sound would be to hear in the imagined scenario. Use the slider on-screen to indicate this rating.

Now press the 'Reference Sound' button, this stops the test sound and begins the reference sound. Again, try to imagine how annoying this sound would be to hear in imagined scenario in your garden. Press the plus or minus buttons to change the level of part of the sound until it is of equal annoyance to the test sound. You can toggle between the two sounds until you are satisfied with your response, but please do not spend too long or think too hard about your answer as an initial answer is often the most natural. Press 'Next' when you are ready to move onto the next trial.

Please remember, your goal should always be to compare annoyance, if the sounds were heard at home, in your garden. Make sure that you are not simply adjusting the reference sound to equally loud as the test sound unless this also happens to be the point of equal annoyance.

**Figure 18.3 Instruction sheet for Final Test and Validation**

# 19. Appendix VII: Participant and observer comments

## 19.1 Participant observations

- Sounds were perceived by a number of participants to be very similar
- It was suggested for future experiments to use the descriptor 'Intrusive' which was one participant's interpretation of the term annoyance.
- It was recommended to introduce timescales together with the rating on how annoyed participants were; i.e. Not at all annoying = I could sit and listen to this all day; Very annoying = This would annoy me within a few minutes. Alternatively, a specific imagined time duration was thought to be useful
- Some participants felt that different answers could be given for the same stimulus on two different instances. – Possibly a lack of consistency in context elicitation.
- One participant mentioned that the swish was reminiscent of prenatal ultrasounds, and as such was a pleasant sound! Had to make sure they were clear as to what wind turbines were and that the context was in the garden. In the participant's individual results, this attitude is apparent as the increase in modulation led to a decrease in annoyance and equal annoyance level.

## 19.2 Observer comments

- 1 participant provided very polarised responses, either not at all annoying or very annoying. This highlights the problem that each participant's range of realistic answers is not quantified, an approximation maybe to use Z-scores, or divide by measured range? Also brings about the dilemma of whether to tell participants to use the full range (currently done), or let them do what comes naturally.
- 1 participant had very narrow range on annoyance slider, around 'not at all'. Also did not understand what annoyance meant, had to explain. Possibly language barrier.

# 20. Appendix VIII: Results of validation test 1

## *20.1 Rating distributions*

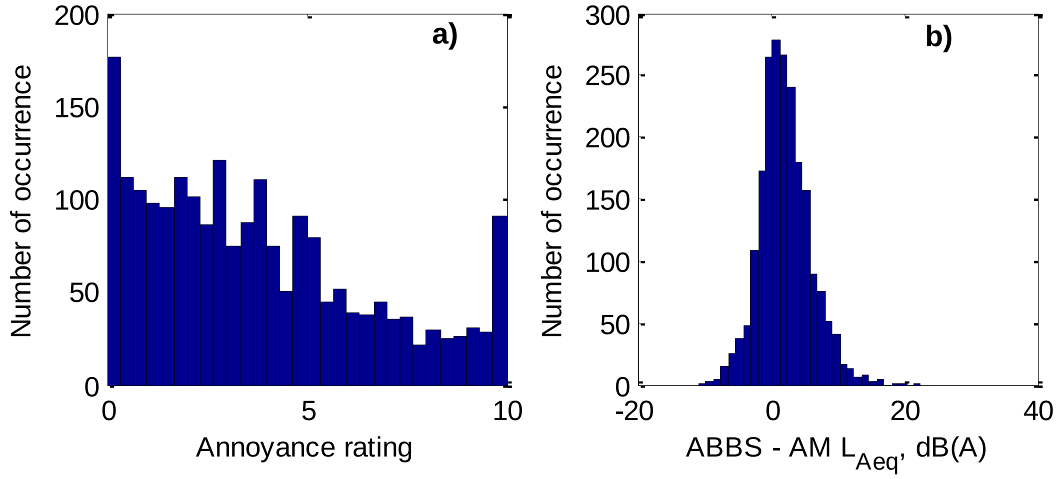



**Figure 20.1 Distribution of ratings a) for absolute annoyance scale, b) for the level difference between the ABBS and the AM stimuli with and without garden noise, mean = 2.2 dB(A), minimum rating = - 11 dB(A) .**

## *20.2 Absolute Ratings*

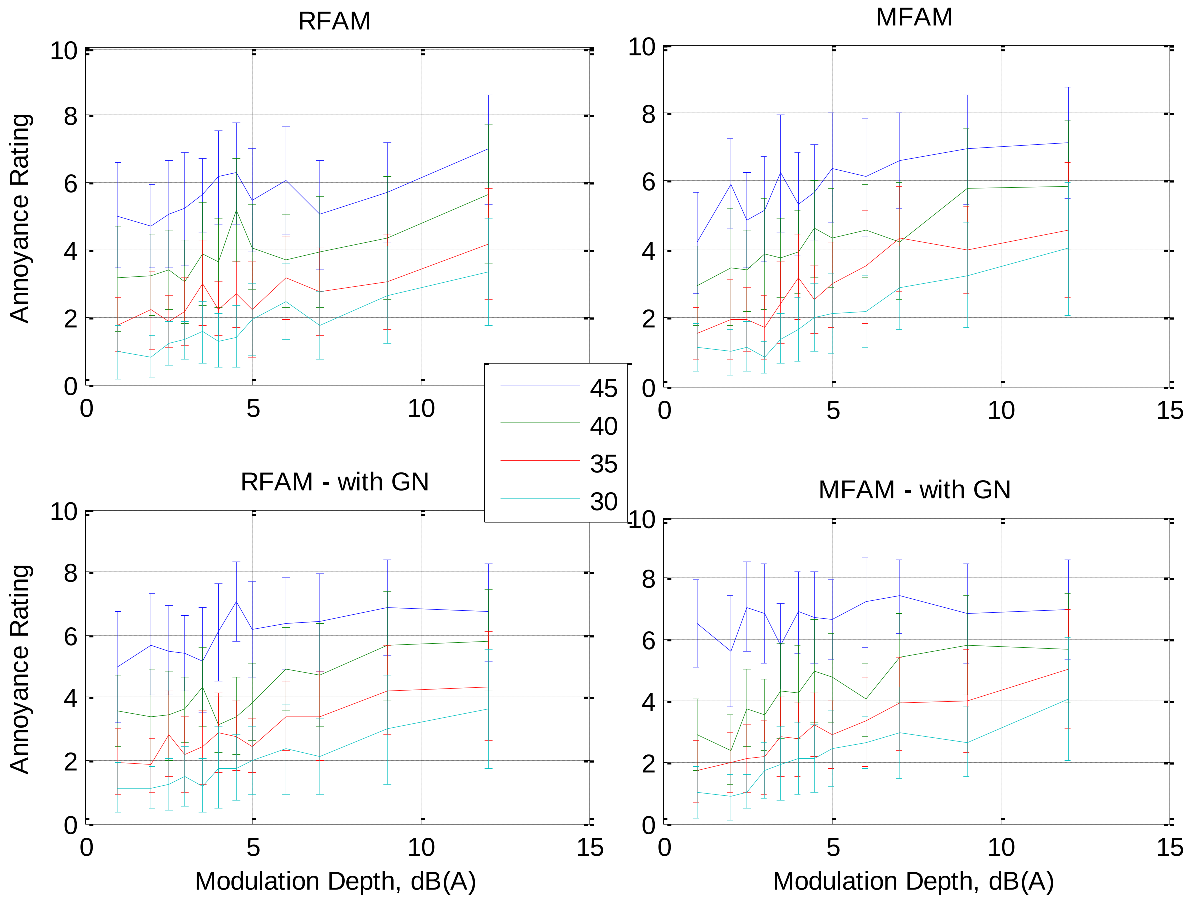



**Figure 20.2 Absolute annoyance ratings as a function of modulation depth for four different scenarios. The legend specifies the $L_{Aeq}$ of the test stimuli in dB(A).**

## *20.3 ABBS results*

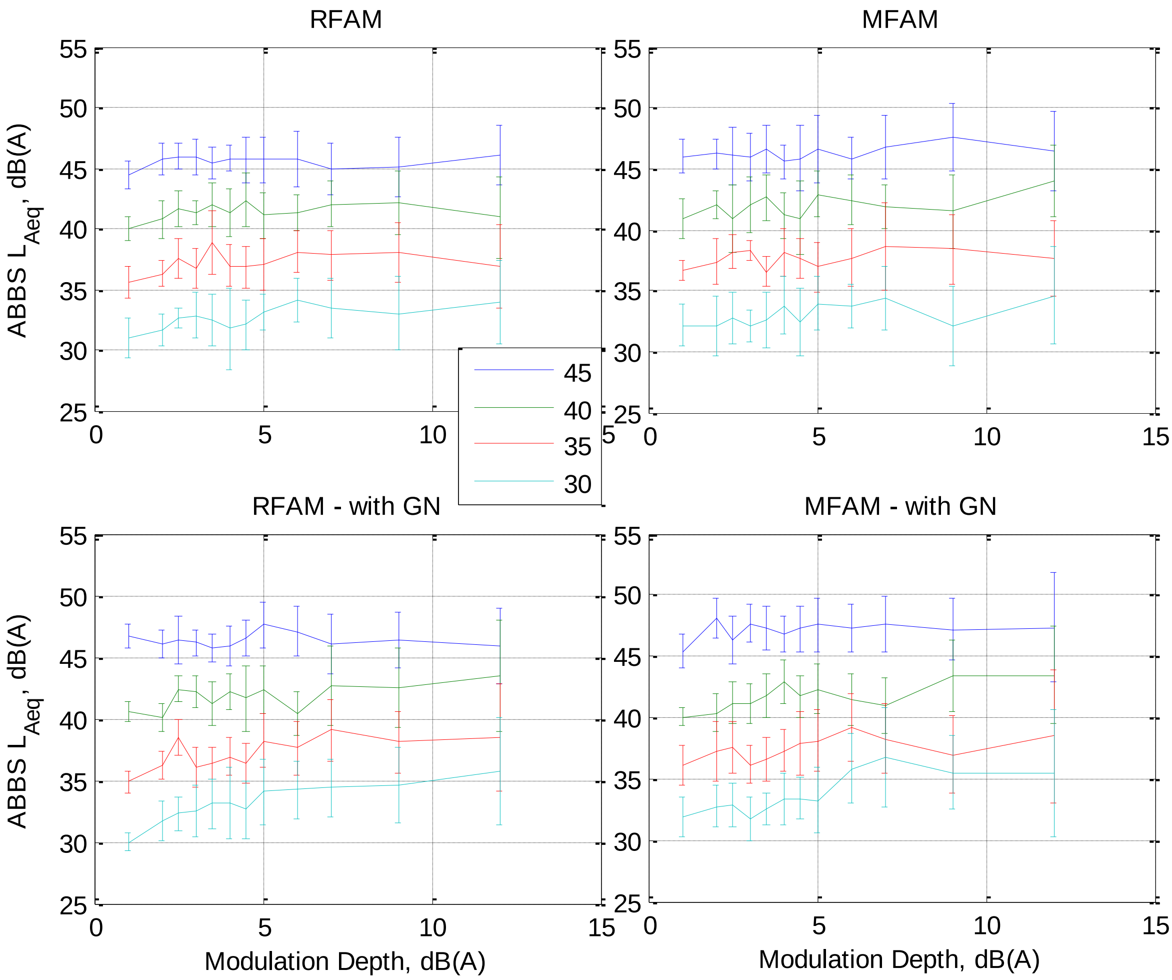



**Figure 20.3 Absolute ABBS $L_{Aeq}$ as a function of modulation depth for two types of modulation and in the presence and absence of masking garden noise.**

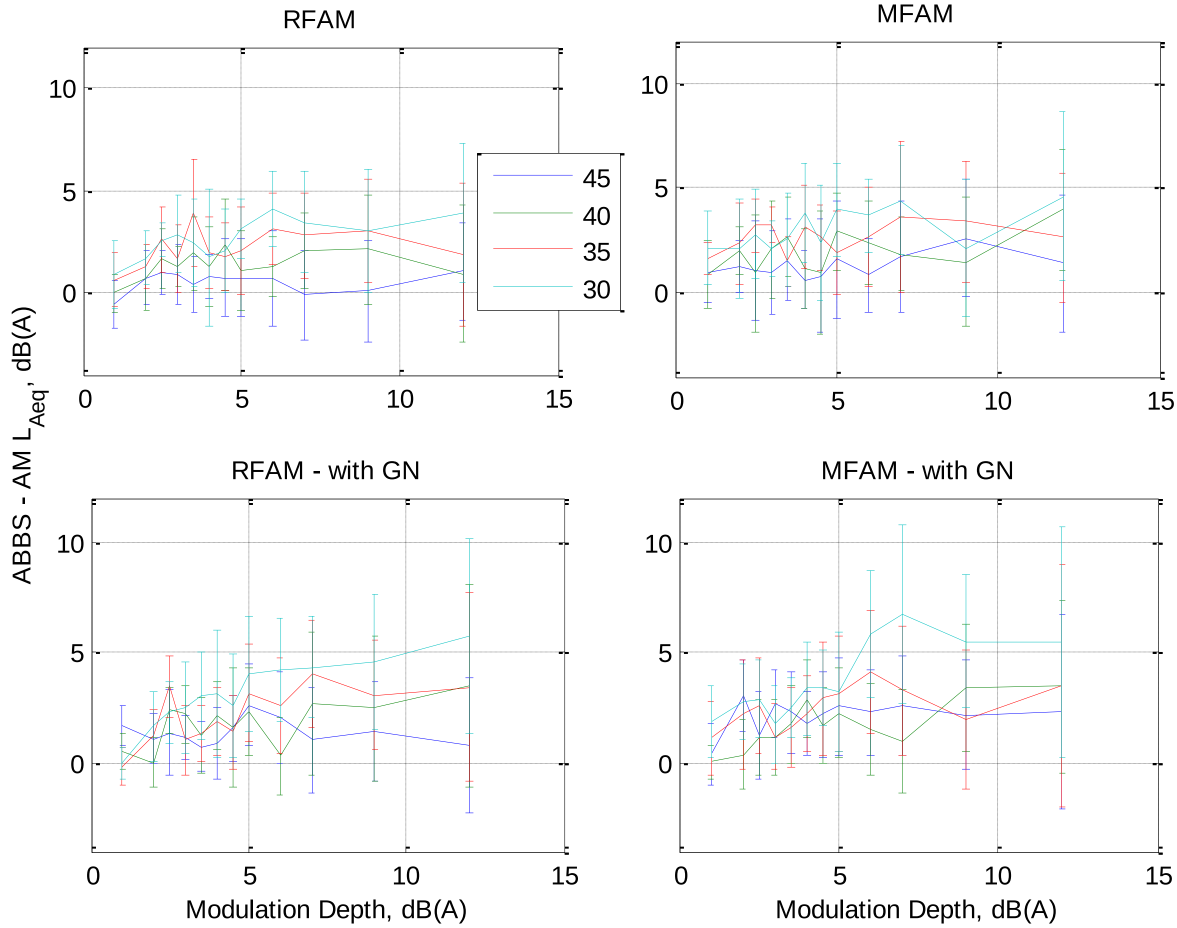



**Figure 20.4 ABBS -AM $L_{Aeq}$ as a function of modulation depth for two types of modulation (RFAM left, MFAM right) and in the presence (bottom) and absence (top) of masking garden noise. The legend specifies the $L_{Aeq}$ of the AM stimuli in dB(A).**

The general features of Figure 20.4 a) -d) are very similar within the 95 % CI. Using two types of AM is not likely to produce different responses of statistical significance whereas the presence/absence of garden noise might become significant for the quieter stimuli for a larger sample of participants.

# 21. Appendix IX: Results of Validation Test II

One concern with stimuli of constant $L_{Aeq}$ was that to keep the level constant with increasing MD the masking noise had to be reduced which might have led to changes in signal character. In Validation Test II a subgroup of nine participants therefore also rated a set of stimuli with the same MD levels and masking levels of 25, 30, 35 and 40 dB(A). They used the procedure of adjusting ABBS for this test.

## *21.1 Rating distributions*

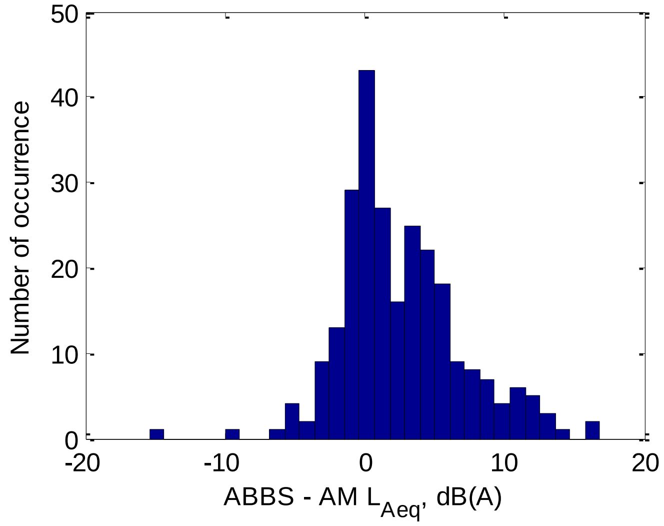



**Figure 21.1 Distribution of ratings for the level difference between the ABBS and the AM stimuli with and without garden noise, mean = 2.6 dB(A), minimum rating = - 15 dB(A). Absolute ratings were not recorded for this validation.**

## *21.2 ABBS results*

In general the rating behaviour of the participants shown in Figure 21.2 was very similar for the two sets of stimuli.

Figure 21.2 a) and c) show the absolute ABBS $L_{Aeq}$ for the constant stimulus $L_{Aeq}$ and the constant masking noise $L_{Aeq}$ stimuli, respectively. The monotonous increase of the rating with modulation depth is evidence that the participants adjusted the louder stimuli to slightly higher ABBS $L_{Aeq}$ compared to a).

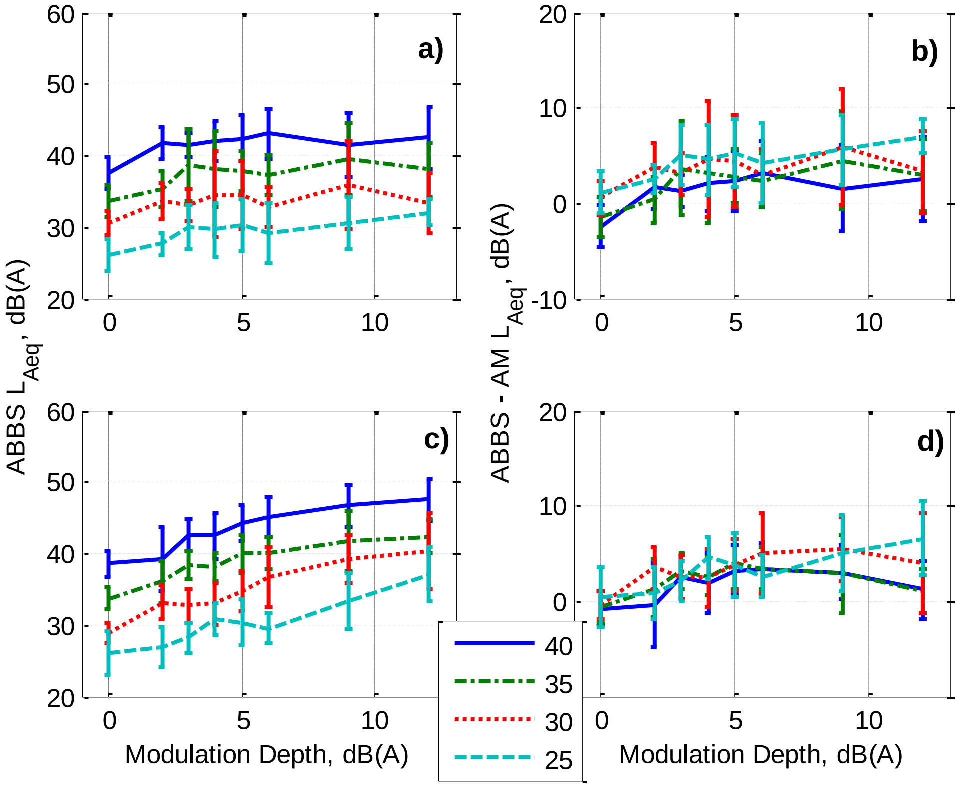



**Figure 21.2 Comparison of equal annoyance ratings between stimuli with constant overall $L_{Aeq}$ (top) and constant masking noise with increasing $L_{Aeq}$ as a function of MD (bottom). Note that the legend has therefore a different meaning for top and bottom graphs. Left panel: absolute ABBS $L_{Aeq}$, right panel: ABBS-AM $L_{Aeq}$**

Figure 21.2 b) and d) show the difference between ABBS and AM $L_{Aeq}$. Importantly, the increase of $L_{Aeq}$ due to the increasing MD was measured and the correct total stimulus $L_{Aeq}$ was used to calculate this difference. Therefore if the nature of the stimuli changed by reducing masking noise then a difference between b) and d) should be seen. However, within the 95 % confidence intervals b) and d) are very similar which suggests that by the nature of the stimuli is not changed significantly in terms of annoyance when reducing the masking noise to retain the total $L_{Aeq}$.

# 22. Appendix X: Annoyance ratings as a function of $L_{90}$

| $L_{Aeq}$ MD | $L_{Aeq}$ | $L_{A10}$ | $L_{A90}$ | $L_{Aeq}$-$L_{A90}$ | $L_{A10}$-$L_{A90}$ |
|---|---|---|---|---|---|
| 30dB(A)-0dB(A) | 29.3 | 29.7 | 28.8 | 0.5 | 0.9 |
| 30dB(A)-2dB(A) | 29.5 | 30.1 | 28.8 | 0.7 | 1.3 |
| 30dB(A)-3dB(A) | 29.5 | 30.5 | 28.5 | 1.0 | 2.0 |
| 30dB(A)-4dB(A) | 29.2 | 30.7 | 27.9 | 1.3 | 2.8 |
| 30dB(A)-5dB(A) | 29.6 | 31.6 | 27.7 | 1.9 | 3.9 |
| 30dB(A)-6dB(A) | 29.4 | 31.9 | 27.2 | 2.2 | 4.7 |
| 30dB(A)-9dB(A) | 29.8 | 33.1 | 26.0 | 3.8 | 7.1 |
| 30dB(A)-12dB(A) | 29.6 | 33.4 | 23.6 | 6.0 | 9.8 |

**Table 22.1 Measured $L_A$ metrics for nominal 30 dB(A) AM stimuli**

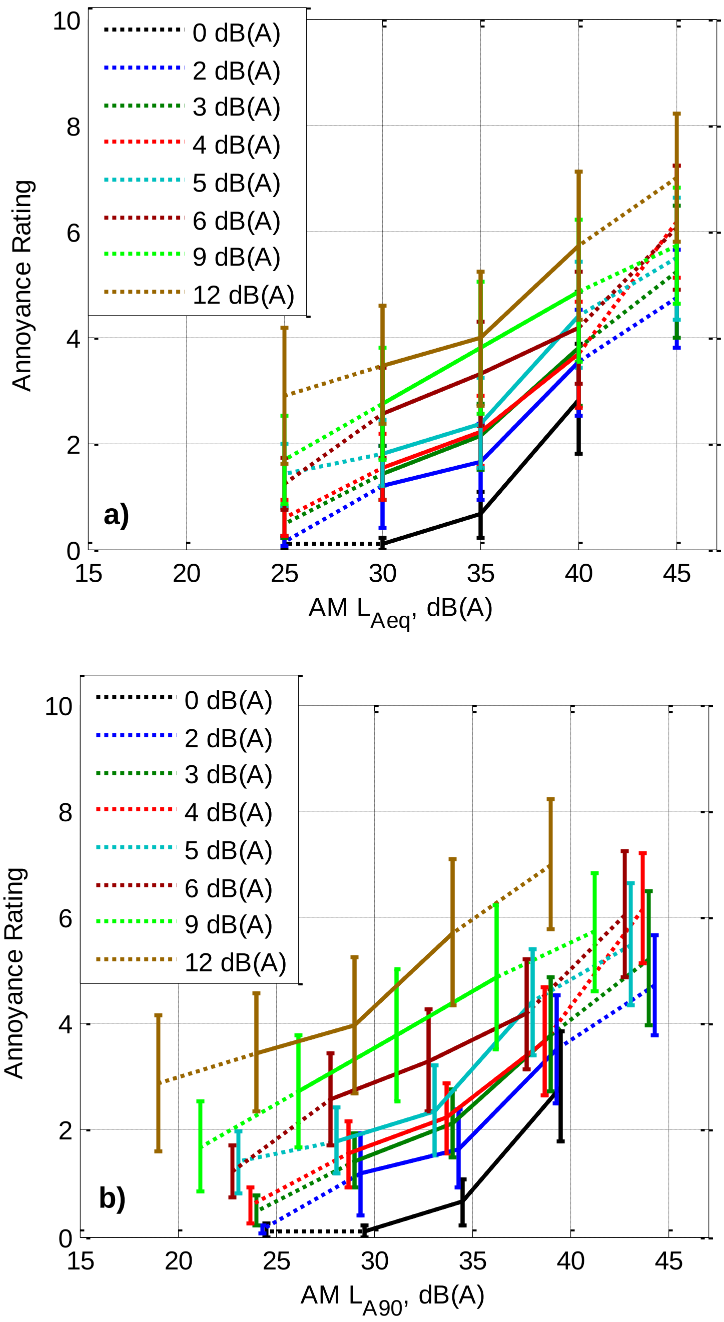



**Figure 22.1 Annoyance ratings as a function of L90 b) in comparison to ratings as a function of $L_{Aeq}$.a) (identical to Figure 9.2)**

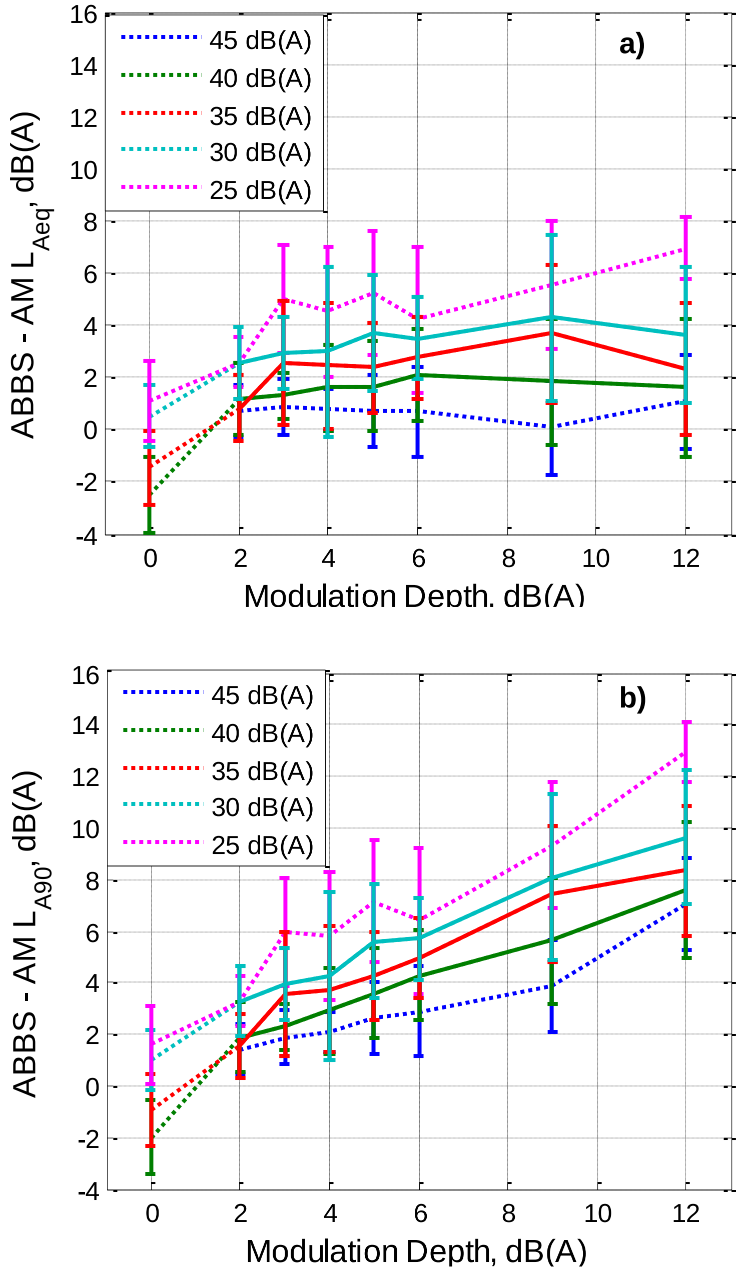



**Figure 22.2 ABBS ratings normalised by a) $L_{Aeq}$ (identical to Figure 9.5), b) $L_{A90}$**

# 23. Appendix XI: Data tables

## *23.1 Final Test composite figures*

| *Modulation Depth, dB(A)* | *RFAM L_Aeq* | *Annoyance Rating* | *95% CI* |
|---|---|---|---|
| **0** | 45 | n/a | n/a |
| **2** | 45 | 4.72 | 0.93 |
| **3** | 45 | 5.22 | 1.25 |
| **4** | 45 | 6.17 | 1.04 |
| **5** | 45 | 5.48 | 1.16 |
| **6** | 45 | 6.05 | 1.18 |
| **9** | 45 | 5.72 | 1.10 |
| **12** | 45 | 6.99 | 1.21 |
| **0** | 40 | 2.81 | 1.04 |
| **2** | 40 | 3.52 | 1.00 |
| **3** | 40 | 3.78 | 1.07 |
| **4** | 40 | 3.67 | 1.00 |
| **5** | 40 | 4.41 | 1.00 |
| **6** | 40 | 4.17 | 1.05 |
| **9** | 40 | 4.86 | 1.34 |
| **12** | 40 | 5.71 | 1.38 |
| **0** | 35 | 0.65 | 0.44 |
| **2** | 35 | 1.64 | 0.73 |
| **3** | 35 | 2.12 | 0.62 |
| **4** | 35 | 2.22 | 0.66 |
| **5** | 35 | 2.36 | 0.85 |
| **6** | 35 | 3.31 | 0.98 |
| **9** | 35 | 3.79 | 1.24 |
| **12** | 35 | 3.97 | 1.26 |
| **0** | 30 | 0.09 | 0.12 |
| **2** | 30 | 1.18 | 0.77 |
| **3** | 30 | 1.42 | 0.51 |
| **4** | 30 | 1.54 | 0.63 |
| **5** | 30 | 1.80 | 0.62 |
| **6** | 30 | 2.56 | 0.86 |
| **9** | 30 | 2.74 | 1.05 |
| **12** | 30 | 3.46 | 1.12 |
| **0** | 25 | 0.11 | 0.14 |
| **2** | 25 | 0.13 | 0.08 |
| **3** | 25 | 0.48 | 0.29 |
| **4** | 25 | 0.58 | 0.34 |
| **5** | 25 | 1.40 | 0.58 |
| **6** | 25 | 1.23 | 0.48 |
| **9** | 25 | 1.69 | 0.84 |
| **12** | 25 | 2.89 | 1.27 |

**Table 23.1 Annoyance rating data for Figure 9.2 and Figure 9.3. Note that the unmodulated stimulus was not included for $L_{Aeq}$ = 45 dB(A).**

| *Modulation Depth, dB(A)* | *RFAM L_Aeq* | *ABBS L_Aeq* | *95% CI* |
|---|---|---|---|
| **0** | 45 | n/a | n/a |
| **2** | 45 | 46 | 1.0 |
| **3** | 45 | 46 | 1.1 |
| **4** | 45 | 46 | 0.8 |
| **5** | 45 | 46 | 1.4 |
| **6** | 45 | 46 | 1.7 |
| **9** | 45 | 45 | 1.8 |
| **12** | 45 | 46 | 1.8 |
| **0** | 40 | 38 | 1.4 |
| **2** | 40 | 41 | 1.3 |
| **3** | 40 | 41 | 0.9 |
| **4** | 40 | 42 | 1.6 |
| **5** | 40 | 42 | 1.7 |
| **6** | 40 | 42 | 1.7 |
| **9** | 40 | 42 | 2.4 |
| **12** | 40 | 42 | 2.6 |
| **0** | 35 | 34 | 1.4 |
| **2** | 35 | 36 | 1.2 |
| **3** | 35 | 38 | 2.4 |
| **4** | 35 | 37 | 2.4 |
| **5** | 35 | 37 | 1.7 |
| **6** | 35 | 38 | 1.6 |
| **9** | 35 | 39 | 2.6 |
| **12** | 35 | 37 | 2.5 |
| **0** | 30 | 30 | 1.2 |
| **2** | 30 | 33 | 1.4 |
| **3** | 30 | 33 | 1.4 |
| **4** | 30 | 33 | 3.2 |
| **5** | 30 | 34 | 2.2 |
| **6** | 30 | 34 | 1.6 |
| **9** | 30 | 34 | 3.2 |
| **12** | 30 | 34 | 2.6 |
| **0** | 25 | 26 | 1.5 |
| **2** | 25 | 28 | 1.0 |
| **3** | 25 | 30 | 2.1 |
| **4** | 25 | 30 | 2.5 |
| **5** | 25 | 30 | 2.4 |
| **6** | 25 | 29 | 2.8 |
| **9** | 25 | 31 | 2.4 |
| **12** | 25 | 32 | 1.2 |

**Table 23.2 Data for Figure 9.4. - Note that the unmodulated stimulus was not included for $L_{Aeq}$ = 45 dB(A).**

| *Modulation Depth, dB(A)* | *RFAM L_Aeq* | *ABBS - RFAM L_Aeq* | *95% CI* |
|---|---|---|---|
| **0** | 45 | n/a | n/a |
| **2** | 45 | 0.7 | 1.0 |
| **3** | 45 | 0.9 | 1.1 |
| **4** | 45 | 0.8 | 0.8 |
| **5** | 45 | 0.7 | 1.4 |
| **6** | 45 | 0.7 | 1.7 |
| **9** | 45 | 0.1 | 1.8 |
| **12** | 45 | 1.1 | 1.8 |
| **0** | 40 | -2.5 | 1.4 |
| **2** | 40 | 1.2 | 1.3 |
| **3** | 40 | 1.3 | 0.9 |
| **4** | 40 | 1.6 | 1.6 |
| **5** | 40 | 1.7 | 1.7 |
| **6** | 40 | 2.1 | 1.7 |
| **9** | 40 | 1.8 | 2.4 |
| **12** | 40 | 1.6 | 2.6 |
| **0** | 35 | -1.4 | 1.4 |
| **2** | 35 | 0.8 | 1.2 |
| **3** | 35 | 2.6 | 2.4 |
| **4** | 35 | 2.4 | 2.4 |
| **5** | 35 | 2.4 | 1.7 |
| **6** | 35 | 2.7 | 1.6 |
| **9** | 35 | 3.7 | 2.6 |
| **12** | 35 | 2.3 | 2.5 |
| **0** | 30 | 0.5 | 1.2 |
| **2** | 30 | 2.6 | 1.4 |
| **3** | 30 | 3.0 | 1.4 |
| **4** | 30 | 3.0 | 3.2 |
| **5** | 30 | 3.7 | 2.2 |
| **6** | 30 | 3.5 | 1.6 |
| **9** | 30 | 4.3 | 3.2 |
| **12** | 30 | 3.6 | 2.6 |
| **0** | 25 | 1.1 | 1.5 |
| **2** | 25 | 2.6 | 1.0 |
| **3** | 25 | 5.0 | 2.1 |
| **4** | 25 | 4.5 | 2.5 |
| **5** | 25 | 5.3 | 2.4 |
| **6** | 25 | 4.2 | 2.8 |
| **9** | 25 | 5.5 | 2.4 |
| **12** | 25 | 6.9 | 1.2 |

**Table 23.3 Data for Figure 9.5 - Note that the unmodulated stimulus was not included for $L_{Aeq}$ = 45 dB(A).**