Effects of wind farm noise characteristics on subjective human responses

Mahmoud Ahmed Alamir

A thesis in fulfilment of the requirements for the degree of
Doctor of Philosophy

College of Science and Engineering

31st May 2022

Acknowledgements

First and foremost, profound gratitude goes to Almighty Allah, who guided and helped me perform this task.

I would like to thank Dr Kristy Hansen, Prof. Peter Catcheside and Dr Branko Zajamsek for their help and supervision. I would like to thank my Prof. Mark Taylor; whose expertise was invaluable to get this work done. I also acknowledge the financial support of Flinders University, the Australian Research Council (ARC), Project 180100022, and the Australian National Health and Medical Research Council (NHMRC), Project 1113571.

My deep appreciation goes to the students and staff members at the Adelaide Institute for Sleep Health. My thanks and deep appreciation are also conveyed to my friends, who made me excited to continue this research.

This thesis is dedicated to my family. My father's words, inspiring situations and advice always stick in my mind despite his absence. My mother and sister (Hanem and Eman) always provide me with great power, love and support. My elder brother (Ahmed) has always been an example in my academic and non-academic life. His helpful advice and time have always been exceptionally useful to this thesis. My brothers Mohammed and Ahmed and their families are also a source of pleasure and support to me. I am blessed to be a member of such a noble family. No words can express my love and gratitude to them. The last words go for Raneem, my wife, who became the light of my life and who has given me the extra strength and motivation to get things done.

Abstract

Wind energy is a clean, economically viable and environmentally friendly renewable energy source. However, questions have been raised concerning potential adverse health effects of wind farm noise (WFN). This thesis sought to determine the extent to which subjective responses to WFN are judged differently from road traffic noise (RTN) responses. It also investigated whether WFN characteristics such as tonal and broadband amplitude modulation (i.e., WFN level variability) and masking background noise (i.e., noise added to WFN), including long- and short-range RTN, pink noise and room criteria curve noise, affect sleep acceptability threshold levels (SATs). Relative SATs to a baseline stimulus representing the general spectrum of WFN for tonal and broadband amplitude-modulated (AM) WFN were also presented. This study included groups of participants of different ages and localities. Each group also had different self-reported noise sensitivity, allowing, for the first time, a comparison between group responses to WFN through listening tests. Participant groups were WFN-sensitive, WFN-insensitive, quiet rural area and RTN-sensitive participants defined on the basis of residential location and self-reported sensitivity to WFN or RTN.

WFN and RTN with different characteristics, including tonal AM WFN, were presented at six levels to examine dose-response relationships with annoyance, loudness and potential to disturb sleep. WFN with tonal amplitude modulation was perceived as more annoying and yet less loud than RTN.

Tonal AM WFN affected absolute and relative SATs. Increasing modulation depth decreased absolute and relative SATs (i.e., the noise was less acceptable for sleep). The older group and WFN-highly sensitive participants had the lowest SATs with no difference in relative SATs. Interestingly, the road-traffic-noise-sensitive group simultaneously showed the lowest relative SATs and the highest absolute SATs. Low relative SATs of 7.5 ± 2.1 dBA (mean ± CIs) were

obtained for less sleepy participants at a modulation depth and tonal audibility of 15 dBA and a modulated frequency of 110 Hz.

Absolute and relative SATs with broadband AM WFN were also examined. The increase of modulation depth of broadband AM WFN decreased absolute and relative SATs. Relative SATs started at 0 for a modulation depth of 2 dBA and increased gradually and reached relative SATs of 6.8 ± 1.8 dBA (mean ± CIs) at a modulation depth of 10 dBA and a baseline SAT range below 30 dBA.

Different types of masking background noise also affected absolute SATs of WFN (F (3, 422) = 21.5, $p < 0.001$). Short-range RTN masking increased WFN sleep acceptability by (mean ± CIs) 3.0 ± 1.5 dBA for noise-sensitive groups. Unlike other groups, RTN-sensitive participants found WFN more acceptable for sleep when mixed with short-range RTN at 47 ± 1.5 dBA.

The effects of WFN acoustic characteristics and related human factors on sleep acceptability are important considerations for future WFN guideline developments. The outcomes of this thesis work provide important new data towards guiding evidence-based improvements to noise guidelines and penalties for WFN characteristics and choosing suitable sites for wind farms based on pre-existing ambient noise.

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Signed: ---------------------------------

Mahmoud Ahmed Alamir

Date: 1 / 7 / 2021

Table of contents

List of figures

List of tables

List of appendices

List of abbreviations

AM	Amplitude modulation;
ATLs	Acceptability thresholds levels;
BGN	Background noise;
HTLs	Hearing thresholds levels;
LF	Low frequency;
LFN	Low-frequency noise;
MD	Modulation depth;
MI	Modulation index;
NAM	Normal amplitude modulation;
OAM	Other amplitude modulation;
RTN	Road traffic noise;
SATs	Sleep acceptability threshold levels
SNR	Signal to noise ratio;
SPLs	Sound pressure levels;
WHO	the World Health Organization;
WTs	Wind turbines;
WFN	Wind farm noise.

List of publications

Journal Articles:

1. Chapter 2: Alamir, M. A., Hansen, K. L., & Catcheside, P. (2021). Penalties applied to wind farm noise: Current allowable limits, influencing factors, and their development. Journal of Cleaner Production, 126393. https://doi.org/10.1016/j.jclepro.2021.126393
2. Chapter 2: Alamir, M. A., Hansen, K. L., Zajamsek, B., & Catcheside, P. (2019). Subjective responses to wind farm noise: A review of laboratory listening test methods. Renewable and Sustainable Energy Reviews, 114, 109317. https://doi.org/10.1016/j.rser.2019.109317

Conferences:

1. Chapter 7: Alamir, M., Zajamsek, B., & Hansen, K. (2020). Perceived sleep acceptability of broadband amplitude-modulated wind farm noise, Inter-noise 2020, E-conference, Seoul, South Korea.
2. Chapter 4: Alamir, M., Catcheside, P., Hansen, K., Zajamsek, B., & Micic, G. (2019). Human response to wind farm noise compared to road traffic noise based on focused listening tests, Internoise 2019, Madrid, Spain.
3. Chapter 2: Alamir, M. A., Hansen, K. L., & Zajamsek, B. (2018). The effect of wind farm noise on human response: An analysis of listening test methodologies, Proceedings of Acoustics 2018, Adelaide, Australia.

Papers in preparation at the time of submission:

1. Chapter 3: A listening test design to measure acceptability for sleep: A novel methodology.
2. Chapter 4: Annoyance, sleep disturbance and loudness of wind farm noise compared to road traffic noise – A laboratory study.
3. Chapter 5: Sleep acceptability of tonal amplitude-modulated wind farm noise.
4. Chapter 6: Sleep acceptability of broadband amplitude modulated wind farm noise.
5. Chapter 7: Sleep acceptability of wind farm noise in the presence of different masking background noise types.

Chapter 1 Introduction

1.1. Introduction

The world has expanded its use of renewable energy due to increasingly stringent emission targets and more favourable economic factors promoting investment in renewable more than traditional energy sources. One of the most viable renewable energy sources is wind energy, which can be clean, economically profitable and environmentally friendly (Kaldellis et al., 2011). However, questions have been raised regarding the potential adverse health effects of wind farm noise (WFN) (Schmidt et al., 2014).

The use of wind energy as a renewable source of energy is expanding rapidly. For example, the global wind power capacity was 743 GW (59.25% onshore and 4.75% offshore) in 2020 with a growth of 14% compared to 2019, while it was 24 MW in 2001 (Global wind energy production, 2021). Onshore wind power capacity in 2020 was the largest in China, the US, Germany, India and Spain respectively (Global wind energy production, 2021). These countries produce 74% of the world's total onshore wind energy. Australia was among the top ten countries in expanding installations of wind power onshore in 2020 (*Global wind energy production*, 2021).

The increase in wind turbine size increases wind energy production with a maximum efficiency of 58.3% (Betz, 1966). During the last three decades, the size of wind turbines, in terms of blade length, has increased 10-12 times, with a rotor diameter in modern wind turbines ranging from 17 to 90 m, and power output has increased 30-40 times, reaching 5 MW (Leung and Yang, 2012). It is predicted that power output will be 10-20 MW in the near future (Islam, Mekhilef and Saidur, 2013).

Along with the increase in new facilities and installed power capacities of wind farms, WFN emissions are also likely to increase. Søndergaard (2013) presented a prediction model of SPLs depending on the power output for modern wind farms in Denmark. At 1.00 MW, the emitted A-weighted SPL predicted by this model at a distance between 304 and 810 m was over 100 dBA, and the infrasound level was around 90 dBA. WFN emissions can also be influenced by many factors such as wind farm power output, distance from the nearest turbine, design of the blades, local topography and atmospheric conditions (Micic et al., 2018).

Noise attenuation over distance and through building structures is lower with WFN compared to road traffic noise (RTN) due to the more predominant low-frequency content of the noise. Attenuation during propagation can occur due to many factors (e.g. air attenuation, ground absorption, meteorological effects, and reflection). At low frequencies, air attenuation of sounds is small and other attenuating factors such as absorption by the ground, barrier shielding, and refraction can also be negligible (Hubbard and Shepherd, 1991). For example, Leventhall (2003) found an attenuation of only 0.1 dB/km at 63 Hz, compared to 1.1 dB/km at 250 Hz, supporting that low-frequency noise (LFN) levels are minimally affected by large propagation distances. WFN has a higher low-frequency content than RTN (Hansen et al., 2015; Schäffer et al., 2016), and thus shows lower propagation attenuation, helping to explain why WFN can be audible indoors at distances of 3.5 km from WTs (*ISO 389-7:2005*).

The noise level of WFN can also vary significantly over time, making it more noticeable and intrusive compared to other noise types. Figure 1 shows an example of an amplitude-modulated signal with an amplitude varying over time. This variation often occurs due to the rotation of the blades at the blade pass frequency and is referred to as amplitude modulation (AM), which can be a significant contributor to perceived annoyance (Lee et al., 2011; Virjonen et al., 2019). WFN also contains tonal amplitude-modulated components, which could significantly affect the human response to WFN.

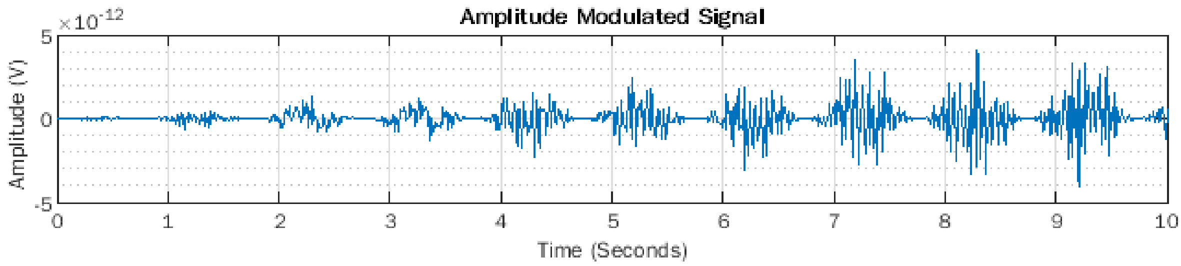


Figure 1.1. Amplitude modulated signal as a demonstration of wind farm noise amplitude modulation. The signal was generated through MATLAB with a modulation frequency of 1 Hz and a sinusoidal modulating function.

As with other noise types, WFN has the potential to cause annoyance, disturb sleep and impact health and well-being (Schmidt and Klokker, 2014). These potential effects, along with continuing wind energy developments, highlight the need for acceptable threshold levels (ATLs) along with adequate noise limits and penalties to help ensure that WFN emissions remain within reasonable limits.

At distances greater than approximately 2 km from a wind farm, WFN has special characteristics, including low-frequency (LF) and infrasound components and tonal amplitude modulation (AM) (Hansen et al., 2019). At close distances from wind farms, WFN can contain broadband AM components often referred to as "swish noise" (Hünerbein et al., 2013). These WFN acoustic characteristics appear to increase perceived annoyance, even at lower sound pressure levels (SPLs) compared to other environmental noise sources (Janssen et al., 2011).

Besides WFN acoustic characteristics, non-acoustic factors such as gender, noise sensitivity and visual effects can also affect perceived responses (Virjonen et al., 2019). However, identifying the level of contribution of these factors compared to acoustic characteristics on human responses through laboratory-controlled studies remains important to help guide the need for and design of potential future improvements to noise guidelines and mitigation strategies.

1.2. Current evidence of WFN effects on humans

In addition to annoyance, other symptoms such as stress and sleep disruption may be attributed to WFN (Pohl et al., 2018). It has been reported that most complaints about WFN occur during the nighttime (van den Berg, 2005). This could be attributed to a higher signal-to-noise ratio between WFN and background noise compared to the daytime (van den Berg, 2005).

Further work is required to establish whether additional penalties are necessary, especially for some WFN characteristics such as tonal AM and for other responses such as acceptability for sleep. The impact of WFN on sleep has been the subject of considerable debate. Some studies have reported that WFN has no significant effects on sleep (Jalali et al., 2017; Michaud et al., 2013), while another recent laboratory study has shown that there could be effects on sleep in terms of prolonged REM sleep latency, reduced REM sleep and reduced self-assessed sleep quality (Smith et al., 2020). Given known SPL-dependent sleep disruption effects from other noise types, such as road traffic noise (Basner et al., 2011) a clear need remains for further studies on potential effects of WFN on sleep.

Although there is evidence that acoustic characteristics of WFN can cause high annoyance, the degree to which other confounding factors affect annoyance is largely unknown (Council of Canadian Academies, 2015). Confounding factors include sensitisation of participants, attitudes to wind turbines (WTs), presence versus absence of economic benefits, dwelling place (i.e. rural or urban areas) and WT visibility (Janssen et al., 2011; Pedersen et al., 2009). Controlled listening tests are particularly useful for examining the effects of acoustic versus non-acoustic factors more reliably than in field surveys, which may be confounded by multiple factors not possible to control in real-world environmental exposure settings (Schäffer et al., 2018).

1.3. Aims

The work presented in this thesis aimed to:

1. Examine the effects of WFN and its characteristics on subjective human responses, including annoyance, loudness and potential to disturb sleep compared to road traffic noise (RTN) in Chapter 4. Hypothesis: The type of WFN and SPLs have different effects on human perception as compared to RTN.
2. Obtain absolute and relative sleep acceptability thresholds of AM WFN acoustic characteristics (i.e., LF tonal AM and swish noise) in Chapters 5 and 6. Hypothesis: Tonal and broadband AM WFN and its parameters affect sleep acceptability differently.
3. Examine the effects of different background masking noise types and signal-to-noise ratios on the absolute and relative sleep acceptability in Chapter 7. Hypothesis: Masking background noise type and its ratio with WFN can affect sleep acceptability of WFN.
4. Investigate the effects of non-acoustic factors on subjective sleep acceptability of noise in participants exposed to WFN in Chapters 4, 5, 6 and 7. Hypothesis: Non-acoustic factors such as gender and noise sensitivity contribute to perceived responses in addition to specific WFN acoustic characteristics.
5. Order WFN acoustic and non-acoustic factors in terms of their relative contribution to sleep acceptability based on the best performing machine learning models in Chapters 5, 6 and 7. Hypothesis: Non-acoustic factors contribute to perceived responses differently compared to WFN acoustic characteristics.

1.4. The contribution of the thesis to knowledge

The work presented in this thesis is the first to systematically investigate the effects of WFN characteristics on sleep acceptability based on listening tests. This approach more specifically tests if sleep acceptability, which may be judged differently from annoyance. The relative contribution of some non-acoustic factors to perceived responses is also presented. This information helps determine which factors may be the optimal targets for potential future noise mitigation strategies.

Knowledge regarding the effects of WFN and its characteristics on human responses is sparse compared to other noise sources (Council of Canadian Academies, 2015). Besides WFN perceived annoyance and sleep disturbance, acceptability for sleep could be an essential response to establish the responses to WFN. Noise context, instructions and subjective expectations affect responses. For example, Moorhouse et al. (2005), through listening tests, reported that the acceptability threshold levels of real sounds at nighttime could be lower than their acceptability threshold levels during the daytime. Inukai et al. (2000) showed that different situations led to annoyance differences. This indicates the importance of measuring acceptability for sleep, which could help establish subjective responses to WFN and its acoustic characteristics.

Sleep acceptability threshold levels of WFN and its characteristics were, for the first time, obtained to address inconsistencies surrounding night-time responses due to WFN. As discussed in Chapter 2, WFN allowable limits are inconsistent, and a difference of 25 dBA can be found between guidelines. Comparing sleep acceptability results to current guidelines provides further information on whether current penalties are adequate to protect the amenity of people residing near wind farms. Human responses to some WFN characteristics such as LF tonal AM and masking background noise were also studied for the first time. Therefore, the work presented in this thesis makes several noteworthy contributions to improving the fundamental understanding of human response to WFN.

The primary study methodologies involved laboratory listening tests, which provide a controlled environment to examine the perceptual effect of non-acoustic factors on SATs and their relative contribution compared to acoustic characteristics. Understanding these parameters' relative contributions is important to target the potential main factors contributing to increased negative responses in future treatment strategies.

The location of wind farms could affect perceived responses (in terms of geographical properties, features of airflow and noise propagation and background noise). For example, Pedersen et al. (2010) showed that sites exposed to RTN are a good choice for wind farms with moderate SPLs between (35-40 dBA L_{den}) based on a survey in the Netherlands in 2007, especially when RTN is 20 dB higher than WFN. Therefore, acoustic characteristics of masking background noise and its level have the potential to affect perceived human response of WFN. The current work presented in Chapter 7 provided WFN SATs using different masking background noise types at different signal-to-noise ratios (SNRs) for the first time. This data could improve understanding of the effects of masking background noise in a more controlled environment. Identifying the type of masking background noise and its level that make the perception of WFN less annoying can be beneficial for pre-construction planning of wind farms.

Chapter 2 Literature review[1]

Laboratory listening tests are a fundamental part of identifying and quantifying human perceptual responses to noise, including wind farm noise (WFN); a low-frequency environmental noise with time-varying components with the potential to impact more people as wind farm facilities continue to expand worldwide. Design characteristics of WFN listening tests vary between studies. This likely impacts WFN listening test results and makes quantitative comparisons difficult between studies. Accordingly, this review presents the available literature regarding WFN listening test methods, their overall characteristics and potentially important differences in noise stimuli and rating methods used. Key design variables explored include participant selection, stimuli duration, signal synthesis methods, noise reproduction methods and listening room characteristics. Listening test results from studies that have investigated the perceptual effects of various WFN components such as low-frequency noise and infrasound, tonality and amplitude modulation are presented. The impact on listening tests of factors unrelated to noise such as sensitivity, attitudes/beliefs and visual effects are also explored. It is shown that some WFN characteristics have received limited attention to date in listening tests. These include broadband low-frequency noise, tonal noise, tonal amplitude modulation and amplitude modulation parameters such as modulation frequency and intermittency. The relative importance of acoustic and non-acoustic factors to human perception is also largely unknown and requires well-designed listening tests to help elucidate.

[1] This chapter was published in two journal articles:
a) Alamir, M. A., Hansen, K. L., Zajamsek, B., & Catcheside, P. (2019). Subjective responses to wind farm noise: A review of laboratory listening test methods. Renewable and Sustainable Energy Reviews, 114, 109317. https://doi.org/10.1016/j.rser.2019.109317
b) Alamir, M. A., Hansen, K. L., & Catcheside, P. (2021). Penalties applied to wind farm noise: Current allowable limits, influencing factors, and their development. Journal of Cleaner Production, 295, 126393. https://doi.org/10.1016/j.jclepro.2021.126393

2.1. Introduction

Questions have been raised about the potential effect of wind farm noise on human health (Caddick et al., 2018). WFN has unique characteristics such as the dominance of low-frequency noise (LFN) and infrasonic components that have been shown to be more annoying than medium to high-frequency components at the same A-weighted sound pressure levels (SPLs) (Leventhall, 2003; Schmidt et al., 2014). This low frequency (LF) content could be particularly problematic due to possible issues with the A-weighting for low-frequency noise (LFN) (Kelley, 1987), significant variability in LF hearing threshold levels among individuals, and compression of equal loudness contours at LFs (Moller, 1987). Moreover, WFN can contain potentially annoying components such as amplitude modulation (Ioannidou et al., 2016; Schäffer et al., 2016) and tonality (L. S. Søndergaard et al., 2013; S. Yokoyama et al., 2016).

Subjective human responses to WFN are commonly assessed using surveys and laboratory studies. Surveys provide more ecologically valid results as they reflect the actual conditions and impressions of long-term exposure, which cannot be provided by a laboratory study (Moorhouse et al., 2005; Schäffer et al., 2016, 2018). However, many confounding factors cannot be controlled through survey studies, such as problems associated with recall, sensitisation of participants, attitudes to wind turbines (WTs), the presence of variable economic benefits, uncertainty in propagation modelling, variation in acoustic parameters through variation in environmental conditions and other influencing factors, dwelling location (i.e. rural or urban areas) and wind turbine visibility, all of which have the potential to influence perceptual outcomes (Janssen et al., 2011; Pedersen et al., 2009).

Many of the factors affecting subjective annoyance ratings during field measurements can be minimised through laboratory listening tests. These provide a substantially more controlled environment, allowing for more control over stimulus characteristics and the ability to identify more specific effects of key WFN acoustic characteristics (i.e. LFN and/or infrasound, broadband and tonal amplitude modulation (AM), and tonality) on subjective responses. However, many factors may affect the reliability of WFN listening tests, such as participant selection methods, the type of stimulus exposure and the method used to assess annoyance ratings.

Recent research has focused on isolating and quantifying WFN characteristics and their contribution to annoyance ratings (Ioannidou et al., 2016; Schäffer et al., 2016). Along with this growth in WFN annoyance assessment, there has been increasing concern over the suitability of current allowable noise limits and penalties for WFN and its characteristics (Hansen et al., 2020). Further work is required to better establish whether additional penalties are necessary, on which noise features these should be applied, and which LFN and infrasound weighting strategies (e.g., the A-, C- or G-weightings) may be the most appropriate for WFN and its characteristics.

This review focuses on relevant factors drawn from published WFN listening test studies and provides recommendations towards more robust, comprehensive and standardised methods relevant to WFN listening tests. Methods for quantifying annoyance due to key WFN acoustic characteristics identified through listening tests are also reviewed.

2.2. Design characteristics of WFN listening tests

Different design characteristics of WFN listening tests may be needed depending on the specific study purpose. However, this makes quantitative comparison between studies difficult. Therefore, this section describes key relevant study design parameters of WFN listening tests such as participant selection, stimuli duration, stimuli exposure and rating methods.

2.2.1. Participant selection

Annoyance ratings are subjective and may be strongly influenced by the characteristics of study participants selected for investigation. For example, although hearing thresholds levels (HTLs) and annoyance are expected to vary between participants (Dittrich et al., 2009; Moorhouse et al., 2005; Sakamoto et al., 2014), all participants should have normal hearing unless the purpose of a study is to investigate hearing loss (e.g. reduced high-frequency hearing acuity) effects on lower frequency WFN perception or responses in participants self-reporting annoyance irrespective of hearing deficits. Normal hearing can be verified by comparing the HTL of study participants to the reference equivalent threshold using an audiometric test (Frank, 1997) or questionnaires (Schäffer et al., 2016). However, it is worth noting that conventional audiology tests rarely cover hearing acuity below 250 Hz, falling well short of a potentially important frequency range to consider in the context of WFN.

The number of participants included in published WFN listening tests varies between 14 and 97 as shown in Tables 2.2 and 2.3. Depending on the study's purpose, small sample sizes can be sufficient, but given expected between-participant variability and potential biases associated with sample selection methods, results from relatively small studies may not be representative or generalisable to the broader population. Thus, carefully selected and larger samples may be necessary for reliable population estimates of annoyance rating distributions and to minimise the potential for Type II statistical error risks in studies comparing study interventions or between-group effects.

When selecting the sample population, other factors may also be important to take into account, including age, sex and exposure history to WFN (Moorhouse et al., 2005; Perkins et al., 2015). Personality factors, attitudes, beliefs and potential sensitisation to WFN may also importantly influence overall annoyance ratings (Leventhall, 2010; Maffei et al., 2015; Moorhouse et al., 2005). Crichton and Petrie (2015) investigated the nocebo effect of participants exposed to infrasound and audible WFN by comparing how the response to WFN varied depending on whether the participants had been shown material demonstrating the nocebo effect or material demonstrating some pathophysiological theories of symptoms. The authors found that the reported symptoms from WFN considerably reduced when nocebo effects were demonstrated to participants. Tonin et al. (2016) also investigated the effects of WT infrasound and the placebo/nocebo effect and found that people with prior concerns regarding infrasound reported more WFN symptoms than people without prior concerns. However, the results reported by Crichton and Petrie (2015) and Tonin et al. (2016) may have been influenced by the acoustic stimuli design, as discussed in Section 2.2.2.

Prior exposure, anticipatory and attitudinal effects may also be important. Maffei et al. (2015) compared auditory recognition of WFN using an equal number of people living near wind farms and people who had never been exposed to WFN (N=20 per group). The two groups responded differently to WFN exposure at large distances from WTs, where people with prior WFN exposure tended to falsely detect noise unrelated to WTs as WFN.

WFN studies often include people living near wind farms to control for potential effects of exposure history to WFN, in terms of long-term exposure and sensitisation, on participant responses. However, most previous studies have not considered inter-individual variation or potential interaction effects involving factors such as age, sex, and exposure history to WTs. For example, some studies have not included people living near wind farms (Lee et al., 2011), while others have specifically tried to include subjects with high sensitivity to LFN, albeit with small participant numbers (N=3) (Moorhouse et al., 2005) and typically without specifically testing for between-group effects, which may also be underpowered. Schäffer et al. (2016) conducted the most wide-ranging laboratory study to date of WFN using a relatively large sample size of 60 (29 females and 31 males) of different ages, with participants selected from locations with various background SPLs and types. However, no subjects living near WTs were involved. Moorhouse et al. (2005) included only three subjects with high sensitivity to LFN such that it remains unclear if numerically higher acceptability levels meaningfully differed from the remaining participants. Szychowska et al. (2018) studied participants who had previously experienced seeing and/or hearing WTs compared to WT naïve participants. However, annoyance responses were not presented or compared between groups, and participants were all young students unlikely to be representative of typical WFN exposure groups.

Test-retest and between-participant variability both impact effect-sizes detectable with any given sample size; so are also important to consider with any measure, including subjective measures such as annoyance ratings expected to vary quite widely both within- and between-individuals. However, intra- and inter-participant measurement variabilities such as the standard deviation of repeat measurements, intra-class correlations (ICCs) (Hallgren, 2012) and 95% confidence limits of group estimates are rarely reported in WFN annoyance studies. This makes it difficult to evaluate if sample sizes are sufficient in terms of statistical power and limits comparisons between studies.

2.2.2. Stimuli differences

Listening tests typically involve relatively short-term exposures and corresponding measurements compared to more real-world settings with much more prolonged exposure times. The exposure duration of the stimuli used in WFN-related listening tests also varies between experiments. For example, Lee et al. (2011) and Ioannidou (2016) used a stimulus duration of 30 sec, while for non-focused listening tests, Waye and Öhrström (2002) used a stimulus duration of 600 sec, and Renterghem et al. (2013) used a stimulus duration of 450 sec. For listening tests involving exposure to infrasound, Tonin et al. (2016) used a stimulus duration of 23 min. Greater standardisation of WFN stimulus exposure times would likely be useful to facilitate between-study comparisons.

Some studies have more specifically examined stimulus duration effects on listening test outcomes. For focused listening tests, Schäffer et al. (2016) found no significant effects of stimulus duration on annoyance ratings. They also used a sample of road traffic noise with various durations ranging from 10 to 40 sec. Corresponding effects of each stimulus duration on perceived duration ratings were classified into three categories: too short, spot-on and too long, and demonstrated a significant positive correlation between stimuli duration and perceived duration ratings. The authors concluded that a stimulus period of 20 sec was both convenient and appropriate for subjects to rate responses with minimal confounding from more prolonged exposure effects. For non-focused listening tests, Poulsen (1991) found stimulus durations ranging from 1-30 minutes had no significant effect on the perceived annoyance of road traffic noise and synthesised gunfire noise. Given Schäffer et al.'s work (2016), relatively short 20 second exposures may be a good choice for examining acute WFN exposure effects for focused listening tests. Short exposures allow for increased noise sample types and replication without an overly prolonged protocol where fatigue, prolonged exposure and other temporal effects could further influence outcomes. On the other hand, longer exposure times may be required in non-focused listening tests to establish acceptable levels for quiet activities and falling asleep. Much longer exposure times may also be needed to establish chronic exposure effects but may be impractical in a laboratory study setting.

A rest period between stimuli could help to reduce participant rating fatigue, which may otherwise occur when stimuli are rated consecutively. On the other hand, rest periods between stimuli also prolong study protocols, potentially also contributing to within-study fatigue. Nordtest guidelines (2002) recommend at least a 10-sec rest period between consecutive stimuli. However, rest periods vary from 10 sec (Lee et al., 2011) to a few minutes (Hafke-Dys et al., 2016; Van Renterghem et al., 2013), or are sometimes variable and left to be self-determined by study participants (Bolin et al., 2014).

The acoustic characteristics of pre-recorded WFN stimuli can vary due to a number of factors, including distance from WTs, time of day, meteorological conditions, WT type, and wind farm power output (Micic et al., 2018). It should also be noted that outdoor-to-indoor noise reduction ratios affect WFN spectral content (Hansen et al., 2015). To reproduce stimuli of WFN similar to those inside buildings, Thorsson et al. (2018) presented differences in outdoor-indoor SPLs at different frequencies to be used in laboratory studies. There could also be difficulties with the measurements of LFN and infrasound of WFN due to wind-induced noise. In a recent study, D'Amico et al. (2021) designed and tested a wind-shielding dome. Their design showed low insertion loss in the low-frequency range, especially at high wind speeds for frequencies up to 0.5 Hz.

In a laboratory testing environment, stimuli may be comprised of processed or unprocessed pre-recorded WFN (Poulsen, 1991) or synthesised WFN constructed to reproduce key features or real-world WFN (Yokoyama et al., 2013b). Signal synthesis has been adopted in several studies to allow for better manipulation and control of LFN and infrasound characteristics such as AM and tonality (Hünerbein et al., 2013; Ioannidou et al., 2016; Lee et al., 2011; Schäffer et al., 2016) not possible with pre-recorded WFN. Some studies have proposed realistic synthesis methods. For instance, Thorsson et al. (2019) synthesised amplitude modulated (AM) components of WFN, maintaining many of the unique features of WFN. Auralisation methods have also been developed in the context of WFN to synthesise emission characteristics, propagation and vegetation noise (Heutschi et al., 2014; Pieren et al., 2014). Using synthesised stimuli enables better separation of WFN characteristics to more specifically examine the effects of key WFN characteristics on human responses. However, synthesis methods vary between studies, and it is not clear if this importantly influences outcomes (Hünerbein et al., 2013; Lee et al., 2011).

Tonin et al. (2016) synthesised WFN based on recordings of WFN near the Shirley Wind Farm in Wisconsin, USA. However, the noise was synthesised using a 0.8 Hz trapezoidal-shaped waveform with 16 harmonics, which may not be representative of real WFN infrasound. Also, Crichton and Petrie (2015) used synthesised infrasound at 9 Hz and 50.4 dB, which is not representative of measured WFN, which contains multiple harmonics of the blade pass frequency in the infrasound range (Bray et al., 2011; Zajamšek et al., 2015). The American National Standard (ANSI/ASA-S12.9-2016/Part7, 2017) outlines measurement and synthesis methods that ensure preservation of LFN and infrasound time-domain features, as these may affect perceptual responses. Irrespective of the synthesis method, accurate signal synthesis is clearly necessary to faithfully reproduce real-world signal characteristics.

Stimuli can be reproduced using either headphones (Ioannidou et al., 2016; Lee et al., 2011) or loudspeakers (Crichton et al., 2015; Van Renterghem et al., 2013). Loudspeaker-based annoyance ratings may be more realistic than headphone-based annoyance ratings (Hafke-Dys et al., 2016; Hünerbein et al., 2013). On the other hand, headphones have some advantages, including increased portability and utility in areas with higher ambient noise, while loudspeakers need a more specific listening room with careful attention to room characteristics (Faller et al., 2013; Taghipour et al., 2019). Besides the audible effects of noise on human responses, vibrational effects may also be important (Freiberg et al., 2019). Body resonance may occur due to tactile sensations of sound, at frequencies between 5-10 Hz (Brownjohn et al., 2011). Yeowart and Evans (1974) showed that headphones could be more reliable for quantifying hearing threshold levels as LFN exposure is limited to a smaller area compared to loudspeakers, where body resonances may contribute to detection via non-aural sensory pathways. However, further work is required to clarify and quantify potential differences in perceived responses when using headphones and loudspeakers.

As WFN contains relatively high levels of LFN and infrasound, special attention should be focused on reproduction methods used in listening room experiments (Cooper et al., 2017). Previous studies reproduced infrasonic and LF components of WFN using either loudspeakers (Crichton et al., 2015) or headphones (Tonin et al., 2016). Recently, Cooper and Chan (2017) concluded that more studies are needed to investigate the effect of WFN infrasound and LFN and recommended the use of a more realistic stereo signal approach for recording and reproduction of WFN infrasound and LFN. The same authors validated their method through a listening test with nine LFN-sensitised participants who accurately detected inaudible LFN stimuli, with stronger perception when noise was exposed to the whole body.

Besides acoustic stimuli, human perception could also be influenced by visual effects of WTs (e.g., audio-visual effects) (Ruotolo et al., 2012; Slater, 2009). Methods have been developed to simulate realistic visual stimuli of WTs (Jallouli et al., 2009; Manyoky et al., 2016; Manyoky et al., 2014), where acoustic stimuli can be reproduced by loudspeakers (Jallouli et al., 2009) or headphones (Slater, 2009), the latter providing the advantage of head-mounted display for spatialised WFN (Slater, 2009).

Stimulus presentation order and replication warrant careful consideration. Most studies recommend the use of non-ordered stimuli testing to avoid potential participant guessing and rating bias from exposure to ordered signals (Ioannidou et al., 2016; Sakamoto et al., 2014; Schäffer et al., 2016). To help make annoyance ratings more consistent, stimuli are often rated more than once, and the average, or some other measure of central tendency, is used to rate each participant's overall response (Hafke-Dys et al., 2016; Søndergaard et al., 2013; Yokoyama et al., 2016). Although one study found no significant effect of the number of stimuli repetitions on overall annoyance ratings (Danish Environmental Protection Agency, 2002), stimulus replication, study sample size and appropriate statistical methodology all remain important considerations for meaningful study outcomes.

WFN is typically directional, so the reproduction system should consider this (Hünerbein et al., 2013). Some studies use more than one loudspeaker to ensure WFN directivity is accurately reproduced (Yokoyama et al., 2013a). In any listening test, sound calibration is important and should be carried out to measure the effectiveness of the reproduction system and to demonstrate if any modifications are needed at the position of the participant's head for faithful noise reproduction.

2.2.3. Rating methods

Many methods have been used to assess WFN-induced annoyance in listening tests (Fastl et al., 2001). Questionnaire scales (verbal or numerical scales) are most commonly used for rating annoyance to WFN. Adjustment methods can also be used, where subjects are asked to adjust SPLs to achieve a specified noise condition such as noise becoming audible or acceptable to the participant. This method is particularly useful for determining specific acceptability threshold levels (ATLs) as discrete SPLs for each stimulus, rather than via interpolation from dose-response curves from other responses. Other rating methods include paired comparisons, where one stimulus is adjusted relative to another stimulus so that the loudness or annoyance is the same, and an adaptive method where SPLs of stimuli are increased and then decreased in up-down trials depending on the participant responses in each trial. Responses from each participant are then averaged once small size steps are reached.

The scale and questions used in the rating method can be an important factor influencing perceived annoyance ratings (Fields et al., 2001). To help make social survey annoyance results more directly comparable, the International Commission on the Biological Effects of Noise

(ICBEN) (Fields et al., 2001, 2006) presented guidelines on survey questions and scales. ISO/TS 15666:2003 (ISO/TS-15666, 2003) also provided these standardised questions and scales. Eleven-point and 5-point rating scales are often used in listening tests (ISO/TS-15666, 2003; Lee et al., 2011), although 3, 7, 9 or 10-point scales have also been used (Tonin et al., 2016; Waye et al., 2002; Yokoyama et al., 2014), either as continuous (Hünerbein et al., 2013) or discrete point numerical rating scales (Lee et al., 2011). Continuous ratings avoid quantisation and rounding to the nearest discrete choice and thus may have some advantages over discrete point scales for detecting small changes in annoyance. However, both rely on subjective psychophysiological responses and thus a broad range of human factors may influence outcomes, where discrete scales may have sufficient resolution and be simpler for participants to provide meaningfully interpretable responses. Anchor descriptors (subjective responses associated with any given point on the rating scale) are also important to consider in the interpretation of annoyance rating scales. Using different descriptors for 11-point scales makes direct comparisons between studies difficult, given the extent to which different descriptors influence participant responses is unclear. For example, the maximum annoyance rating was defined as 'very annoyed' by Lee et al. (2011), while it was defined as 'extremely annoyed' by Hünerbein et al. (2013). Hence, ICBEN (Fields et al., 2001, 2006) and ISO/TS 15666:2003 (ISO/TS-15666, 2003) recommended standardised anchor descriptors (0 = not at all annoyed, 10 = extremely annoyed).

The use of suitable physical or potentially descriptive imaginary situations and specific instructions to participants are other key factors in listening tests. Inuiki et al. (2000) found that participants imagining different situations led to observable differences in the perceived annoyance of pure tones. Moorhouse et al. (2005) instructed participants to rate the same WFN exposure as if it were being presented at night as compared to during daytime situations, from which night-time situation annoyance ratings were considerably higher than daytime situation annoyance ratings. Noise context and subjective expectations regarding night-time situations will likely differ substantially from daytime exposure where situational variables, such as participants wishing to sleep, may be particularly important. Thus, night-time and daytime exposure scenarios may be needed for developing separate WFN dose-response relationships according to the time of day and the context within which noise exposure can occur in real-world settings. The type of instructions given to the participants may also affect the ratings. For example, the instruction of imagining a certain context can be different from being asked to perform some actions (e.g. trying to sleep) during noise exposures.

To help evaluate the effect of focussed attention on noise outcomes, subjects can be asked to deliberately listen to the noise in a 'focused listening test', as opposed to 'non-focused listening tests,' where subjects listen to noise stimuli while performing familiar activities such as reading that may potentially distract and thus reduce the impacts of noise exposure (Van Renterghem et al., 2013; Waye et al., 2002). Despite their comparatively small study number, non-focused WFN listening tests may more closely mimic real-world situations of WFN exposure. In addition, each stimulus is usually longer in non-focussed than in focused listening tests and may thus be more representative of real-world noise exposure and annoyance ratings.

2.3. Comparison between WFN and RTN

Key differences between road traffic noise (RTN) and WFN are summarised in Table 2.1. RTN is typically more problematic in modern cities during the daytime (Halperin, 2014). In contrast, WFN sites are usually located in rural areas with more problematic noise at night-time when background SPLs are low.

The modulation frequency for AM components in WFN is generally between 0.5 Hz and 2 Hz (Hünerbein et al., 2013), compared to the modulation frequencies of RTN between 0.1 and 0.3

Hz (Kim et al., 2010). However, higher modulation frequencies can contribute to higher annoyance (Fastl and Zwicker, 2001).

Table 2.1. A comparison between different characteristics of WFN and RTN.

Noise	Most intrusive periods	Active places	Common AM scheme	Modulation frequency range	Effective SPL range	LF content	Attenuation
WFN	Night-time	Rural areas	Periodic	0.5: 2 Hz	35 dBA: 55 dBA	Higher	Lower
RTN	Daytime	Modern cities	Random	0.1: 0.3 Hz	40 dBA: 60 dBA	Lower	Higher

Based on RTN in Europe, the World Health Organisation (WHO) guidelines (Hurtley, 2009) recommend housing construction should achieve an outdoor-to-indoor noise reduction of 30 dBA with windows closed. However, in rural Australian residences, the windows closed outdoor-to-indoor reduction was found to be as low as 12 dBA for WFN at night (Hansen et al., 2015). Substantial differences in outdoor-to-indoor noise reductions likely reflect higher LF content in WFN compared to RTN studied by the WHO, and housing construction differences in rural Australia compared to Europe, where additional insulation required for cold winters provides superior acoustic insulation. Many rural Australian residences are over 50 years old and contain poor insulation, air gaps and thin, single-glazed windows.

Using listening tests to directly compare WFN with RTN responses, with reduced potential for non-acoustic factors that may affect field measurements, Schäffer et al. (2016) found that WFN was consistently rated as more annoying than RTN, especially with amplitude modulated (AM) WFN. WFN with AM (usually periodic AM) with a higher modulation frequency (0.75 Hz) was more annoying than the AM RTN (usually random AM) with a lower modulation frequency (0.14 Hz). These results are consistent with survey studies showing WFN to be more annoying compared to other environmental noise sources including RTN, railway noise and aircraft noise (Janssen et al., 2011). In contrast, van Renterghem et al. (2013) reported WFN was less annoying than RTN. However, the RTN tested had stronger AM fluctuations than other tested noise sources including WFN, which was also modelled and only tested at 40 dBA. It remains unclear how representative these noise types may be compared to real-world environments. Usually, WFN contains stronger periodic AM components and more variable SPLs than RTN (Schäffer et al., 2016). Moreover, the results obtained by Janssen et al. (2011) included substantially more participants as they were derived using data from two surveys in Sweden (N=341, N=754) and one survey in the Netherlands (N=725), while van Renterghem et al. (2013) included only 50 participants, but in a more controlled listening test study design.

In general, RTN and WFN have sufficiently different acoustic characteristics to differentially impact nearby residences. WFN may be more annoying than RTN at the same SPLs due to the unique characteristics of WFN such as LF content, AM properties, high signal-to-noise ratio in rural areas, and prominent night-time occurrence (Schäffer et al., 2016). This supports that different penalties should potentially be applied to WFN to better account for different annoyance and potential sleep disturbance effects compared to other environmental noise sources such as RTN.

2.4. Results of annoyance quantification

A major advantage of listening tests, compared to survey studies, is that they can isolate, control, and quantify different noise characteristics. Laboratory listening test outcomes show a strong correlation between annoyance and SPL as opposed to field studies, where uncontrolled variables in the field, such as noise characteristics, situational, environmental and human factors, may importantly confound (Schäffer et al., 2018). Reviewing quantification methods for different WFN characteristics is also important and is particularly relevant for the assessment of potential WFN penalties (Schäffer et al., 2018).

Although some research has been carried out on the effect of LFN (20-200 Hz) and/or infrasound (<20 Hz) on humans (Møller et al., 2011), there is very little scientific understanding of LF WFN related annoyance (Leventhall, 2010). Several studies have used acceptability threshold levels to assess LFN effects on humans, and these are summarised in Table 2.2. LFN and infrasound of WFN have the potential to affect other human responses. For example, Chiu et al. (2021) found that heart rate variability was associated with LFN of WFN for a sample of 30 participants living within 500m from a wind farm in western Taiwan.

Moller (1987) studied equal annoyance curves for pure tones in the frequency range of 4 Hz to 31.5 Hz and found that when LFN becomes audible, a slight increase in SPLs leads to a large increase in annoyance. This phenomenon is also evident in the work of Subedi et al. (2009), as shown in Figure 2.1. Persson et al. (1985) also showed that noise with high LF content, at equal A-weighting levels, is more annoying and more difficult to adapt to compared to noise with lower LF content. Higher LF content noise also gave rise to measurable physiological effects. Similarly, Kjellberg et al. (1984) investigated noise with SPLs in the range of 49-86 dBA and frequencies ranging between 15 and 50 Hz in twenty subjects. At the same A-weighted levels, LFN was perceived as 4-7 dB louder and 5-8 dB more annoying than higher-frequency noise.

Various countries with published WFN standards use different criteria for LFN, as shown in Figure 2.2. The lowest frequency included in these criteria ranges from 8 Hz for German criteria to 31.5 Hz for Swedish criteria. It is also notable that ATLs for all curves are lower than the standard HTLs at frequencies below 31.5 Hz. Although these curves appear to differ, methodological differences between curve derivations could be important. For example, the Polish curve was derived with background noise considered. In addition, the Netherlands criteria were derived to predict HTLs rather than ATLs. Consequently, these criteria may be more similar than they appear in Figure 2.2 (Moorhouse et al., 2005).

Table 2.2. Quantification of WFN LFN and infrasound in listening tests.

Study	N	Rating Method	Stimuli Type	T	Outcomes
(Moorhouse et al., 2005)	18	Adjustment method	Synthesised	90 s	Acceptability threshold levels (ATLs) and hearing threshold levels (HTLs) were studied for steady and fluctuating sounds. ATLs were found to be 4-5 dB higher for sounds with strong fluctuations compared to steady sounds.
(Danish Environmental Protection Agency, 2002)	22	11-point scale	Real	120 s	A group of 4 sensitive people reporting LFN in their homes reported higher annoyance levels than 18 non-sensitive people. Comparison between 8 LFN samples indicated that they were not equally annoying at the same SPL due to differences in the frequency content of each sample.
(Crichton et al., 2015)	66	Verbal questionnaires	Synthesised	120 s	The reported symptoms and mood status from WFN infrasound were investigated in terms of the nocebo response of the participants. Demonstrating the nocebo effect to people nearby WTs was proposed to reduce the reported symptoms related to WFN.
(Inukai et al., 2000)	39	Adjustment method	Pure tone	------	Equal unpleasantness curves were obtained using pure tones in a 4-step rating scale of unpleasantness. Instructions for different situations were given to the participants, and the various scenarios yielded different equal unpleasantness curves.
(Persson et al., 1985)	59	Questionnaire (verbal and numerical)	Real	30 min	Using two different LFN signals at the same SPL, annoyance ratings of the noise with lower frequency content were more significant when using two questionnaire types.
(Tonin et al., 2016)	72	7-point scale	Synthesised	23 min	Infrasound levels did not have a statistically significant effect on reported symptoms that have been attributed to WFN. The subjects with a pre-exposure history to WFN had more symptoms compared to those who had never been exposed to it due to some psychogenic responses.
N: Number of participants		T: Stimuli duration			

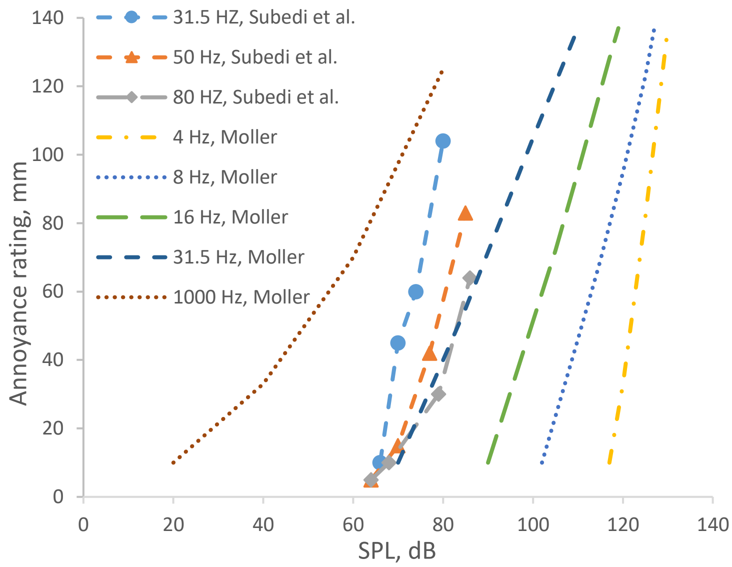


Figure 2.1. Equal annoyance contours as obtained by Moller (1987) and Subedi et al. (2009) using a 150 mm horizontal annoyance scale. The left end of the scale was marked "not at all annoying", while the right end was marked "very annoying".

Moorhouse et al. (2005) investigated ATLs and HTLs for pure tones at low frequencies between 25 and 160 Hz, including separate ATLs of LFN for the daytime and night-time. Their criteria curve is also shown in Figure 2.2. However, participants were randomly chosen from different categories (i.e., LFN sufferers and non-sensitised participants), and infrasound annoyance was not quantified.

Some studies have produced estimates of LFN ATLs using pure tones. However, ATLs of broadband LFN can be different. To investigate the effect of broadband LFN, the spectral content of LFN was quantified by Nilsson (2007) and Schäffer et al. (2018) by calculating the difference between the C- and A-weighted SPLs ($L_{C\text{-}A}$). Their results suggested a relationship between $L_{C\text{-}A}$ and increased annoyance ratings. The human response to filtered broadband WFN (e.g. in the octave bands, which is correlated with human perception) and pure tones could also be compared to investigate possible differences.

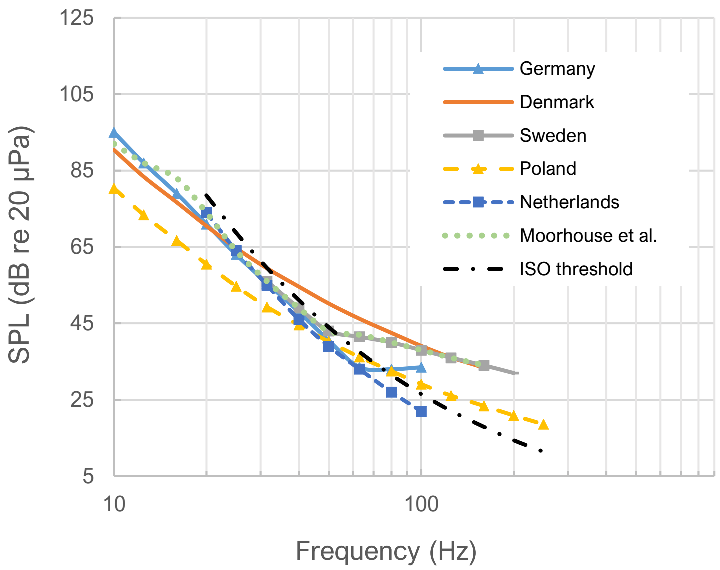


Figure 2.2. Low-frequency noise (LFN) criteria in some countries compared to the ISO threshold (Moorhouse, Waddington and Adams, 2005).

Penalties applied to LFN and infrasound may be more reliable if the real broadband nature of LFN and infrasound is considered. For example, the characteristics of pure tones may be substantially different from infrasound and LFN contained in most WFN (Leventhall, 2010). ISO used filtered WFN (e.g. in octave or 1/3-octave bands) to get HTLs (*ISO 389-7:2005*, 2005). Using the same concept of filtering WFN, the effect of LFN and infrasound can be determined more accurately, based on objective and subjective measurements.

Recent studies sought to quantify LFN emissions. For example, Blumendeller et al. (2020) found that WFN has tonal components at LFs modulated at the blade pass frequency. These components were dominant in stable atmospheric conditions and increased with the increase of wind speeds. Similar findings were found by D'Amico et al. (2021). Zagubień et al. (2019) proposed that lattice supporting towers of wind turbines can generate lower levels of LFN, compared to tabular supporting towers.

WFN AM represents the variability of sound pressure levels in the time domain. The characteristics of AM, including spectral content, modulation frequency, modulated frequencies, SPL, modulation depth (MD) and temporal variations, can affect perceptual responses. The Independent Noise Working Group (INWG) reviewed knowledge regarding excessive WFN AM in terms of its quantification, production techniques and planning conditions (Cox, 2015; Stigwood et al., 2015). Their review showed that more research is needed to investigate the effect of AM on human responses, to improve current AM quantification metrics, especially the IOA metric (Bass et al., 2016), and to update current AM penalties. Several studies have investigated AM characteristics, as summarised in Table 2.3. Lee et al. (2011) used two WFN samples with different spectral content, SPL and MD. Five MD spectra were generated for each sample by increasing the background noise. For both noise samples, there was a statistically significant relationship between SPL and annoyance rating and MD and annoyance rating. However, SPL had a stronger effect than MD.

Additional penalties regarding WFN AM components have been proposed in some guidelines. For example, *NZS 6808* (2010) proposed additional penalties for AM with a maximum penalty of 6 dBA. Renewable UK (2013) proposed a penalty for WFN AM, which applies a linear penalty between 3 dBA and 5 dBA for modulation depths (MDs) between 3 dBA to 10 dBA. This penalty is based on the results of Hünerbein et al. (2013) and Yokoyama et al. (2013), who found a gradual increase of annoyance penalties up to an MD of 10 dBA.

Recently, Virjonen et al. (2019) presented a comprehensive study of the effects of broadband periodic AM on responses using modulation frequencies between 0.25 and 16 Hz and MDs between 1 and 14 dBA and two modulated spectra to represent WFN and RTN. Suggested penalties increased with higher modulation depth and modulation frequency and varied between 4 and 14 dBA where higher modulated frequencies (i.e., RTN spectrum) had higher penalty values than lower modulated frequencies (i.e., WFN spectrum).

Several studies have discussed methods for improving Renewable UK penalties (Renewable UK, 2013). For example, Bowdler et al. (2018) proposed penalties could start at 2 dBA, that the IOA metric (Bass et al., 2016) could be useful for quantifying and penalising WFN AM, and that modulation frequency effects warranted further investigation. Hansen et al. (2018) also proposed further development of current AM penalties related to WFN by including the effects of modulation frequency, tonal modulation, and random AM.

The relationship between SPL and MD was studied by Hünerbein et al. (2013). MD was defined as the difference between the mean peak level and the mean trough level based on 100 ms L_{Aeq} readings. The results using an 11-point rating scale were consistent with Lee et al. (2011), whereby increasing the SPL and MD significantly increased annoyance. A paired comparison method was also used whereby participants matched the annoyance of steady broadband WFN with annoyance from an AM WFN stimulus. A stronger relationship between MD and annoyance was observed when using the 11-point rating scale method compared to the paired comparison method. In addition, adjustments were smaller at high SPLs, where adjustments to achieve equal annoyance between AM vs. no AM stimuli were 1.7 dBA at 40 dBA compared to 3.5 dBA at 30 dBA indicating greater annoyance when MD was present in a lower SPL stimulus. This finding is consistent with the study of Schäffer et al. (2016), in which annoyance ratings of periodic and random AM WFN stimuli became closer to those of

WFN stimuli without AM components at high SPLs.

Annoyance due to the AM WFN was also studied by Yokoyama et al. (2013b), who used the pairwise comparison method, similar to Hünerbein et al. (2013). However, the modulated sound was adjusted to match the level of unmodulated sound. The modulation index (MI) was used and represents the difference between the maximum and minimum SPLs. Increasing MI increased the amount of adjustments. In contrast to the study of Hünerbein et al. (2013), the adjusted levels were higher at high SPLs. Therefore, the effect of SPL on adjustment for AM warrants further research.

Table 2.3. Quantification of WFN AM in listening tests.

Study	N	Rating method	X	T	R	f_m	Outcomes
(Ioannidou et al., 2016)	19	11-point scale	400:700 m	30 s	200:1200 Hz	0.5:2 Hz	Temporal variations and intermittence of AM WFN were investigated using an SPL of 60 dBA. MD was found to have the most significant effect compared to other AM properties.
(Lee et al., 2011)	30	11-point scale	62 m	30 s	100: 8000 Hz	0.865 Hz	SPLs from 35 dBA to 55 dBA using a step of 5 dBA were examined with different MDs for each SPL. Both the SPL and MD were observed to significantly affect annoyance.
(Hünerbein et al., 2013)	20	11-point scale & Paired comparison method	------	20 s	500-1000 Hz& 200:600 Hz	0.8 Hz&1.5 Hz	The SPL of stimuli had a more significant effect than the MD, which also had an observable effect on perceived annoyance. Also, a modulation frequency of 1.5 Hz was more annoying than 0.8 Hz.
(Schäffer et al., 2016)	60	11-point scale	60:600 m	25 sec	50:8000 Hz	0.75 Hz	WFN was more annoying compared to road traffic noise, particularly with AM stimuli included. The annoyance of WFN was attributed to its acoustic characteristics rather than psychological effects.
(S. Yokoyama et al., 2013b)	10	Paired comparison method	252:908 m	10 &15 s	1:8000 Hz	------	With an increase in the AM index, the hearing sensation of AM WFN increased, and the SPL adjustment required for equal annoyance for stimuli with and without AM increased.
(Hafke-Dys et al., 2016)	19	11-point scale	------	8.4 s	500:8000 Hz & Broadband noise	1, 2, 4 Hz	The modulation frequency and MD were found to have a major effect on human response for different spectral content of the noise.
(Yoon et al., 2016)	12	Adaptive method	------	------	------	------	AM WFN was investigated in terms of its loudness and annoyance, compared to unmodulated WFN. AM WFN loudness was perceived as higher. In addition, increasing the MD lowered HTLs.
(Schäffer et al., 2018)	52	11-point scale	400 m	20 s	50:8000 Hz	0.75 Hz	Besides SPLs, the spectral content, AM type and MD can affect annoyance ratings. Hence, they should be considered for WFN compliance testing.

N: Number of participants X: Distance from wind turbine T: Stimuli duration R: Frequency range fm: Modulation frequency

Broadband AM can be classified into “normal” AM (NAM), and “other” AM (OAM), where OAM occurs intermittently at frequencies below 400 Hz with higher MD values (6 to 12 dB) than NAM (3 to 6 dB), which occurs in the frequency range from approximately 500 - 1200 Hz (Oerlemans, 2013). Ionandou et al. (2016) studied these two AM types and found that MD influenced perceptual responses to a greater degree than other studied characteristics such as modulation frequency and intermittence properties. Intermittent periods of OAM with larger MDs did not lead to larger annoyance differences than those observed using periodic NAM with lower MDs. The effect of the modulation frequency on perceived annoyance for WFN stimuli containing intermittent periods of AM showed no significant differences in contrast to the results obtained by Hünerbein et al. (2013) using periodic AM characteristics.

Many factors can affect human annoyance to tonal components, including tonal frequency, tonal audibility and SPL of the stimulus (Oliva et al., 2017; S. Yokoyama et al., 2016). Yokoyama et al. (2016) studied adjusted SPL differences of WFN without tones, compared to WFN with tonal components at an SPL of 45 dBA and tonal audibility ranging from 3 to 15 dBA. The required adjustment of broadband WFN to match annoyance for tonal sounds increased with increasing tonal audibility ranging between 0.2 and 6.1 dBA. There was no noticeable difference between the adjustments required for frequencies of 100, 200 and 400 Hz over the range of tonal audibility investigated.

In contrast, Oliva et al. (2017) found that annoyance increased as a function of frequency for frequencies of 50, 110, 290, 850 and 2100 Hz and tonal audibility ranging from 5 to 25 dBA. The authors proposed relatively higher penalties for high-frequency tones and suggested that no penalty would be required for tones at frequencies of 50 and 110 Hz. Higher penalty values were obtained for 25 dBA compared to 35 dBA at frequencies of 290, 850 and 2100 Hz and tonal audibility of 10 dBA and above, while results at lower frequencies and tonal audibility were not significant. Although this study was not specifically focused on WFN, many of the frequencies investigated have been identified in WFN spectra (K. Hansen et al., 2014; S. Yokoyama et al., 2016).

So far, however, there has been limited research on the potential annoyance of WFN tonal components. There has also been limited quantitative analysis on the effect of tonal components on perceived annoyance, especially the effect of using different frequencies at different SPLs, and the combination of two or more tonal components.

WFN can contain tonal AM components at frequencies below 200 Hz (Hansen et al., 2014). The main physical differences between tonal and broadband AM are shown in Table 2.4. Tonal AM occurs at low frequencies and may be the reason why it travels up to distances of 4 km from wind farms (Zajamšek et al., 2015). Persson Waye et al. (2001) showed that exposure to low-frequency tonal AM noise modulated in the frequency range between 31.5 and 125 Hz at a rate of 2 Hz and level of 40 dBA for 30 minutes could significantly affect perceived annoyance and performance of verbal grammatical reasoning and proofreading tasks as compared to a reference noise with flat frequency content. However, some current guidelines penalise tonal WFN components without consideration of AM, and this may underestimate the annoyance potential of tonal AM (Radun et al., 2019). Recently, Jurado et al. (2019) found that modulated tonal stimuli are perceived louder than tonal stimuli, especially at low modulation frequencies below 2 Hz, which is relevant to WFN (Hünerbein et al., 2013). A pilot study on this topic showed that the penalty of tonal AM WFN could reach five dBA at a modulated frequency of 50 Hz (Hansen et al., 2019).

Table 2.4. Physical differences between tonal and broadband amplitude modulation.

Amplitude Modulation (AM) Type	Modulated frequencies (range 1)	Modulated frequencies (range 2)	Modulation frequencies	Modulation depth	Modulating function	distances
Tonal AM	46 Hz	109 Hz	0.5-2 Hz	Up to 15 dBA	Tone	Up to 4 km
Broadband AM	200:1200 Hz	200:400 Hz	0.5-2 Hz	Up to 15 dBA	Broadband noise	Up to 2 km

2.5. Factors affecting subjective ratings

Besides listening test design characteristics and WFN acoustic characteristics, WFN annoyance ratings can be influenced by other factors, such as human factors (e.g., sensory acuity and sensitivity and attitudes/beliefs) (Radun et al., 2019) and visual effects (Schäffer et al., 2019).

Available evidence is conflicting, but individual noise sensitivity may importantly influence perceived annoyance of WFN. Some survey studies suggest that people living in higher noise-exposed areas are more sensitive to WFN (Arezes et al., 2014; Leventhall, 2010), whilst others suggest that people living in lower noise-exposed areas have higher WFN sensitivity (Moorhouse et al., 2005; Pedersen et al., 2010). However, surveys can only establish associations between variables that may indicate incidental rather than causal relationships with selected outcomes (Shaughnessy et al., 2011). Causality can only be confirmed via interventional experiments manipulating variables relevant to potential causal pathways whilst controlling for potential confounding factors. Differences in exposure group demographics, attitudes and beliefs and many other factors, including previous noise exposure, could all be important factors helping to explain discrepant observational findings.

The visibility of wind farms can also affect human responses. Using virtual reality techniques, Maffei et al. (2013) showed that the distance from wind farms was the most important visual factor, followed by the colour of WTs and finally their number. Similarly, using 3D GIS-based simulation, Schäffer et al. (2019) also showed that annoyance could be increased by WT visibility, but was decreased by landscape visibility.

The relative importance of visual effects, SPL and type of noise on perceived annoyance was investigated by Szychowska et al. (2018) using seven audio samples and pictures of seven landscapes. Their results suggested that the SPL of stimuli was the main factor influencing annoyance ratings; followed by visual effects reproduced during the experiment, and finally, the audio sample (their study also included transportation and natural-like noise samples)., The main contributing/causal factors (i.e. acoustic and non-acoustic factors) and the relative magnitude of their potential effects on human responses is an important target for systematic future listening tests.

2.6. Regulations and guidelines regarding WFN

Recent trends in the development of wind farms have led to a proliferation of guidelines, standards and regulations related to WFN. Allowable limits and metrics used in different countries during the day and night were summarised by Alamir et al. (2021). Some allowable WFN limits depended on the time of day, topography and location specified as the setback distance, recognising that the percentage of highly annoyed people due to WFN decreases with distance from the nearest wind turbine (Pawlaczyk-Łuszczyńska et al., 2018). However, regulations vary substantially, with a difference in the maximum allowable threshold of up to 25 dBA between the lowest compared to the highest allowable limit.

In addition to variable allowable limits, WFN guidelines and standards also vary in the choice of metrics used to quantify noise and the application of penalties for certain audible characteristics. For example, the WHO issued guidelines for environmental noise for Europe in 1999 (Berglund et al., 1999), 2009 (Hurtley, 2009) and 2018 (The World Health Organisation (WHO), 2018) based on spectra dominated by medium- to high-frequencies. The maximum allowable night-time indoor SPL (L_{Aeq}) is 30 dBA, while the maximum allowable night-time outdoor SPL is 45 dBA. Also, an outdoor noise level of 40 dBA should not be exceeded to help minimise potential sleep disturbance. Recently, the WHO released updated guidelines for WFN based on (L_{den}) (The World Health Organisation (WHO), 2018). WFN is recommended to be below 45 dBA. Although somewhat outdated, ETSU-R-97 (1996) specifies a maximum outdoor limit (L_{Aeq}) for the daytime of 35 to 40 dBA, and a maximum outdoor limit

(L_{Aeq}) for the night-time of 43 dBA based on indoor allowable limits of 35 dBA recommended by the WHO. ETSU-R-97 (1996) recommended that the allowable limit (L_{Aeq}) be reduced by 10 dBA for an open window and 2 dBA for the use of L_{A90} instead of L_{Aeq}. As the night-time limit was based upon WHO guidelines, which were updated to 30 dBA indoors for the night-time after ETSU-R-97 (1996) was published, the relevant (L_{Aeq}) night-time limit should be 37 dBA to account for this change. Furthermore, the allowable (L_{Aeq}) limit of 43 dBA for the night-time is greater than the limits for the daytime, unlike many other regulations. In addition, penalties for WFN characteristics such as AM are not considered. *NZS 6808* (2010) is the most recently published WFN standard and presents SPL penalties for specific characteristics of WFN. For example, the maximum allowable level is 40 dBA based on L_{A90}, while for areas with low background SPLs, especially at night, the maximal allowable limit is only 35 dBA.

2.7. Insights into development of WFN penalties

Given substantial variability in noise regulations to date, it remains unclear if current penalties are appropriate (Fowler et al., 2013). Consequently, noise features with the strongest potential to contribute to adverse effects on humans warrant more specific and systematic attention towards more evidence-based allowable limits and penalties for WFN and its potentially most problematic characteristics (NHMRC, 2015).

Since prominent LFN noise and specific WFN acoustic characteristics such as AM and beating tones appear to produce greater levels of noise intrusion and annoyance compared to other noise types, more specific penalties applied to WFN characteristics beyond overall WFN SPL warrant further research and development. Moreover, in addition to penalties for more problematic WFN characteristics, suitable penalties for combinations of WFN characteristics also warrants future research, given that WFN can have two or more of these characteristics occurring simultaneously (Zajamšek et al., 2015).

Objective human responses to WFN and its characteristics also warrant more specific consideration beyond subjective responses in the design of more evidence-based penalties. However, objective measurements of human responses to WFN and its characteristics, particularly in relation to potential sleep disturbance effects, are still lacking (Micic et al., 2018). More objective measures are likely to help reduce uncertainty and debate around acceptable limits of WFN and to help guide the design of more clearly defined and evidence-based penalties designed around dose-response relationships with human annoyance and sleep disturbance (Carlile et al., 2018). Combined vibrational and acoustic effects of LFN could potentially affect allowable limits, and seismological studies have proposed to investigate the propagation of LFN vibrations depending on the soil composition (Pohl et al., 2018).

Different methods have been developed to derive penalties for WFN and its characteristics. Guidelines mainly depend on the overall A-weighted SPL only for deriving these values. In addition to encompassing other noise features, more appropriate limits and penalties may need to better account for human factors such as noise sensitivity and age. For example, most listening test studies have used participants aged between 20 and 40 years for deriving penalties for WFN characteristics. Given ageing effects on sensory acuity, sleep and other human factors, further studies spanning wider age groups more representative of exposed and potentially vulnerable resident groups are recommended to examine if these factors warrant consideration in future penalties.

Previous studies have commonly used the percentage of highly annoyed individuals in a sample population to help define allowable noise limits (Fredianelli et al., 2019). This dichotomised outcome is typically based on a classification of individual annoyance ratings of little annoyed (%LA), annoyed (%A) or highly annoyed (%HA) into the proportion of the sample highly annoyed by a specific noise type and SPL (Miedema and Oudshoorn, 2001) from which a threshold can be used to define an allowable limit. 10%HA (Davy et al., 2018) is a common choice of limit, although alternatives such as 8%HA have been used in Dutch (Bowdler et al., 2018) and 21%HA in Norwegian guidelines (Klæboe and Sundfør, 2016).

Current noise metrics may not adequately account for WFN features and contribute to problems in defining appropriate allowable limits (Council of Canadian Academies, 2015). Thus, considerable opportunities remain to determine which noise metrics are most strongly related to psycho-acoustic human responses. For example, Hansen et al. (2019) showed that both tonal audibility and modulation depth significantly influence human perceptual responses and thus warrant improved metrics for assessing tonal AM that consider both tonal audibility and modulation depth.

The variability of WFN and its characteristics may also warrant consideration in future guidelines. WFN and its characteristics depend on many factors such as weather and operating conditions. Ashtiani (2013) suggested that the occurrence probability of SPLs could be considered for the assessment of wind farm compliance. However, no studies have yet defined the occurrence probability of WFN characteristics likely contributing to annoyance. In addition, the effect of occurrence probability on health is still unknown. In future studies, it might be possible to more clearly define ATLs for annoying WFN SPLs and other relevant acoustic characteristics and then establish acceptable time percentages over which these limits could be exceeded based on field measurement studies. The degree to which intermittent characteristics may affect human impacts also appears to warrant investigation. Clearly, a range of further research on WFN effects on humans is needed to better understand noise impacts and what specific noise features dominate annoyance and sleep disturbance effects before suitable allowable limits and penalties can be developed.

2.8. Summary

Substantial attention has recently been focused towards investigating human responses to WFN. Laboratory experiments importantly complement field surveys as they provide a more controlled environment needed to test for causal relationships between WFN characteristics and perceptual outcomes such as annoyance.

This review highlights the importance of considering a range of design parameters relevant to listening tests, such as participant selection, stimuli duration, control of stimuli characteristics and rating methods. This is needed to ensure that quantitative comparisons between WFN listening test results can be made to meaningfully inform real-world issues most relevant to noise impacts on humans. For instance, instructions and scenarios described during listening tests can measurably influence response ratings. Therefore, careful instruction is likely to help develop separate dose-response relationships for daytime versus night-time scenarios.

Listening tests are particularly useful for investigating the effects of specific acoustic characteristics of WFN such as LFN, infrasound, tonality and amplitude modulation. However, it remains unclear whether the tonal LFN used to establish ATLs in many countries is representative of WFN, which is broadband in nature. Few studies have explored tonal WFN, and while there is agreement that increased tonal audibility increases annoyance, the effect of tonal frequency is less clear and warrants further investigation, particularly at low frequencies. The impact of WFN tonal AM on annoyance has not been explored in listening tests to date, and thus, further studies are also warranted. Further research is required to establish the effect of modulation frequency and intermittency of broadband AM on responses.

WFN has several unique characteristics compared to other types of noise that could contribute to greater perceived annoyance at the same SPLs, compared to other major environmental noise sources such as RTN. Comparisons with RTN demonstrate that WFN has higher LF content, higher modulation frequencies, includes periodic AM, has higher signal-to-noise ratios, and shows less abatement during the night. Thus, allowable limits and penalties designed around traditional industrial and environmental noise sources may be inadequate for WFN and its characteristics.

The investigation of annoyance quantification through listening tests shows that more research is still needed to better understand the effect of acoustic and non-acoustic factors on perceived responses. Moreover, it would be useful to investigate the relative response of these factors through listening tests. This would help to determine which factors may be most relevant for future and more targeted noise assessment and mitigation strategies.

WFN has several unique characteristics compared to other types of noise that could contribute to greater perceived annoyance at the same SPLs, compared to other major environmental noise sources such as RTN. Comparisons with RTN demonstrate that WFN has higher LF content, higher modulation frequencies, includes periodic AM, has higher signal-to-noise ratios, and shows less abatement during the night. Thus, allowable limits and penalties designed around traditional industrial and environmental noise sources may be inadequate for WFN and its characteristics.

Given substantial variability in current WFN penalties and an ongoing lack of evidence to support that current measures and limits adequately account for WFN impacts on humans, further studies are clearly needed towards more consistent and evidence-based WFN penalties. This approach is likely to be more widely acceptable, representative, and reliable than currently variable noise guidelines and penalty approaches. Representative penalties and allowable limits of WFN would also ultimately contribute to more sustainable developments of wind energy as a renewable source of energy.

Chapter 3 Methodology

3.1. Listening test design

Four separate listening tests were conducted in Chapters 4, 5, 6 and 7. The design parameters of these tests are summarised in Table 3.1. This section presents participant demographics, listening room design, response measurement methods and experimental procedures.

Table 3.1. Design parameters of the listening tests conducted in this thesis work.

Chapter	4	5	6	7
Test types	Focussed listening tests	Non-focussed listening tests		
Character studied	Type and level of noise	Tonal AM WFN	Broadband AM WFN	Masking noise of WFN
Responses studied	Annoyance, sleep disturbance and loudness	Sleep acceptability threshold levels (SATs)		
No. of participants	24 (group 1)	48 (group 2)	36 (group 2)	44 (group 2)
Reproduction	Loudspeaker	Headphones		
No. of stimuli	30	15*	7	13
Stimuli type	Real	Synthesised		
Stimuli duration	20 seconds	3 minutes		
Rating method	Questionnaires	Adjustment Method		
Listening room	Psychology Sleep Lab	Adelaide Institute for Sleep Health (AISH)		
Total duration	~60 minutes	~50 minutes *	~25 minutes	~45 minutes
Test time	Morning	Evening (before participants' bedtime)		

*The test was done over two days. The numbers indicate the number of stimuli and duration in one day.

3.1.1. Participants

Two groups of participants took part in this study. A hearing acuity test by checking hearing thresholds was done to check if the participants had normal hearing in the frequency range of 125 and 8000 Hz. Questionnaires were also used to check if there were any abnormalities in their hearing. Participants with hearing aids were excluded from the study analysis. All included participants had normal hearing threshold levels. To determine hearing normality, the hearing levels of participants were measured by specialised audiometric tests for frequencies between 125:8000 Hz. Participants were considered with normal hearing if their hearing levels at each frequency were within 20 dBA from normal hearing threshold levels defined in *(ISO 226:2003*, 2003). Participant noise sensitivity was assessed using the 21-question noise sensitivity survey originally designed by Weinstein (Weinstein, 1978).

3.1.1.1. Group 1

Twenty-four participants (14 females and 10 males) took part in the focused listening test, comparing subjective responses to WFN and RTN with different characteristics (Chapter 4). The participants were English speakers, had normal hearing and were 18-75 years old (mean = 25, SD = 12).

3.1.1.2. Group 2

This group of participants took part in the non-focused listening tests measuring sleep acceptability threshold levels (Chapters 5, 6 and 7). Table 3.2 shows the number of participants, their age and gender. This is the first time that different groups of participants have been included to explore the effects of WFN and its characteristics on acceptability for sleep. Participants were classified based on residential location and self-reported sensitivity to WFN or RTN into 4 different groups (WFN sensitive, WFN insensitive, Quiet rural area and RTN sensitive). All the participants were English speakers living in the states of South Australia and Victoria in Australia. Participants who wore hearing aids were excluded. Data were collected until Feb 2021. However, only data collected up to December 2020 are included in this thesis. All participants were supposed to do the same tests. However, some participants did not continue doing the planned protocol. This explains why a different number of participants were included in these tests. Another reason was that some result files had some technical problems and were excluded from the analysis.

Table 3.2. Number and age distribution of participant groups in this study.

Chapter	Group/character	All groups	WFN sensitive	WFN insensitive	Quiet rural area	RTN sensitive
Ch.5 (Tonal AM WFN effects)	n (age mean ± SD)	48(47±19)	6(69±7)	14(52±17)	13(48±19)	15(34±13)
	Females [n (mean ± SD)]	26(48±18)	2(73±8)	7(58±11)	11(45±18)	6(35±14)
	Males [n (mean ± SD)]	22(46±20)	4(67±5)	7(45±19)	2(65±9)	9(32±12)
Ch.6 (Broadband AM WFN effects)	n (age mean ± SD)	36(53±18)	7(64±14)	13(53±17)	9(51±23)	7(43±15)
	Females [n (mean ± SD)]	20(52±20)	2(73±11)	7(58±11)	7(47±24)	4(40±18)
	Males [n (mean ± SD)]	16(54±18)	5(61±15)	6(47±22)	2(65±13)	3(48±10)
Ch.7 (Masking BGN effects)	n (age mean ± SD)	44(48±18)	3(72±6)	13(55±14)	14(49±18)	14(35±13)
	Females [n (mean ± SD)]	26(48±18)	1(80±0)	7(59±9)	12(46±18)	6(35±14)
	Males [n (mean ± SD)]	18(17±18)	2(69±1)	6(50±17)	2(65±9)	8(34±12)

3.1.2. Listening room and reproduction methods

3.1.2.1. Bedford Park room (listening room 1)

The focused listening test in Chapter 4 was done in a bedroom of the Flinders University College of Education, Psychology and Social Work sleep laboratory in Bedford Park, South Australia, with background noise (BGN) of 23 dBA. A sound card (RME Babyface Pro) was used and an amplifier of type (Lab Gruppen) with a loudspeaker of type (Krix ex). Sound calibration was carried out to ensure that the signal was faithfully reproduced at the participant's head.

3.1.2.2. AISH rooms (listening room 2)

Non-focused listening tests in Chapters 5, 6 and 7 were done in two bedrooms located in the sleep laboratory of the Adelaide Institute for Sleep Health (AISH), Mark Oliphant Building, Bedford Park (Chapters 5, 6, and 7). The room had a BGN of 19 dBA, and it was dominated in the low-frequency range, as shown in Figure 3.1.

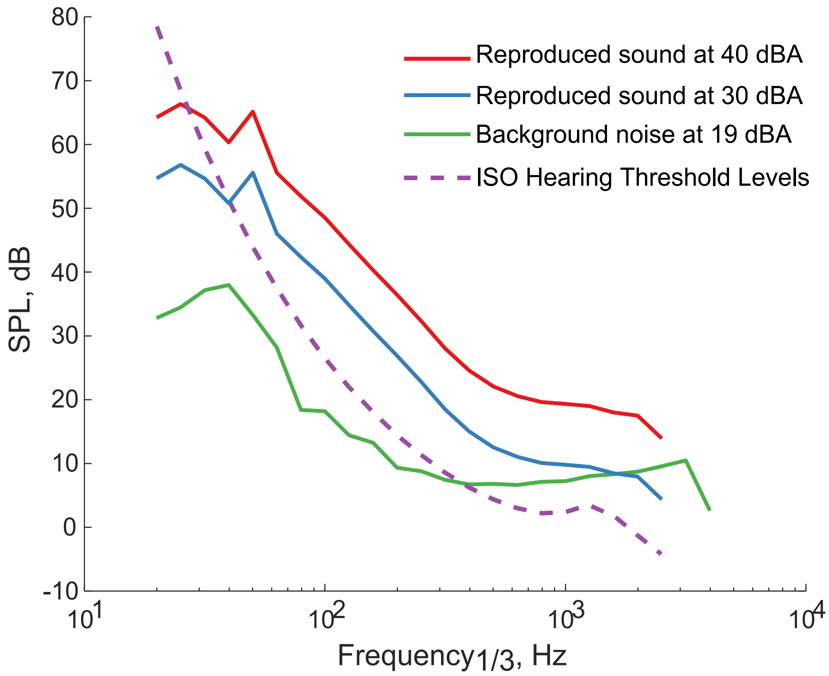


Figure 3.1. 1/3-octave band spectra for background noise (BGN) in the listening room at AISH (listening room 2) at 19 dBA compared to a randomly selected stimulus played in non-focused listening tests at levels of 30 and 40 dBA.

Stimuli were reproduced using Bose Quiet Comfort II headphones connected to an RME Babyface Pro sound card. The frequency response of the headphones as provided by the manufacturer was within 0.5 dB between 0 Hz - 20.8 kHz. However, this error margin of the frequency response could be a bit small (please see https://www.rtings.com/headphones/1-5/graph#565/7913). The headphones had the option of noise cancellation, which was turned on during the experiments to reduce LF background noise.

MATLAB Graphical User Interface (GUI) was used to help guide participants through each test, as shown in Appendix B. The GUI was designed to present step-by-step instructions for each test, present each noise stimulus, capture participant responses, including sleepiness ratings and help to ensure that participants remained awake.

3.1.2.3. Signal calibration

Sound calibration was carried out to ensure that the overall SPL was faithfully reproduced at the participant’s head when noise was reproduced through the loudspeaker (Chapters 4). A

SVANTEK 979 sound level meter was initially used to calculate the amplification factors needed to play the stimuli at the required levels. A P8004 Prosig data acquisition device was used to record and refine the amplification factors that were initially obtained.

The headphones were calibrated using an artificial head (HEAD acoustics HMS III) and a GRAS 45 CA-2 test fixture with RA0039 ear simulators and GRAS 40AD microphones with a frequency range from 3.15 Hz to 10 kHz (Chapters 5, 6 and 7). The calibration was done at a frequency of 250 Hz with an amplitude of 94 dB using a 42AG G.R.A.S. calibrator. The signal outputs from the head and torso simulator and test fixtures were connected to a P8004 Prosig data acquisition device to allow measurement of the raw time signal. Each sample from MATLAB was first played through the reproduction system. The audio output from a tablet used for each test was recorded through MATLAB and Prosig device. The SPL of the recorded signal through MATLAB was matched with the SPL of the recorded signal through Prosig by multiplying the recorded signal through MATLAB with the appropriate calibration factor needed to match the calibration standard. Following checks to ensure the correct output SPL with a repeated known input SPL that factor subsequently was used for all further testing with that sample. This operation was repeated for all the samples to calculate calibration factors for each sample. The calibration procedure for each sample is shown in Appendix C.

3.1.3. Stimulus design

3.1.3.1. Real stimuli

Selected stimuli from field measurements were collected using specialised low-frequency microphones for the focussed listening test in Chapter 4. Five stimuli were used, including three stimuli of WFN. WFN stimuli included "swish", measured at a short range and "rumbling", measured at a long range. A long-range recording of WFN was also included after filtering out AM components ((WFN-NOAM)). Two stimuli were also used for RTN, representing long-range (measured more than 1 km) and short-range RTN (measured less than 1 km) stimuli. The source for RTN was a typical road in South Australia for moving cars at average speeds of less than 60 km/hr. Each stimulus had six levels as shown in Table 3.3 with a duration (length) of 20 sec. The spectra of the stimuli used in this test are shown in Figure 3.2 for the five stimuli used at 33 dBA.

Table 3.3. Levels and types of stimuli included in the focused listening test.

Type	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Tonal AM WFN	33 dBA	36 dBA	39 dBA	42 dBA	45 dBA	48 dBA
Broadband AM WFN						
WFN-no-AM						
Long-range RTN						
Short-range RTN						

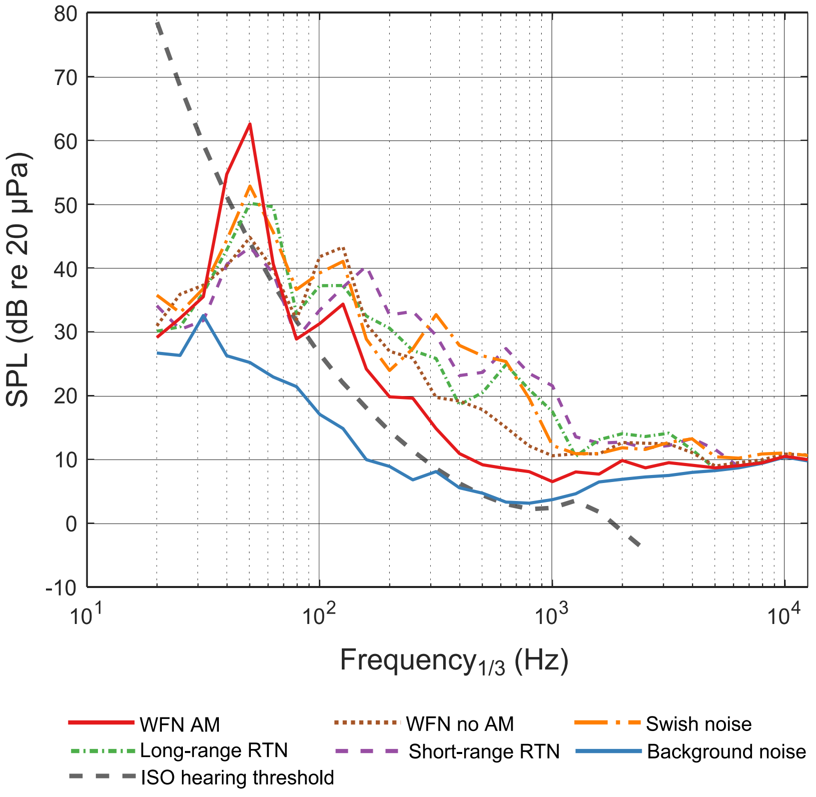


Figure 3.2. Spectra of the stimuli tested at 33 dBA, compared to the background noise in the listening room 1 at 23 dBA and ISO hearing threshold.

3.1.3.2. Synthesised stimuli

Synthesised signals were used instead of real recorded WFN for non-focused listening tests in Chapters 5, 6 and 7 to allow control of stimuli characteristics such as modulation depth (MD), modulated frequency range and tonal audibility. This is because real WFN contains variable SPLs and types of BGN, and it is hard to obtain the desired parameters.

Synthesised samples are summarised in Table 3.4. For LF tonal AM test, the relevant frequencies of tonal AM from field measurement, which were 50 and 110 Hz (Hansen et al., 2014), were used at 4 tonal audibility (TAs) and 4 MDs at each frequency as shown in Table 3.4a. For the swish noise test, 6 stimuli were used at 6 MDs from 2 to 12 dBA with a step of 2 dBA, as shown in Table 3.4b. In the masking background noise test, masking BGN was used at three different noise to signal ratios (SNRs), which are shown in Table 3.4c. The synthesised BGN of WFN without tonality or AM was also included as a baseline stimulus for tonal and broadband AM WFN. For masking background noise tests, a realistic tonal AM WFN was used as a baseline stimulus. The total duration of each stimulus was three minutes. This period was selected as it has previously been found to be sufficient for determining sleep acceptability for most participants in non-focused listening (Hansen et al., 2019).

Table 3.4. Stimuli included in a) tonal AM WFN listening tests where TA: tonal audibility and MD: modulation depth b) swish noise listening test and c) masking noise listening test where BGN: masking background noise type and SNR: signal to noise ratio. A baseline stimulus was also added to these tests.

a)

TA \ MD	0	5	10	15
0	S1	S2	S3	---
5	S4	S5	S6	---
10	S7	S8	S9	S10
15	S11	S12	S13	S14

b)

MD	2	4	6	8	10	12
Stimulus	S1	S2	S3	S4	S5	S6

c)

BGN \ SNR	0	5	10
RC curve	S1	S2	S3
Pink noise	S4	S5	S6
Short-range RTN	S7	S8	S9
Long-range RTN	S10	S11	S12

3.1.3.2.1. Tonal and broadband AM WFN synthesis

To reproduce tonal and broadband AM components in Chapters 5 and 6, stimuli were synthesised according to previously reported methods (Nguyen et al., 2020) to imitate real WFN recorded from previous measurements (Zajamšek et al., 2015). Firstly, the acoustic characteristics of real WFN were analysed. For example, real broadband WFN is modulated in the frequency range between 200-1200 Hz and has a modulation frequency of 0.8 Hz. Therefore, stimuli were synthesised to achieve parameters of the same values. Also, previous measurements of WFN showed that tonal AM consistently occurs at modulated frequencies of 50 and 110 Hz and a modulation frequency of 0.8 Hz (Hansen et al., 2014). Therefore, synthesised samples of tonal AM had the same parameters.

The general spectrum was first obtained using a moving average method. This spectrum was multiplied by white Gaussian noise and then transformed into the time domain using an Inverse Fast Fourier Transform (IFFT). This approach is described in more detail by Lee et al. (2011). The modulation depth of synthesised background noise was less than 1 dBA.

For LF tonal AM and broadband AM tests, the modulated part was first synthesised. Then, synthesised BGN was used to control the overall MD or audibility of synthesised samples by increasing or decreasing the factor (α) according to the following equation (Nguyen et al., 2020):

$$WFN_{AM} = \beta(p_{AM} + \alpha\,(p_{BGN})) \tag{3.1}$$

where WFN_{AM}, is the final amplitude-modulated stimulus, p_{AM}, is the modulated part, α, a factor that controls signal to noise ratio, and β is a factor to control the overall noise level.

The modulated part was synthesised using a sinusoidal wave as the modulating signal. Sound pressure with a frequency, f, can be given by,

$$p_{AM} = A\big(1 + \mu m(t)\big)\cos\,(2\pi f t + \varphi) \tag{3.2}$$

Where μ is the modulation index, φ is an arbitrary phase angle and set to zero. The modulation function, $m(t)$, was represented as white Gaussian noise for broadband AM filtered in the required frequency range between 200-1200 Hz, and it was set to a simple cosine function for tonal AM as follows,

$$m(t) = \cos\,(2\pi f_m t + \varphi_m) \tag{3.3}$$

where f_m is the modulation frequency of the tone, and φ_m, is an arbitrary phase angle that is set to zero. The parameter, A, is the signal amplitude and is set equal to 1.

The parameter, α, controls masking SPL and is adjusted to give the required tonal audibility for tonal AM WFN and the required MD for broadband AM WFN. The parameter, μ, was adjusted along with the parameter, α, to obtain the design values of MD and tonal audibility of LF tonal AM, while it was set constant for broadband AM. MD was quantified using the Japanese metric (DAM) (Yokoyama et al., 2013b), and tonal audibility was quantified by IEC 61400-11 (2012). The spectra of tonal AM and broadband AM stimuli are shown in Figure 3.3 and Figure 3.4, respectively.

3.1.3.2.2. WFN masking background noise test synthesis

For the masking background noise test in Chapter 7, samples were added to real tonal AM WFN at three SNRs (0, -5, -10), calculated using Equation 3.4. The spectra of the stimuli combined with WFN at the three SNRs compared to real tonal AM WFN are shown in Figure 3.5. The spectra for pink and RC noise were synthesised, according to the method of Lee et al. (2011). Recorded noise was used for short- and long-range RTN.

$$SNR = 20\log_{10}\frac{p_{WFN}}{p_{BGN}} \qquad (3.4)$$

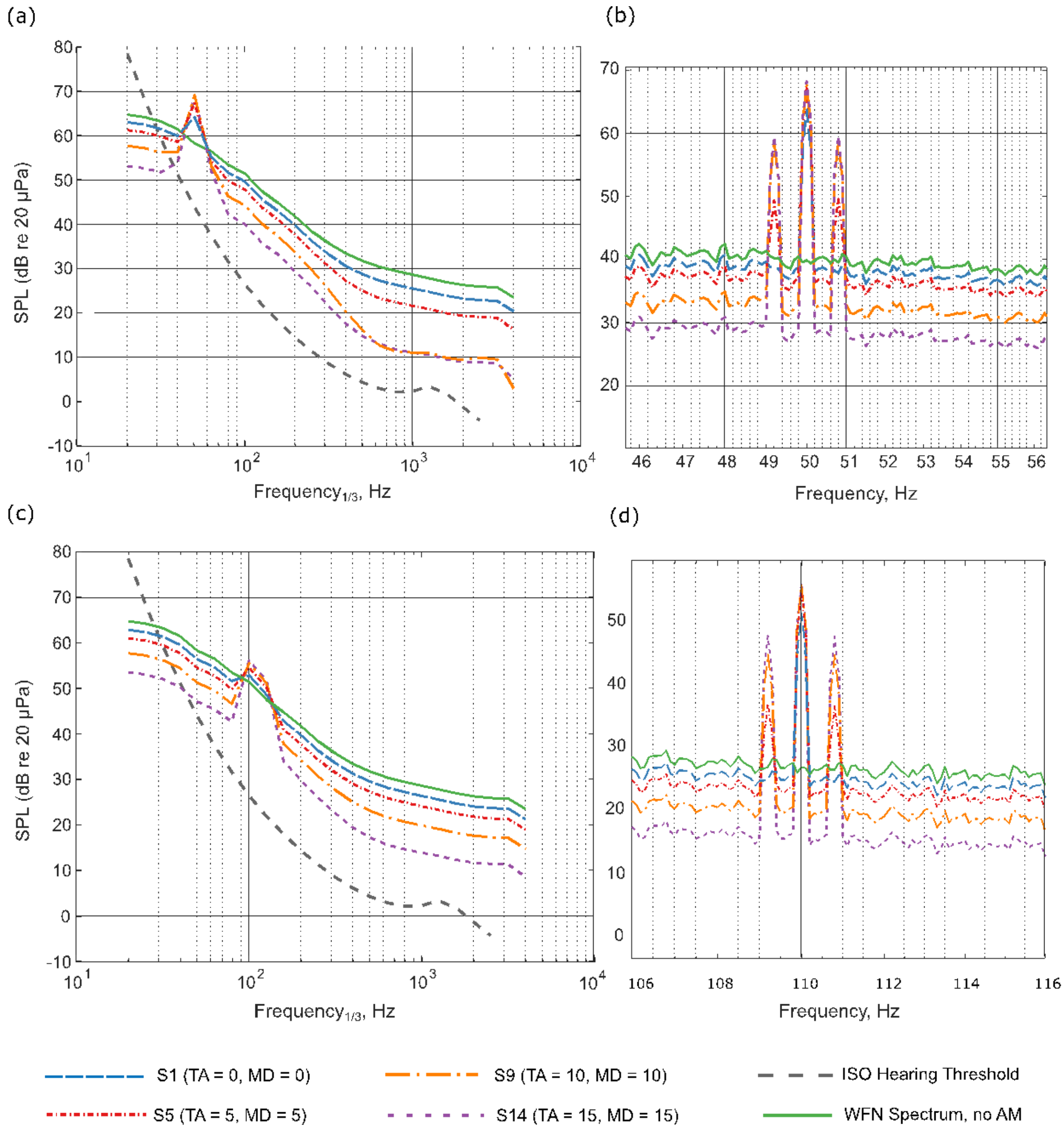


Figure 3.3. Tonal AM stimuli (Chapter 5) at 40 dBA compared to the ISO hearing threshold and the baseline spectrum without AM at 40 dBA (a) third-octave band spectra, modulated frequency of 50 Hz (b) narrowband spectra, modulated frequency of 50 Hz (c) third-octave band spectra, modulated frequency of 110 Hz (d) narrowband spectra, modulated frequency of 110 Hz.

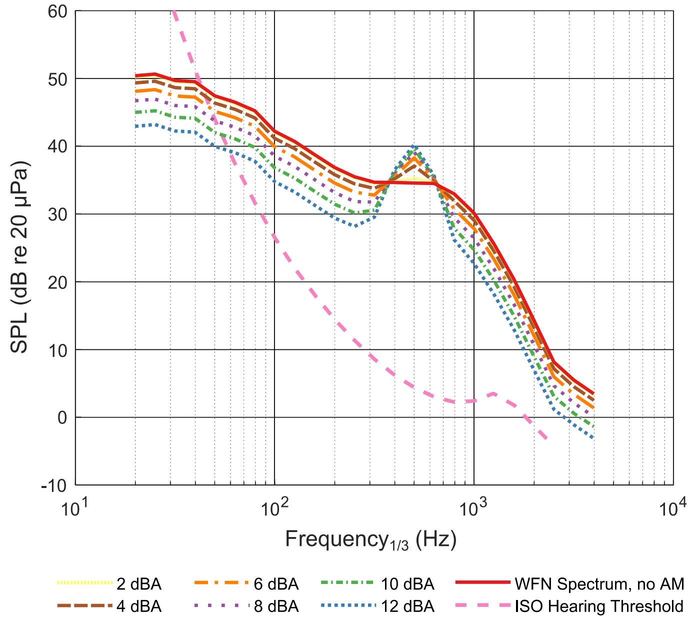


Figure 3.4. Broadband AM stimuli (Chapter 6) at 40 dBA compared to ISO hearing thresholds and the baseline spectrum without AM at 40 dBA.

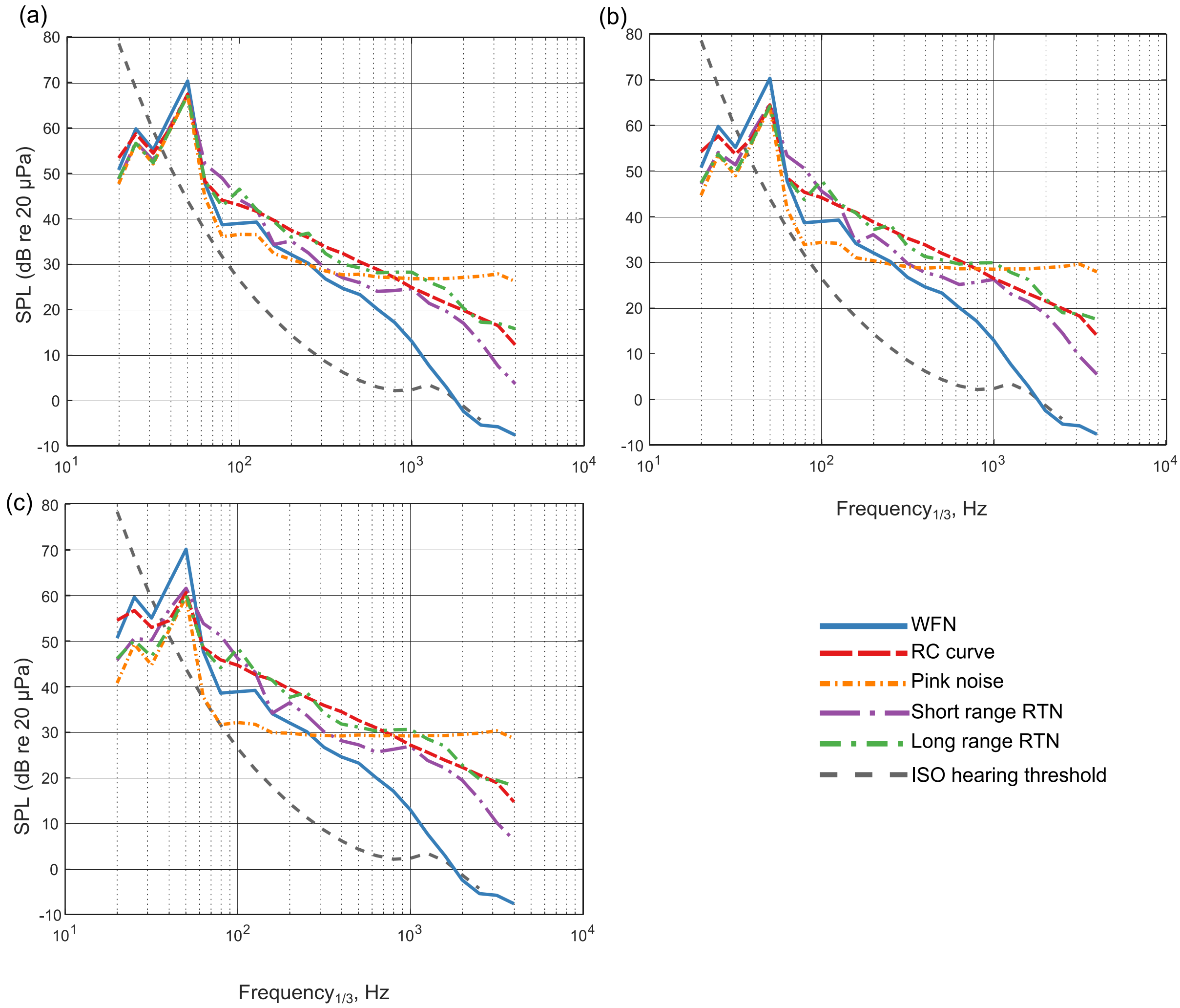


Figure 3.5. Masking noise samples mixed with WFN (Chapter 7) at 40 dBA compared to ISO hearing thresholds and the original WFN at 40 dBA at a) SNR1 of 0 dBA b) SNR2 of -5 dBA c) SNR3 of -10 dBA.

3.1.4. Rating methods

3.1.4.1. Questionnaires

For the focused listening test in Chapter 4, participants rated their responses via questionnaires specifically designed to assess annoyance, sleep acceptability and loudness (i.e., three dependent variables). The questions included were as follows: for annoyance, "*How much did the sound bother, disturb or annoy you?",* for difficulty falling asleep, *"If you are trying to fall asleep in the presence of this sound how difficult would it be?",* and for loudness, *"How loud is the sound?"*. The judgement of loudness, annoyance and sleep disturbance was given on an 11-point discrete scale from *"0"* to *"10"*, with zero labelled as "Not at all" and 10 "Extremely" for annoyance and sleep disturbance, and "No sound" and "Uncomfortably loud" for the loudness. There were six questions included in this test asking about annoyance, sleep acceptability and loudness using absolute and relative scales; however, only three questions with absolute scales (i.e. between 0 and 10) were included in the analysis as they provided a simpler explanation of their results.

3.1.4.2. Adjustment method

The adjustment method was used for the non-focused listening tests in Chapters 5, 6 and 7. This involved the use of a sound card, which had a knob allowing adjustment of stimulus SPL. Participants were asked to adjust the knob to the loudest SPL they considered would be acceptable for sleep. They were also asked to rate their sleepiness after adjusting each stimulus through the question, "*Please indicate how sleepy you felt during the previous trial by selecting the number that corresponds to the words that best describe your state*" on a scale between 0 0 (extremely alert) and 10 (very sleepy). To ensure they were not falling asleep, EEG signals were also tracked, and participants were notified if they fell asleep.

3.1.5. Experimental procedures

3.1.5.1. Focused listening test

For the focused listening test in Chapter 4, participants were asked to deliberately listen to and rate each stimulus. This study was based on a repeated measure design as each participant rated responses. Group 1 participants rated their responses through questionnaires (Section 3.2.4.1), and the listening test was done in listening room 1 (Section 3.2.2.1).

Five stimuli of WFN and RTN (Section 3.2.3.1) were played in random order at six SPLs. The stimuli were reproduced using a MATLAB graphical user interface (GUI). This GUI was developed to guide the participants through the listening test and to record their responses to the questions. First, the participants were exposed to three random stimuli between the quietest and loudest stimuli to train them in using the GUI and to make them familiar with the stimuli used in the test. After training, each participant listened to the thirty stimuli presented in random order (5 types at 6 SPLs). A rest period followed this before participants listened to a second set of the same thirty stimuli presented in random order to test for second exposure effects on response ratings via rating questions after listening to each stimulus. The rest period between exposures was self-determined by the participants with a minimum period of 30 seconds. Training ratings were excluded from analysis.

3.1.5.2. Non-focused listening tests

The non-focused listening tests were conducted in the AISH laboratory and used a repeated measure design in which all participants rated the same stimuli. Group 2 participants stayed in the lab for one week, and listening tests were run in random order on separate days. All tests were done at night before each participant's regular bedtime. After completing questionnaires (Section 3.2.4.1) to assess noise sensitivity, the non-focused listening test was conducted according to the procedure shown in Figure 3.7, with participants responses recorded through the adjustment method.

All instructions were recorded and played aurally through the headphones to ensure that all instructions were delivered consistently to all participants. These instructions welcomed participants, informed them about the aim of the test, which was to investigate the loudest acceptable levels of noise during sleep, then guided participants to familiarise themselves with the room and to try to get comfortable as if they were in their own bed trying to fall asleep.

A practice listening test followed the verbal instructions so that participants could become familiar with the testing procedure and the volume controller. To make sure that participants would not sleep during the test, they were instructed not to fall asleep. After each 3-minute noise sample, there was an alarm. Participants were also familiarised with the alarm during the practice trial. Participants recorded their sleepiness on the touch screen in front of them. The experimenter stayed in the room with participants during the practice test so that participants could ask them any questions at any time. After the practice test, participants did the main listening test, which included between seven and sixteen samples. All aural instructions for non-focused listening tests are provided in Appendix A. To imitate the real sleep conditions, lights inside the room were turned off during the main test, and the test was done in the evening at around 7:30 pm. The duration of each test is mentioned in Table 3.1.

Participants were handed a tablet presenting the GUI and instructed to adjust noise SPLs using the hand-held knob between 0 and 70 dBA above the BGN SPL. After each adjustment, participants were instructed to give themselves time to decide if further adjustments were needed to reach the loudest level likely to be acceptable for achieving sleep. Appendix B has screenshots of the different tabs of the GUI.

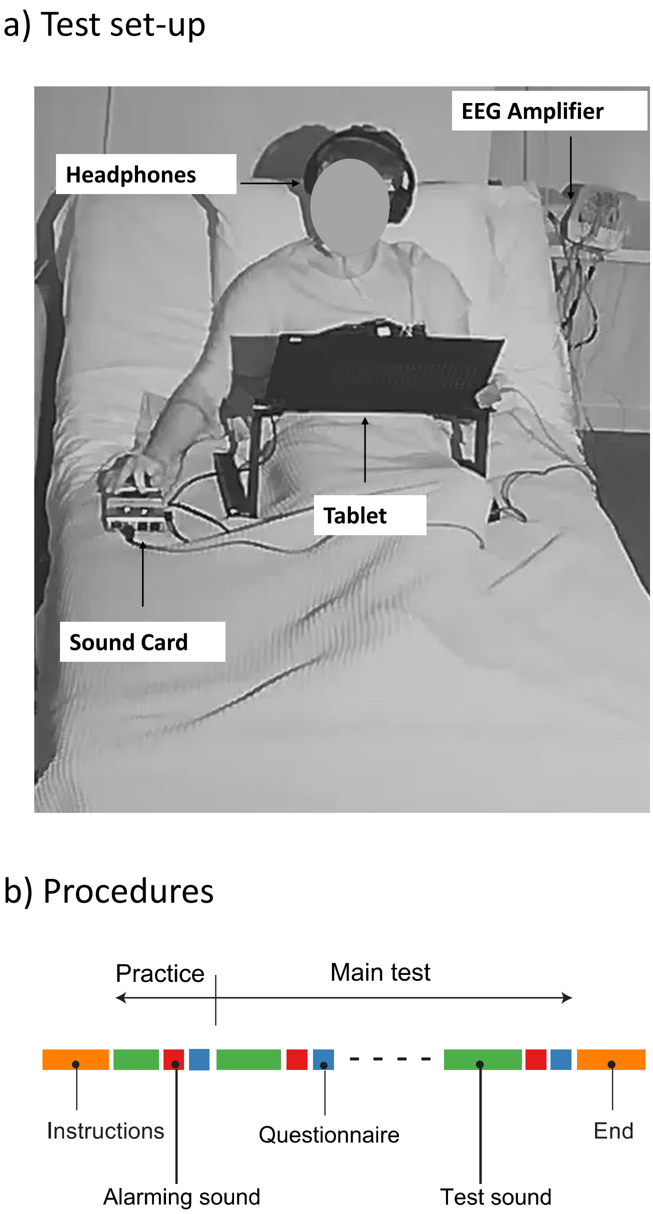


Figure 3.7. Experimental set-up of non-focused listening tests a) actual test set-up b) procedures.

3.1.6. Ethical considerations

The research plan and methodology for these experiments were reviewed and approved by the Social and Behavioural Research Ethics Committee (SBREC project number 7536) and the Southern Adelaide Clinical Human Research Ethics Committee (SAC HREC under project number 55.16). All participants were provided with verbal and written instructions regarding the study and procedures and allowed to have any questions addressed before agreeing to participate and providing informed written consent.

3.2. Modelling methods

The interpretability of models decreases by increasing their performance (Morocho-Cayamcela et al., 2019). For example, highly interpretable algorithms such as classification rules or linear regression are less accurate than highly accurate deep neural networks (DNNs), which are classic examples of black boxes.

High predictive models can be more reliable in identifying the most important factors as compared to statistical models, which could provide other advantages such as drawing inferences (underlying mechanisms) (Bzdok et al., 2018). In this thesis, a range of different statistics and machine learning models were used including statistical mixed models and boosted and bagged trees (further details below). Statistical models were used for drawing inferences and the best performing machine models were used to order the importance of acoustic and non-acoustic factors in terms of their contribution to sleep acceptability.

3.2.1. Statistical models

Statistical models were used to check the effects of different acoustic factors on each specific noise rating outcome (annoyance, difficulty falling asleep and loudness) and sleep acceptability. For focused and non-focused listening test results in this thesis, mixed-model analysis was used to examine fixed effects of acoustic factors, gender and noise sensitivity and their two-way interaction on each specific noise rating outcome. For the non-focused listening tests in Chapters 5, 6 and 7, age, gender, noise sensitivity, participant group and sleepiness ratings were included as covariates. An autoregressive covariance structure (AR1) was used to adjust for serial correlation across trials using SPSS in all analyses. The fixed and random factors included in different chapters are shown in Table 3.5. In all tests, participants were considered as a random effect, each with their own intercept, to account for expected variability between subjects. Where applicable, significant interaction and main effects were examined using Bonferroni adjusted pairwise contrasts within the mixed model. The level of statistical significance was set at $\alpha = 0.05$. Normality check was done to ensure that data was normally distributed.

Normalisation procedures to a baseline can be useful to adjust for inherently variable ratings between participants to more specifically test for consistent within-participant rating effects under different experimental conditions (Shinohara et al., 2014). Ratings in Chapter 4 were normalised to ratings of long-range RTN at 33 dBA for each participant by subtracting the 33 dBA rating value from all other ratings. Long-range RTN was chosen as the primary comparator noise type because it has known effects on human perceptual responses, has relatively high-frequency content, and is one of the most common environmental noise types. In addition, listening tests represent indoor conditions rather than outdoor conditions, and the chosen SPL is more representative of indoor limits (Virjonen et al., 2019).

For non-focused listening tests in Chapters 5, 6 and 7, relative SATs were obtained by subtracting the SATs of the baseline stimulus from the SATs of the other stimuli for each participant. The lower the relative SATs value, the less acceptable for sleep the noise was considered to be compared to the baseline stimulus. The effects of non-acoustic factors, including age, gender, noise sensitivity, participant group and sleepiness ratings, on both absolute and relative SATs, were also examined for each acoustic characteristic examined. Continuous variables such as age, noise sensitivity and sleepiness ratings were classified based on tertiles. Relative SATs were obtained by subtracting SATs of WFN from SATs of all noise types for each participant. Absolute and relative SATs were calculated using A-weighted levels in Chapters 5 and 7, while C-, G-, and Z-weighted levels were also used in Chapter 6.

Pearson product-moment correlations were used to evaluate the strength of relationships between modulation depth and absolute and relative SATs in Chapter 6. The A-, Z-, C- and G-weighted levels were also used to calculate absolute and relative SATs. Tertiles were used in

Chapters 5, 6 and 7 to show the effects of age, noise sensitivity, participant group, gender and sleepiness ratings.

Table 3.5. Factors included in the statistical mixed models included in this thesis.

Chapter	Character studied	Fixed factors	Random factors	Covariates	Outcomes
4	Types and levels of noise	Noise type, its level, gender, noise sensitivity and their two interaction effects.	Participant ID	NA	Absolute and normalised ratings for annoyance, loudness and sleep acceptability
5	Tonal amplitude modulation	Tonal audibility, modulation depth, modulated frequencies and their two interaction effects.		Age, gender, noise sensitivity, participant group and sleepiness rating	Absolute and relative sleep acceptability threshold levels (SATs)
6	Swish noise	Type of noise		Age, gender, noise sensitivity, participant group and sleepiness rating	
7	Masking background noise	Type of masking noise, signal to noise ratio and their two interaction effects.		Age, gender, noise sensitivity, participant group and sleepiness rating	

3.2.2. Machine learning models

Machine learning models in Chapters 5, 6 and 7 were run through the "Regression Learner" application in MATLAB R2020a to predict absolute and relative SATs (https://www.mathworks.com/help/stats/regression-learner-app.html). Table 3.6 shows the factors included in the models. The hyper-parameters of 19 machine learning were automatically tuned to obtain their best combinations. In Chapter 6, only three models were used from that library. They were Gaussian Process Regression, an ensemble of bagged trees and an ensemble of boosted trees. A fourth neural network model was also used in Chapter 6 to identify the most accurate models for relative and absolute SATs due to swish noise. Neural Net Fitting app on MATLAB was used for with 7 neurons in hidden layers using Levenberg-Marquardt training algorithm (https://www.mathworks.com/help/deeplearning/ref/neuralnetfitting-app.html).

Table 3.6. Factors included in the machine learning models included in this thesis. The outcomes were the absolute or relative sleep acceptability threshold levels (SATs).

Chapter	Character studied	Factors included
5	Tonal amplitude modulation	Eight factors: Tonal audibility, modulation depth, modulated frequencies, age, gender, noise sensitivity, participant group and sleepiness rating.
6	Swish noise	Six factors: Type of noise, age, gender, noise sensitivity, participant group and sleepiness rating.
7	Masking background noise	Seven factors: Type of masking noise, signal to noise ratio, age, gender, noise sensitivity, participant group and sleepiness rating.

3.2.2.1. Testing models

10-fold cross-validation was used for predicting the performance of the models. The training was done for 9 folds, and testing was done for the remaining part. This process was run 10 times for each fold. This method has many advantages for predicting reasonable performance of models such as reduced bias, especially for relatively small datasets (Tabe-Bordbar et al., 2018).

3.2.2.2. Rank order importance

The best performing machine learning models were used to identify the relative importance of acoustic and non-acoustic factors to perceived absolute and relative SATs associated with WFN characteristics in Chapters 5, 6 and 7. Ensembles of bagged trees consistently showed high performance. Therefore, this approach was primarily used to identify the relative importance of these factors. For the ensemble of bagged trees, predictor importance values were estimated by permuting out-of-bag observations among the trees and checking resultant uncertainty.

3.2.2.3. Evaluation of machine learning models

For evaluation purposes, four statistical error criteria were used to assess the performance of the proposed models. These were root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE) and coefficient of determination (R^2).

RMSE is defined as the square root of the second sample moment of the differences between predicted and actual data. RMSE defines the discrepancy between the predicted and observed data. RMSE can be calculated using Equation 3.5. The high value of RMSE indicates more variation between predicted and observed data and vice versa. Therefore, the optimal accuracy of the model is obtained when RMSE approaches zero (Shahriari et al., 2016). MSE was also used, and it can be calculated using Equation 3.6.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (Y_i - \hat{Y_i})^2} \tag{3.5}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (\hat{Y_i} - Y_i)^2 \tag{3.6}$$

Where *N* denotes the number of observations; Y_i is the observed value; and $\hat{Y_i}$ is the predicted value.

R^2 is the square of the correlation between predicted and actual data; thus, it ranges from 0 to 1. R^2 equals one for perfectly correlated data and zero for uncorrelated data. R^2 is given by:

$$R^2 = 1 - \sum_{i=1}^{\mathrm{N}} \frac{(Y_i - \hat{Y_i})^2}{(Y_i - \overline{Y}_i)^2} \tag{3.7}$$

Here, $\overline{Y}_i$ is the average of observed values.

Mean absolute error (MAE) defines the accuracy of the model, in which the superior value of MAE must be positive and equal to 1 as shown in Equation 3.8. The negative value of MAE indicates a worse performance of the model.

$$MAE = \frac{1}{\mathrm{N}}\sum_{i=1}^{N} |Y_i - \hat{Y_i}| \tag{3.8}$$

3.2.2.4. Interpretability of machine learning models

Partial dependence plots were generated to interpret the relationships between some input factors used in machine learning models and absolute SATs. This shows how predictions (model outcome) partially depend on one or more input variables based on plots of these relationships. This allowed checking for potential interaction effects between the factors and showing the relative importance of the input factors to the output.

Chapter 4 Perceived annoyance, loudness and difficulty falling asleep with wind farm compared to road traffic noise

4.1. Introduction

A comparison between road traffic noise (RTN) and WFN can show differences in responses to both noise types. RTN is a commonly disturbing noise source with a relatively high LF content. It causes annoyance, mood changes, and degradation of objective and subjective sleep quality (Hurtley, 2009). However, WFN could be considered more annoying than RTN (Schaffer et al., 2016). In this Chapter, real stimuli of WFN and RTN with different acoustic characteristics are used to compare human responses (described in Section 3.1.3.1), allowing a more realistic comparison between noise types. RTN is selected as the main comparator for assessing responses relative to WFN since RTN is a common environmental noise source impacting many people, widely known to affect health, and has a relatively high LF content (Janssen et al., 2011).

Annoyance is usually used to quantify human responses to WFN through listening tests. However, further studies investigating other human responses can help establish the potential effect of WFN on human responses. Previous studies have surveyed the effect of WFN on sleep disturbance (Schmidt et al., 2014). Smith et al. (2020) showed that WFN with amplitude modulation can disturb sleep at high levels through a laboratory sleep study. However, sleep disturbance has not been investigated in listening tests before. Moreover, no studies have investigated WFN dose-response relationships for sleep disturbance before either on a subjective or objective basis (Micic et al., 2018). Perceived loudness can also be an important aspect when comparing responses to noise sources with different characteristics. LFN was found more annoying than medium-to-high frequency content at the same A-weighted SPL

(Moorhouse et al., 2005). We hypothesise that perceived loudness may also be unrepresentative of how much noise is annoying or sleep disturbing.

A more detailed consideration of WFN features and comparisons with other noise sources is needed to identify key characteristics potentially contributing to greater annoyance to WFN, even at lower sound pressure levels (SPLs), compared to other noise sources. Road traffic noise (RTN) is a common environmental noise source with relatively high LF content (Torija and Flindell, 2015), but is generally found to be less annoying than WFN (Schäffer et al., 2016). Therefore, a key aim was to compare RTN and WFN noise characteristics potentially contributing to differential annoyance between noise types, for which current noise guidelines and penalties may be sub-optimal.

This chapter presents WFN and RTN dose-response relationships for annoyance, difficulty falling asleep, and loudness (described in Section 3.1.4.1) in a focused listening test with design parameters introduced in Chapter 3. It also investigated the effects of two non-acoustic factors (i.e., gender and noise sensitivity) and two acoustic parameters (i.e. noise type and sound pressure level (SPL)) on subjective responses. The findings could enhance models of human response to WFN by including influential acoustical and non-acoustical variables to responses. Consequently, this information can help develop current guidelines of WFN, which still vary considerably between jurisdictions.

4.2. Results

4.2.1. Dose-response relationships

The effects of the SPL and type of noise on annoyance ratings are shown in Figure 4.1.a for absolute and normalised ratings. Absolute annoyance ratings showed no statistically significant SPL and noise type interaction effects ($p = 0.09$), but annoyance significantly increased with increasing SPL ($p < 0.001$) and was different between noise types ($p < 0.001$). Normalised annoyance ratings showed similar results to absolute ratings; however, the interaction between SPL and type of noise was significant ($p = 0.022$). Absolute annoyance ratings were significantly higher with tonal AM WFN compared to other noise types except swish noise (absolute difference (mean ± SE) vs swish noise 0.28±0.10, $p = 0.054$; vs long-range RTN 0.21±0.01, $p = 0.036$; vs short-range RTN 0.31±0.08, $p = 0.001$; WFN-NOAM 0.35±0.08, $p < 0.001$). Normalised differences showed that swish and tonal AM WFN had higher annoyance ratings than RTN stimuli. Normalised difference (mean ± SE) of tonal AM WFN vs swish noise 0.00±0.1, $p > 0.05$; vs long-range RTN 0.29±0.08, $p = 0.036$; vs short-range RTN 0.29±0.08 and $p = 0.007$.

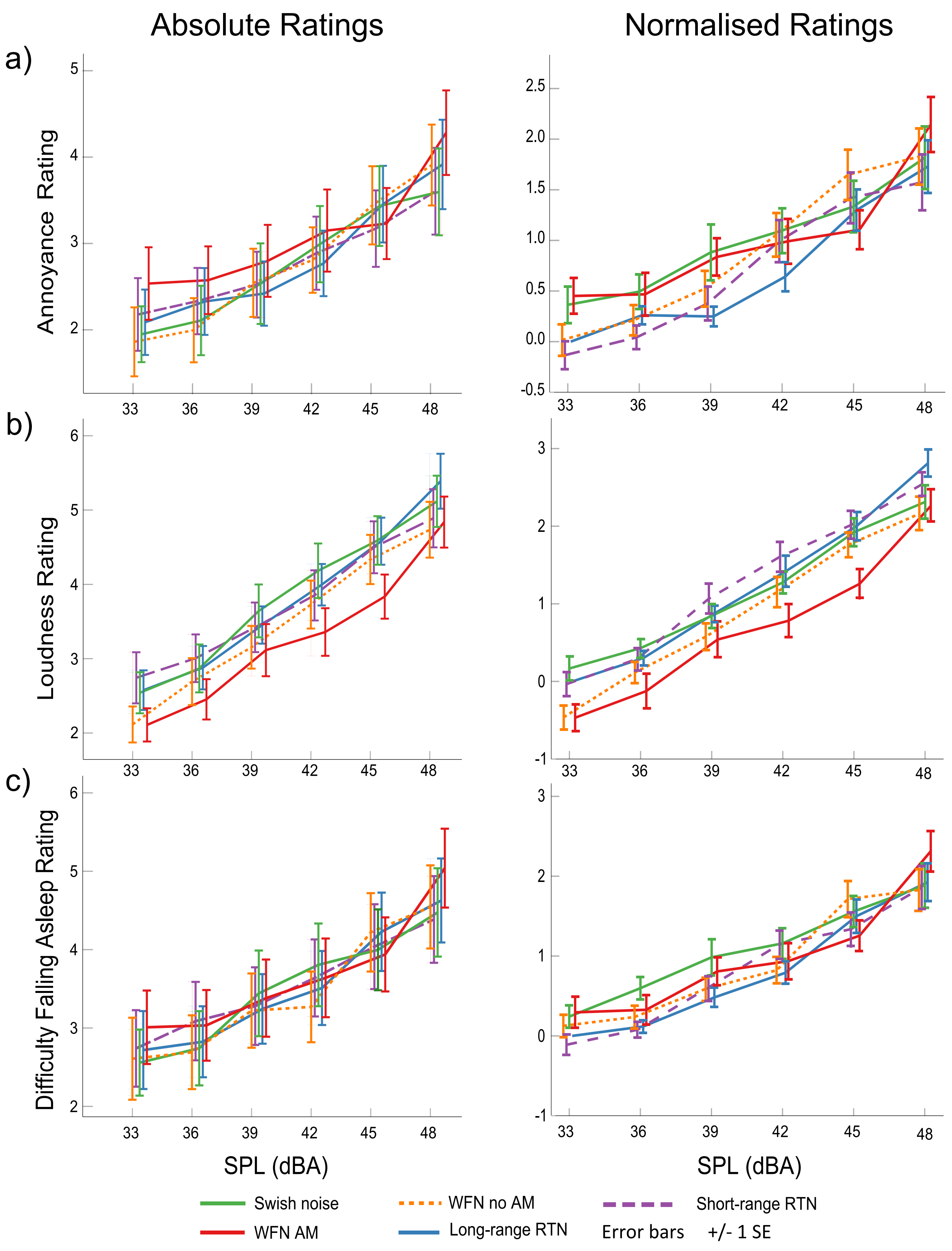


Figure 4.1. Dose-response relationships of WFN and RTN types with different acoustic characteristics for a) Annoyance, b) Loudness, and c) Difficulty Falling Asleep. Normalised ratings (on the right-hand side) are obtained by subtracting ratings of long-range RTN at 33 dBA for each participant from absolute ratings (on the left-hand side).

Loudness ratings (Figure 4.1.b) showed relatively opposite trends when compared to annoyance ratings] . For example, loudness ratings were significantly lower with tonal AM WFN compared to other noise types except for WFN-NOAM (difference (mean ± SE) vs long-range RTN -0.55±0.11, $p < 0.001$; vs short-range RTN -0.61±0.1, $p < 0.001$; vs swish -0.47±0.11, $p < 0.001$; and vs WFN-NOAM -0.17±0.09, $p > 0.05$). Loudness ratings significantly increased with increasing SPL ($p < 0.001$) and showed statistically significant differences between the noise types ($p < 0.001$). Similar results were obtained using normalised ratings, where loudness of tonal AM was perceived differently from other stimuli except for WFN-NOAM (difference (mean ± SE) vs swish noise -0.45±0.09, $p < 0.001$; vs long-range RTN -0.53±0.10, $p < 0.001$; vs short-range RTN -0.54±0.09, $p < 0.001$; WFN-NOAM -0.17±0.09, $p > 0.05$).

The absolute ratings of difficulty falling asleep (Figure 4.1. c) were relatively higher than the absolute ratings of annoyance (differences (mean ± SE) vs annoyance ratings: 0.68±0.04, $p < 0.001$), especially at high SPLs. No significant interaction effects on sleep acceptability ratings were found for SPL and noise type; however, sleep acceptability ratings showed statistically significant differences between noise types ($p = 0.004$) and were significantly increased with increasing SPL ($p < 0.001$). No differences between ratings were found between noise types except tonal AM WFN vs WFN-NOAM (-0.26±0.08, $p = 0.008$). However, normalised ratings showed that likelihood to disturb sleep due to swish noise was higher and different from both types of RTN (difference (mean ± SE) vs long-range RTN 0.27±0.08, $p = 0.016$ and vs short-range RTN 0.25±0.08, $p = 0.025$).

4.2.2. Effects of non-acoustic factors

The effects of noise sensitivity and gender on responses were analysed. Self-reported noise sensitivity showed a positive correlation with annoyance, difficulty falling asleep, and loudness

($p < 0.001$) and statistically significant effects on these responses ($p < 0 .001$), as shown in Figure 4.2.

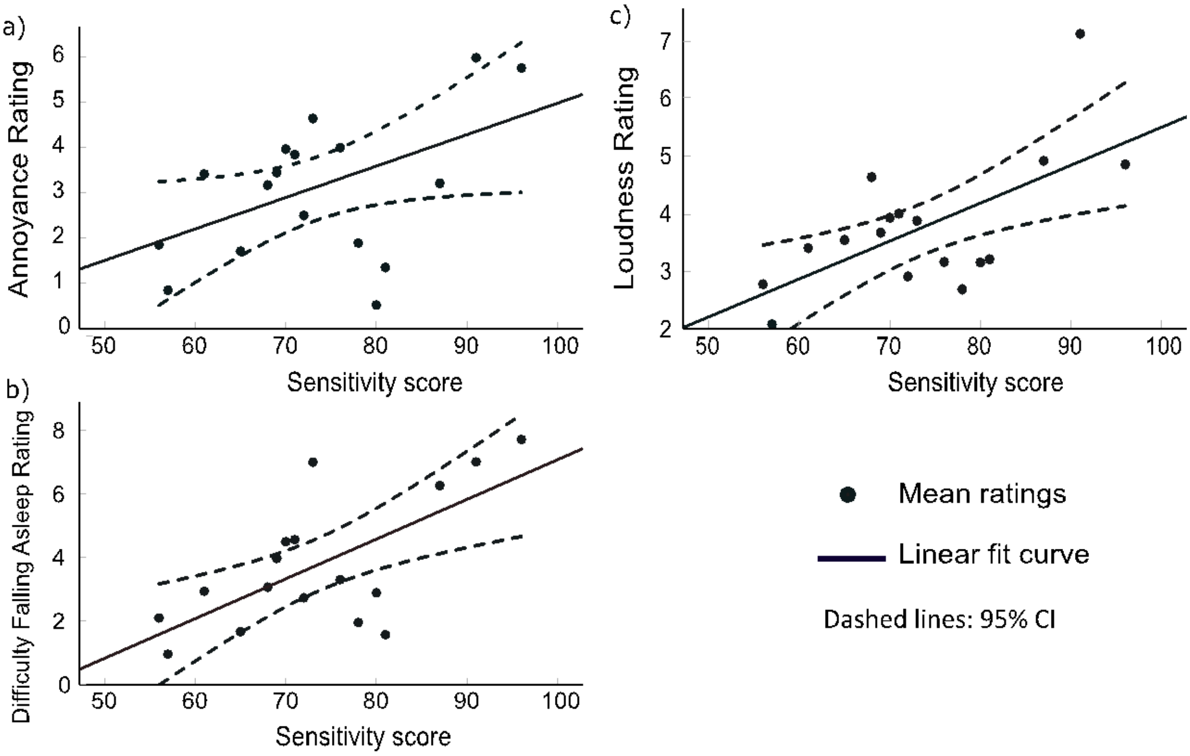


Figure 4.2. The effect of participants' sensitivity score on a) Annoyance ratings, b) Difficulty falling asleep ratings and c) Loudness ratings.

The effect of gender on perceived responses was also significant. As shown in Figure 4.3, females showed higher ratings than males for annoyance (F (1, 87) = 19, (mean difference ± SE) 1.34±0.21, $p < 0.001$), loudness (F (1, 87) = 10, 1.09±0.12, $p < 0.001$) and difficulty falling asleep (F (1, 183) = 28, 1.75±0.23, $p = 0.002$). Males rated tonal AM WFN as more annoying and sleep-disturbing, compared to swish noise ($p < 0.05$). In contrast, females rated swish noise as more annoying and sleep-disturbing than tonal AM WFN ($p < 0.05$) (Figure 4.3).

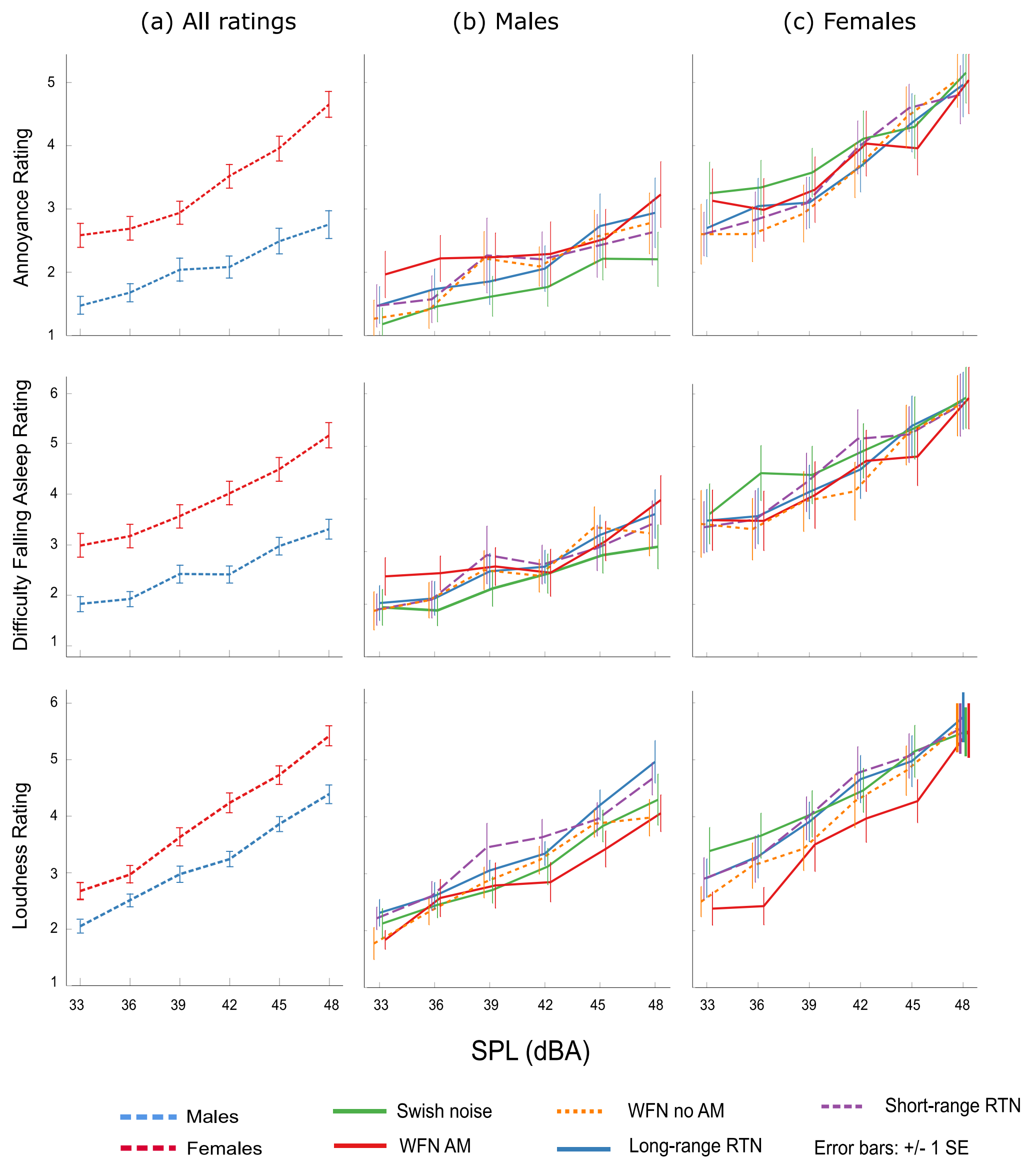


Figure 4.3. The effect of gender on responses for a) All ratings and the mean of all noise types at different levels, b) Male ratings for each noise type at different levels, and c) Female ratings for each noise type at different levels.

4.3. Discussion

WFN dose-response relationships and allowable limits are usually obtained through quantifying annoyance (Fredianelli et al., 2019). However, further control of other responses such as sleep acceptability could yield representative allowable limits, which are still highly variable between jurisdictions as discussed in Chapter 2. This is the first study to undertake a systematic analysis of WFN dose-response relationships of sleep acceptability using real stimuli of WFN and RTN with different acoustic characteristics.

As discussed in Chapter 2, common limits reported for highly annoyed people were (21%HA, 10%HA and 8%HA). However, checking the results in Figure 4.1, it is found that the majority of the scores were below 6 (out of 11 point score, i.e,. less than 60%). Therefore, annoyance results without dichotomisation were presented to report all data for the reader without subjective selection for some of data collected.

In this work, sleep acceptability ratings (Figure 4.1. c) were higher than annoyance ratings (4.1. a) as presented in section 4.2.1. 1this indicates that objective sleep measurements can be used to develop penalties and allowable limits which may be different from penalties and allowable limits to achieve acceptable level of community annoyance. Suggesting separate penalty and allowable limit values depending on different human responses is consistent with other allowable limits used in some guidelines for the daytime and night-time (Alamir et al., 2021).

The low-frequency content of WFN may play an important role in perceived responses. This chapter presented dose-response relationships for the perceived loudness of WFN and RTN with different acoustic characteristics for the first time as shown in Figure 4.1. b. These relationships showed a significant effect of the type of stimulus. Tonal AM WFN was also perceived to be quieter than RTN stimuli. These low loudness ratings may be attributed to the high low-frequency content in these types of WFN stimuli. Only swish noise had relatively high loudness ratings. This could reflect that swish noise is recorded at a short distance from the wind farm and has a relatively higher frequency content than tonal AM WFN and WFN-NOAM.

Tonal AM WFN and WFN-NOAM were perceived as the quietest stimuli. However, this does not mean they were less annoying or less acceptable for sleep. For example, tonal AM WFN was perceived as more annoying (Figure 4.1. a) but quieter noise than RTN stimuli (Figure 4.1. b). These results agree with the findings of other studies in which LFN was perceived as more annoying than medium to high-frequency stimuli at the same SPL (Kjellberg et al., 1984;

Persson et al., 1985). The World Health Organisation (WHO) report (Hurtley, 2009) also stated that LFN could increase perceived annoyance and recommended lower allowable limits for noise sources containing LFN. Annoyance also depends on other psychoacoustic features such as sharpness and roughness (Fastl et al., 2001).

This chapter indicated that SPL of WFN and RTN with different characteristics had a significant effect on response ratings as presented in Section 4.2.1, supporting the findings of previous studies (Lee et al., 2011; Schäffer et al., 2016). This agrees with the findings of Schäffer et al. (2016) in the context of broadband AM WFN, which indicated that WFN with synthesised AM components was more annoying than RTN. Lee et al. (2011) also showed that WFN with different modulation spectra has statistically different annoyance ratings.

Including non-acoustic factors in models of responses can enhance their reliability as they can affect perceived responses to WFN (Janssen et al., 2011). In Section 4.2.2., the increased self-reported noise sensitivity score increased perceived subjective responses such as annoyance, difficulty falling asleep and loudness as shown in Figure 4.2. Noise sensitivity has been identified as a major contributor to noise perception. It has been associated with annoyance of different types of noise, such as airport noise (Gille et al., 2017), RTN (Gille et al., 2016) and railway noise (Vallin et al., 2018), as well as different responses such as food perception (Alamir et al., 2020). Therefore, this study extends knowledge on the effects of noise sensitivity but in the context of WFN and sleep acceptability.

Gender had significant effects on responses as shown in Figure 4.3 where females showed statistically significantly higher ratings of annoyance, potential to disturb sleep and loudness than males. Females have more acute hearing than males (Bies et al., 2017). This trend may contribute to higher negative responses to the noise, but the types of noise should be considered. Recently, Radun et al. (2019) studied the contribution of acoustic and non-acoustic factors to annoyance and sleep disturbance through surveys. Their results suggested that concern for health effects was the most influential factor in indoor and outdoor annoyance and sleep disturbance, followed by gender, where similar to our findings, females were more annoyed than males. Noise sensitivity effects were also significant in Radun et al's investigation, consistent with previous research (Gille et al., 2016, 2017). An important practical implication is that acoustic and non-acoustic factors may both be warranted in models to predict annoyance responses.

4.4. Summary

Human responses to WFN and RTN stimuli were compared in a focused listening test to explore the relationship between SPL and stimulus type and annoyance, potential of sleep disturbance, and loudness. The effects of non-acoustic factors such as gender and sensitivity to human responses were also presented.

WFN and RTN dose-response relationships for annoyance, sleep disturbance, and loudness showed that acoustic characteristics of WFN were associated with increased annoyance and sleep difficulty. Moreover, WFN with amplitude-modulated components was consistently rated as more annoying and yet quieter than RTN stimuli. This study also showed that response ratings were different for perceived annoyance and potential of sleep disturbance.

Gender and noise sensitivity affected perceived responses. Females had higher ratings for annoyance, difficulty falling asleep and loudness compared to males. Moreover, noise sensitivity could play a major role in the studied responses. Specifically, the increase in noise sensitivity scores increased perceived responses (i.e., annoyance, difficulty falling asleep and loudness ratings).

This chapter also provided insights into other relevant non-acoustic factors such as noise sensitivity and gender. More controlled laboratory studies are recommended to examine the relative contribution of other relevant non-acoustic factors such as sensitisation. This could help develop penalties and allowable limits for WFN and its characteristics and enhance models of perceived responses to WFN. Therefore, subsequent chapters focus on perceived responses to WFN acoustic and non-acoustic characteristics using different groups of participants.

Chapter 5 Sleep acceptability of tonal amplitude-modulated wind farm noise

5.1. Introduction

Previous studies on the perception of AM WFN focused on the effects of swish noise (i.e., broadband AM WFN) on annoyance. For example, perceived annoyance due to parameters of swish noise, such as amplitude modulation depth, modulation frequency, modulated spectra and modulation type (i.e., periodic or intermittent) has also been studied extensively (Hünerbein et al.,2013). However, perceived responses due to tonal AM WFN could be different. Jurado et al. (2019) showed that the increase of beating frequency increased perceived loudness from beating tones, compared to individual tonal components. Beating could be similar in its acoustic character to tonal AM (Hansen et al., 2017). Given that AM beating components contribute to louder perception, tonal AM WFN could also contribute to increased negative responses.

Little attention has been paid to the differentiation between modulation depth (MD) and tonal audibility (TA) in the context of tonal AM WFN. Oliva et al. (2017) suggested that there should not be penalties for tonal components at frequencies of 50 and 110 Hz. However, the use of a representative WFN spectrum could make the results differ rather than the use of the inverse A-weighting spectrum. Tonal noise without AM components could elicit different responses compared to tonal AM noise. In Chapter 4, it was shown that tonal AM of WFN could be more annoying and less acceptable for sleep despite being perceived as quieter than road traffic noise (RTN) (see Chapter 4). However, it is unknown whether this increased annoyance is associated with tonal audibility, modulation depth, modulated frequency or a combination of them.

This chapter presents the acceptable threshold levels for sleep (SATs) due to tonal AM WFN with different degrees of tonal audibility and modulation depth using synthesised stimuli described in Section 3.1.3.2 in a novel non-focused listening test with design parameters introduced in Chapter 3 for tonal AM WFN non-focused listening tests. The relevant frequencies of tonal AM concluded from field measurements were used, which were 50 and 110 Hz (Hansen et al., 2014). SATs due to tonal AM WFN relative to baseline WFN stimulus are also presented. The relative contribution of acoustic and non-acoustic factors to the perception of tonal WFN is also presented.

5.2. Results

5.2.1. Statistical models

5.2.1.1. Effects of acoustic characteristics

The effects of tonal audibility on absolute and relative sleep SATs of tonal AM WFN for different degrees of modulation depth and two modulated frequencies are shown in Figure 5.1. Modulation depth (MD) showed statistically significant differences between the absolute SATs (p = 0.026). A post-hoc analysis showed that it was only MD =15 dB that was different from MD = 0 dB ((mean difference in dBA ± SE) 1.88 ± 0.76, p = 0.043), where an increase in MD decreased absolute SATs. A significant interaction effect was found between tonal audibility (TA) and modulated frequency. TAs of 10 and 15 dB were significantly different as compared to TAs of 0 and 5 dB for the two modulated frequencies (46 and 109 Hz) (absolute difference in dBA at 0 -0.94 ± 0.62, p = 0.132; at 5 -1.12 ± 0.62, p = 0.073; at 10 1.30 ± 0.52, p = 0.013; at 15 dBA 1.46± 0.53, p = 0.005). However, overall effect of tonal audibility was not significant (F (3, 379) =0.4, p = 0.866).

MD showed statistically significant differences between the relative SATs (p < 0.001). A post-hoc analysis showed that an MD = 15 dB was different from MDs of 0 and 5 dB, where an increase in MD decreased relative SATs (i.e. noise becomes less acceptable). An MD = 10 dB was also different from MDs of 0 and 5 dB. The difference between MD of 0 dB compared to 10 and 15 dBA was (mean ± SE) 1.46 ± 0.40, $p = 0.003$ and 2.52 ± 0.50, $p < 0.001$, respectively. The difference between MDs of 5 dBA compared to 10 and 15 dBA was (mean ± SE) 1.14 ± 0.39, $p = 0.038$ and 2.52 ± 0.51, $p < 0.001$, respectively. A significant interaction effect was found between TA and modulated frequency as shown in Figure 5.2.b. TAs of 0, 5, 10 and 15 were significantly different for the two modulated frequencies (f (46)-f (109)) (absolute difference at 0 dBA -1.35 ± 0.53, $p = 0.011$; at 5 dBA –1.52 ± 0. 53 $p = 0.004$; at 10 dBA 0.89 ± 0.46, $p = 0.05$; at 15 dBA 1.06± 0.46, $p = 0.021$). The overall effect of tonal audibility was not significant ($p > 0.05$).

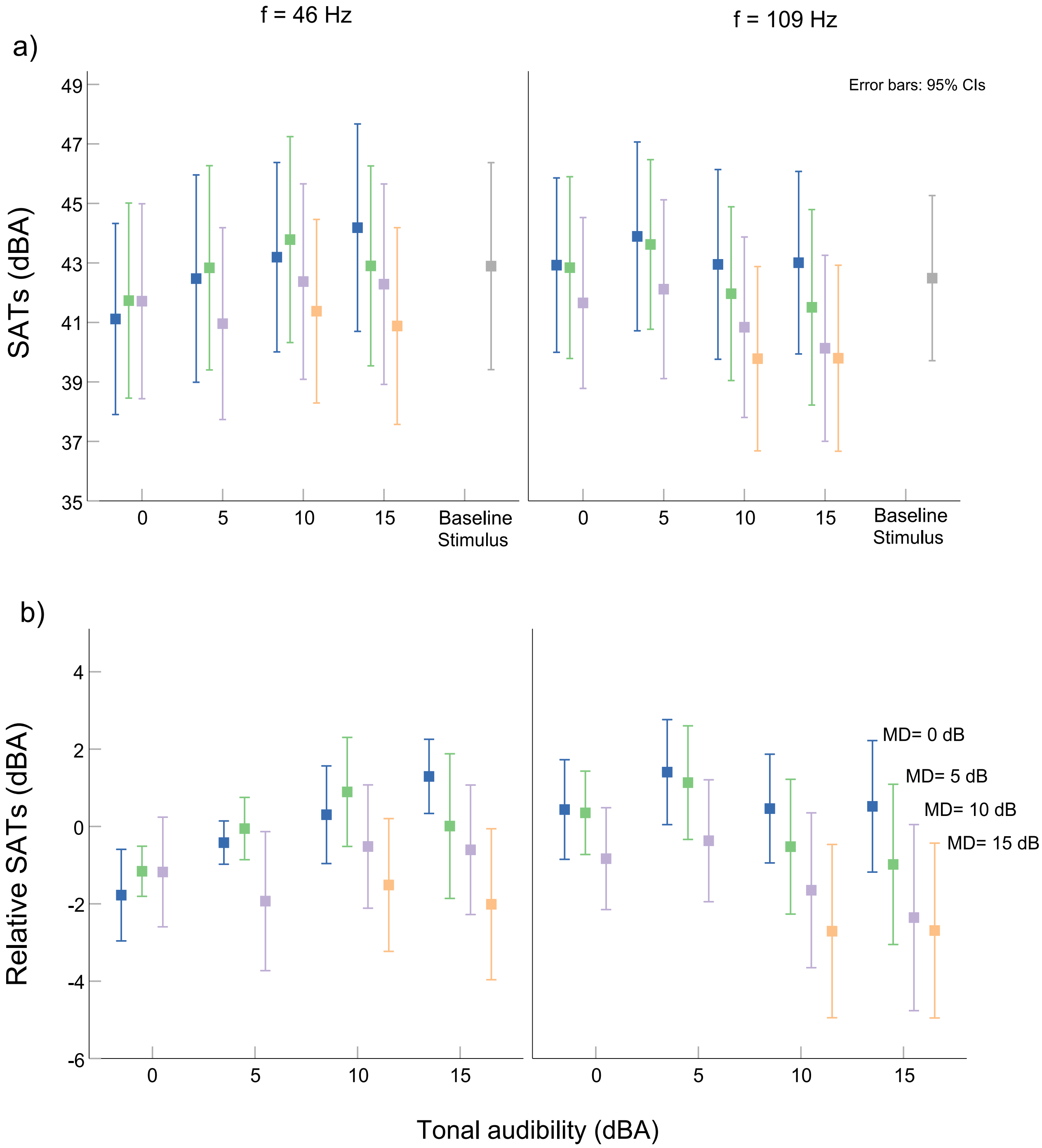


Figure 5.1. The effect of modulation depth (MD) and tonal audibility for two modulated frequencies (f = 50 Hz on the left-hand side and 110 Hz on the right-hand side) on a) sleep acceptability threshold levels (SATs) b) relative SATs. Relative SATs were obtained by subtracting the SATs from the baseline stimulus from the SATs of the other stimuli for each participant. The lower the absolute and relative SATs value, the less acceptable for sleep the noise is as compared to the baseline stimulus.

5.2.1.2. Effects of non-acoustic factors

The effect of non-acoustic factors on absolute and relative SATs is shown in Figure 5.2. The younger participants had higher SATs compared to the older participants. The specific group that participants were assigned to also affected responses. The WFN-sensitive group had the lowest absolute SATs, followed by the WFN-non-sensitive group and quiet rural area group and RTN sensitive group, respectively. Participants with higher sleepiness ratings had lower SATs. Males had lower acceptable SATs than females.

Relative SATs were not necessarily correlated with absolute SATs due to the same group character. For example, the older group of participants had higher differences in the SATs from the baseline (i.e., they could distinguish tonal AM noise types more accurately). In comparison, the younger group had the highest absolute SATs (i.e. noise becomes more acceptable for sleep). This trend was also repeated for the sleepiness rating and participant group. RTN-sensitive and quiet rural area groups had the lowest relative SATs followed by WFN non-sensitive and WFN-sensitive participant groups, respectively. In contrast, RTN-sensitive groups had the highest absolute SATs. The SATs were not different from the baseline stimulus at high sleepiness ratings for sleepiness ratings above 7, while absolute SATs were the lowest at sleepiness ratings above 7.

Appendix D shows detailed differences between absolute and relative SATs for different group characteristics at different degrees of modulation depth and tonal audibility. High relative SATs were obtained at low TAs and MDs (i.e., tonal AM WFN is more acceptable), while they were lower (more negative) at high TAs and MDs (i.e., tonal AM WFN is less acceptable). Low relative SATs of 7.5 ± 2.1 dBA (mean ± CIs) were obtained when the sleepiness score was below five at an MD =15 dBA, TA = 15 dBA and f_m = 110 Hz.

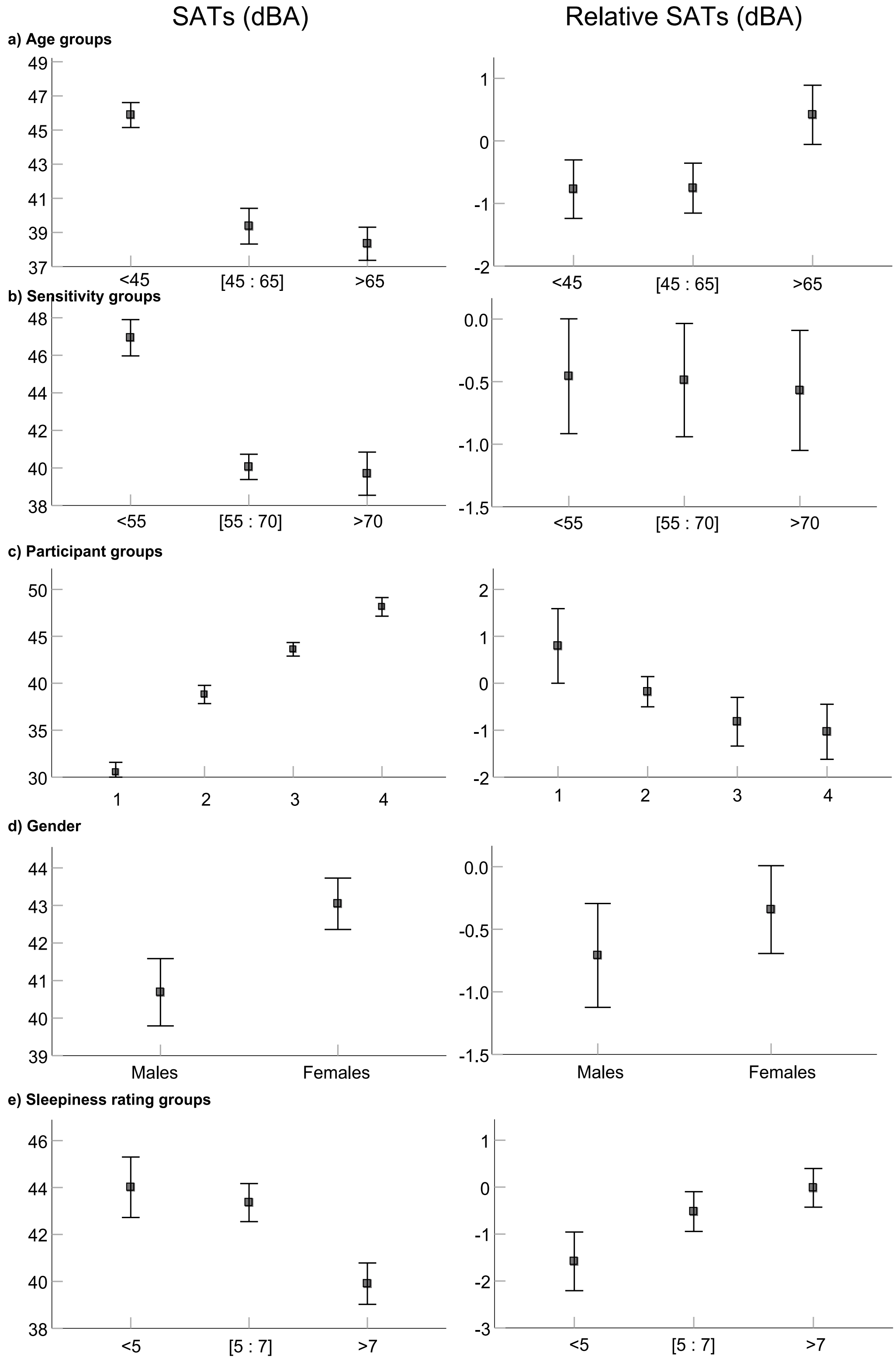


Figure 5.2. The effect of different factors on absolute sleep acceptability threshold levels, SATs (on the left-hand side) and relative SATs (on the right-hand side) is shown for a) Age groups on the x-axis in years b) Sensitivity score groups in years c) Participants groups (1- WFN-sensitive group 2- WFN-insensitive group 3- Quiet rural area group 4- RTN-sensitive groups) d) gender and e) sleepiness rating groups. The relative SATs were obtained by subtracting the SATs of the baseline stimulus from the SATs of the other stimuli for each participant. The lower the absolute and relative SATs value, the less acceptable for sleep the noise is as compared to the baseline stimulus. The error bars represent 95% CIs.

5.2.2. Machine learning models

5.2.2.1. Absolute and relative SAT models

Tables 5.1 and 5.2 show the performance of absolute and relative SAT models, respectively. The ensemble of bagged trees had the best performance with higher R squared and lower mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE).

Table 5.1. Comparison between different models used to predict absolute sleep acceptability threshold levels (SATs) of tonal amplitude-modulated WFN. 10-fold-cross validation was used for all models. Factors included were tonal audibility, modulation depth, modulated frequencies, age, gender, noise sensitivity, participant group and sleepiness ratings. The comparison considered the square of the correlation, mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE).

Model No.	Model	R^2	RMSE	MSE	MAE
1	Linear Regression [linear]	0.34	8.87	78.63	7.19
2	Linear Regression [Interactions linear]	0.40	8.48	71.83	6.79
3	Linear Regression [Robust linear]	0.33	8.91	79.40	7.05
4	Stepwise linear Regression [Stepwise linear]	0.40	8.46	71.50	6.78
5	Tree [Fine tree]	0.82	4.61	21.25	2.92
6	Tree [Medium tree]	0.82	4.59	21.04	3.06
7	Tree [Coarse tree]	0.61	6.82	46.46	5.13
8	Linear SVM	0.30	9.12	83.16	6.88
9	Quadratic SVM	0.36	8.74	76.41	6.39
10	Cubic SVM	0.52	7.54	56.79	5.30
11	Fine Gaussian SVM	0.25	9.49	90.00	7.50
12	Medium Gaussian SVM	0.49	7.78	60.55	5.64
13	Coarse Gaussian SVM	0.30	9.14	83.61	7.03
14	Ensemble [Boosted trees]	0.68	6.16	37.91	4.80
15	**Ensemble [Bagged trees]**	**0.82**	**4.58**	**20.99**	**3.14**
16	Gaussian Process Regression (GPR) [Squared Exponential GPR]	0.63	6.60	43.57	4.84
17	Exponential GPR	0.55	7.36	54.17	5.53
18	Matern 5/2 GPR	0.57	7.14	50.93	5.20
19	Rotational Quadratic GPR	0.64	6.58	43.27	4.79

Table 5.2. Comparison between different models of relative sleep acceptability threshold levels (SATs) of tonal amplitude-modulated WFN. 10-fold-cross validation was used for all models. Factors included were tonal audibility, modulation depth, modulated frequencies, age, gender, noise sensitivity, participant group and sleepiness ratings. The comparison considered the square of the correlation, mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE).

Model No.	Model	R^2	RMSE	MSE	MAE
1	Linear Regression [linear]	0.05	5.35	28.64	3.42
2	Linear Regression [Interactions linear]	0.14	5.07	25.68	3.31
3	Linear Regression [Robust linear]	0.00	5.48	30.00	3.30
4	Stepwise linear Regression [Stepwise linear]	0.14	5.08	25.86	3.33
5	Tree [Fine tree]	0.32	4.38	19.17	2.86
6	Tree [Medium tree]	0.31	4.39	19.28	2.91
7	Tree [Coarse tree]	0.17	4.82	23.24	3.22
8	Linear SVM	0.02	5.42	29.41	3.27
9	Quadratic SVM	0.09	5.23	27.31	3.17
10	Cubic SVM	0.17	4.99	24.94	3.06
11	Fine Gaussian SVM	0.06	5.32	28.32	3.22
12	Medium Gaussian SVM	0.14	5.09	25.91	3.05
13	Coarse Gaussian SVM	0.03	5.39	29.01	3.23
14	Ensemble [Boosted trees]	0.33	4.33	18.81	2.81
15	**Ensemble [Bagged trees]**	**0.38**	**4.18**	**17.50**	**2.65**
16	Gaussian Process Regression (GPR) [Squared Exponential GPR]	0.23	4.79	23.00	3.10
17	Exponential GPR	0.21	4.70	22.07	2.94
18	Matern 5/2 GPR	0.23	4.65	21.60	2.93
19	Rotational Quadratic GPR	0.23	4.80	23.00	3.10

5.2.2.2. Relative importance of non-acoustic factors

Given its high performance, the ensemble of bagged trees was used to identify the most important parameters to absolute and relative SATs, as shown in Figure 5.3. The rank order of the most important parameters for predicting absolute and relative SATs due to tonal AM WFN are shown in Table 5.3.

Table 5.3. The importance rank order of acoustic and non-acoustic factors associated with tonal AM WFN on absolute and relative sleep acceptability threshold levels (SATs).

Rank	1	2	3	4	5	6	7	8
SATs	noise sensitivity	participant group	age	modulated frequencies	sleepiness ratings	gender	modulation depth	tonal audibility
Relative SATs	noise sensitivity	age	sleepiness ratings	modulated frequencies	participant group	modulation depth	gender	tonal audibility

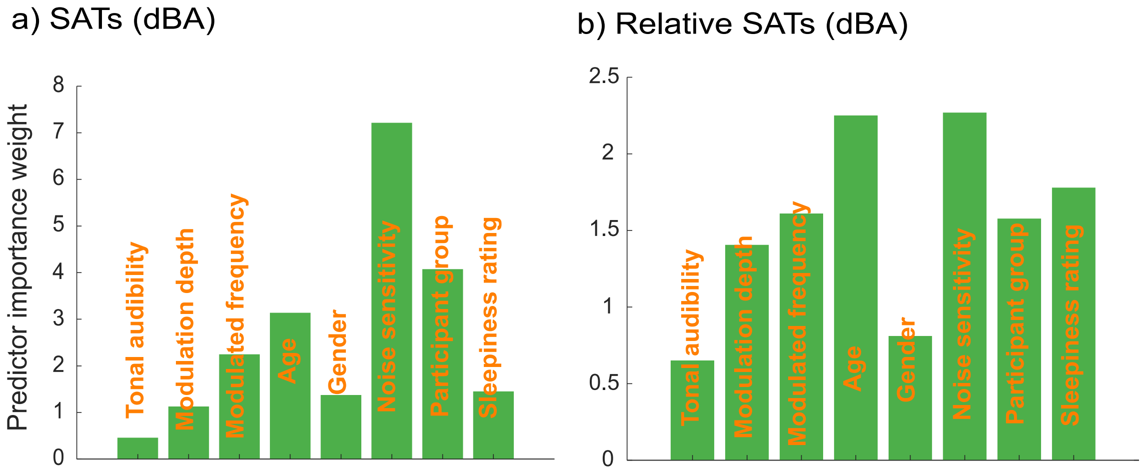


Figure 5.3. The relative importance of acoustic and non-acoustic factors associated with tonal AM WFN on a) absolute sleep acceptability threshold levels (SATs) b) relative SATs.

5.2.2.3. Interpretability of machine learning models

Appendix E shows the results of partial dependence plots. The figures in this appendix show how two factors can affect absolute SATs at the same time. Therefore, effects of individual factors, interaction effects and relative importance of factors can be identified. For example, absolute SATs for age increased gradually until SATs reached maximum values around 30 years. After 30 years, SATs decreased with the increase of age. When age is plotted versus noise sensitivity and groups, it can be seen that these factors affected responses more than the age at the same time, confirming the results in the previous section (Section 5.2.2.2) about the relative importance of factors to SATs.

5.3. Discussion

Modulation depth had a significant effect on SATs. Specifically, the increase of modulation depth decreased the acceptability for sleep. This effect agrees with previous results on the effect of modulation depth on other responses such as annoyance with different amplitude-modulated noise types such as swish noise (broadband AM WFN) (Lee *et al.*, 2011; Ioannidou et al., 2016; Schäffer *et al.*, 2016). However, the effect of tonal AM components at low frequencies has not been investigated before, especially for sleep acceptability. Therefore, these results extend our knowledge of low-frequency tonal AM components of WFN and suggest that a penalty could be required for this character of noise. Further mitigation strategies for lowering modulation depth can also be developed to avoid negative responses.

Interaction effects were found between tonal audibility and modulated frequency. Tonal audibility was less acceptable in terms of absolute and relative SATs at an f_m = 110 Hz as compared to an f_m = 50 Hz. However, the one-way effects of tonal audibility and modulated frequencies did not affect the results (i.e., changing the values of these parameters did not affect the results). These results support the results of previous research at low frequencies, which showed that tonal audibility does not need a penalty at low frequencies of 50 and 110 Hz (Hongisto et al., 2019; Oliva et al., 2017). Therefore, amplitude modulation could be a more annoying part of tonal AM WFN.

These findings imply that acoustic characteristics of WFN such as modulation depth, tonal audibility and modulated frequencies may affect sleep acceptability. Subjective responses to noise are correlated with long-term sleep disturbance (Beutel et al., 2020). Therefore, the acoustic characteristics of tonal AM WFN could also be associated with negative health and wellbeing over a long-term period.

Results of relative SATs showed how different groups of participants can distinguish between stimuli with different characteristics compared to a reference baseline stimulus. For, younger participants had the acceptability for sleep, while they could distinguish tonal AM WFN more than the oldest group. This could be due to their higher hearing acuity than the older group (Bies et al., 2017). Another example is that less sleepy participants had higher acceptability for sleep, while their relative SATs were higher than highly sleepy participants. Therefore, relative responses can be considered in future listening test studies.

Gender plays an important role in the perception of different types of noise. It was shown in Chapter 4 that females found swish noise more annoying than tonal AM, while males found tonal AM more annoying. This effect was confirmed in Chapter 5 as females found tonal AM more acceptable for sleep than males. However, the relative SATs showed no differences between each gender. Females have more acute hearing than males (Bies et al., 2017). This trend may contribute to higher negative responses to the noise, but the types of noise should be considered. Moreover, females are generally more annoyed by WFN than males (Radun et al., 2019).

The increased self-reported noise sensitivity score decreased sleep acceptability for tonal and broadband AM WFN and WFN in the presence of masking noise. This increase also increased perceived subjective responses such as annoyance, difficulty falling asleep and loudness. Noise sensitivity has been identified as a major contributor to noise perception. It has been associated with annoyance of different types of noise, such as airport noise (Gille et al., 2017), RTN (Gille et al., 2016) and railway noise (Vallin et al., 2018), as well as different responses such as food perception (Alamir et al., 2020). Therefore, this study extends knowledge on the effects of noise sensitivity but in the context of WFN and sleep acceptability.

Participants found tonal AM WFN less acceptable for sleep when they had high sleepiness ratings. The test duration of tonal AM WFN was around 60 minutes and participants could have been tired by the end of the test, especially with the small differences between the stimuli types. It has been shown that losing sleep can unleash anger (Krizan et al., 2020). This could explain why the range of high sleepiness ratings obtained in the test of tonal AM WFN could have affected acceptable limits. Most complaints occur during nighttime. Besides lower masking background noise compared to the daytime, this could be explained by expectations as participants with high sleepiness at night could bear noise less than when they are awake. However, the relative SATs showed no differences between sleepy participants. This could confirm that participants find noise less acceptable regardless of its characteristics when they are sleepy.

Older participants had lower absolute SATs (i.e. lower sleep acceptability) than younger participants. This effect could be partially explained by the fact that older participants are more distracted by the noise than younger participants. Older people were more distracted than middle-aged groups by acoustic and visual stimuli for some activities such as driving (Karthaus et al., 2019). Higher distraction levels were also observed for older participants than younger participants (Leiva et al., 2016). Another explanation could be that older participants tend to prefer quiet areas over loud areas for some activities. For example, the liking of food in the presence of background noise was lower for older groups of participants compared to younger participants (Alamir et al., 2020).

Relative SATs of the older age group were not different from the baseline stimulus for some characteristics such as tonal AM WFN, while relative SATs of the younger age group were different. This effect suggests that younger participants could distinguish different noise types more effectively, which is consistent with higher hearing acuity.

Participant selection is an important listening test design characteristic when studying the effects of WFN. However, previous studies have not compared the effects of WFN characteristics on subjective responses for different participant groups (i.e., residential location and self-reported sensitivity) (Alamir et al., 2019). Results showed that wind farm participant groups had low absolute SATs. For example, RTN Sensitive and Quiet Rural Area Groups had higher SATs for tonal AM WFN. Following WFN groups, people living in quiet rural areas generally had low SATs. This could be because the quiet rural area group could be more sensitised to low levels of noise compared to RTN sensitive group, which were predominantly from suburban areas exposed to higher ambient noise levels. However, relative SATs for WFN participants were lower for tonal AM WFN stimuli compared to the quiet rural area and RTN-sensitive participant groups. This suggests that the WFN participant group was generally more annoyed with WFN regardless of its characteristics. These results are consistent with the results of Smith et al. (2020), which suggested that the WFN participant group generally had worse self-reported sleep disturbance in both control and exposure nights compared to a reference group.

5.4. Summary

This chapter showed that the modulation depth of tonal AM WFN affected absolute and relative SATs differently. Specifically, the increase of modulation depth decreased absolute and relative SATs. Tonal audibility and modulated frequency also had significant interaction effects. However, tonal audibility and modulated frequency did not affect absolute and relative SATs, suggesting that modulation depth is a more annoying part of the signal. Therefore, the mechanisms leading to generating tones and high modulation depth should be targeted for more sustainable production of wind energy at future mitigation strategies. Modulated frequency was the most important acoustic parameter for absolute and relative SATs, followed by modulation depth and tonal audibility, respectively. Some non-acoustic factors such as noise sensitivity, age and participant group were more important to perceived responses than acoustic factors. These results could help explain the wide discrepancy in allowable limits applied to WFN by understanding the contribution of different factors. They also help prioritise factors in future mitigation strategies in terms of their contribution to absolute and relative SATs.

Chapter 6 Sleep acceptability of broadband amplitude-modulated wind farm noise

6.1. Introduction

Previous studies have shown the potential effect of broadband AM on perceived annoyance through listening tests. For example, Yokoyama et al. (2013) showed that the increase of modulation depth (MD) increased the adjusted levels that match perceived noisiness between modulated stimuli and unmodulated stimuli at two levels of 35 and 45 dBA. Ioannidou et al. (2016) found that MD had the most significant effect on perceived annoyance, compared to other AM parameters such as intermittency of AM and modulation frequency. Hünerbein et al. (2013) also showed that the increase of MD increased annoyance responses using both questionnaires and adjustment methods for MDs between 0 and 12 dBA at SPLs between 25 and 45 dBA. Virjonen et al. (2019) developed penalties for broadband AM noise at a wide range of MDs and modulation frequencies and two modulated frequencies at a level of 35 dBA where the penalties varied from 4 to 14 dBA.

Nighttime responses to WFN and its characteristics could help establish the effects of WFN. Moorhouse et al. (2005) reported that the acceptability threshold levels of real sounds at nighttime could be lower than their acceptability threshold levels during the daytime through listening tests. While some studies investigated the effect of swish noise on annoyance through listening tests, no previous has been presented to show its effect on acceptability for sleep to check if it is the same as annoyance.

This chapter presents absolute and relative sleep acceptability threshold levels (SATs) due to broadband amplitude-modulated wind farm noise (swish noise) for the first time using different degrees of AM modulation depth described in Section 3.1.3.2. A novel non-focused listening test was undertaken with design parameters introduced in Chapter 3 for broadband AM WFN non-focused listening tests. This chapter also shows the relative contribution of acoustic characteristics and non-acoustic factors on the perception of broadband AM WFN. Different statistical and machine learning models are used to identify differences between modulation depth degrees and relative contribution between the factors. The correlation between modulation depth and absolute and relative SATs was also presented using different weighted levels.

6.2. Results

6.2.1. Statistical models

6.2.1.1. Effects of acoustic characteristics

Figure 6.1 shows absolute and relative SATs at different degrees of modulation depth (MD) using different weightings. The correlation (described in Section 3.3.1) between MD and absolute and relative SATs using different SPL weightings can also be shown. The lowest correlation was obtained when using the A-weighting for absolute and relative SATs, while higher correlations were obtained when using the Z-, C- and G-weightings. All correlations were significant ($p < 0.001$) except for the A-weighted level correlation with absolute SATs ($p > 0.05$).

Statistical mixed models (described in Section 3.3.1) using A-, C-, G-, and Z-weighted levels showed a relationship between modulation depth and relative SATs. An MD of 10 dBA was different from other MDs except for 12 dBA using all models. However, for the Z- and A-weighted level models, relative SATs of MD 10 dBA was not also different from an MD of 8 dBA.

Figure 6.2 shows the distribution of absolute SATs due to the swish noise. The SATs of the baseline were divided into three different ranges depending on the tertiles of obtained SATs. Figure 6.3 shows relative SATs for the three baseline SAT ranges. Significant response differences were found for different MDs ($p < 0.001$) as well as different ranges of the baseline SATs ($p < 0.05$). The presence of broadband AM was less acceptable for sleep for participants with baseline SATs below 30 dBA compared to participants in other baseline SAT ranges ($p < 0.001$). There were also no significant relative SAT differences between different degrees of modulation depth for baseline SATs over 40 dBA. The range below 30 dBA had the lowest relative SATs, and low relative SATs of 6.8 ± 1.8 dBA (mean ± CIs) were obtained at an MD of 10 dBA.

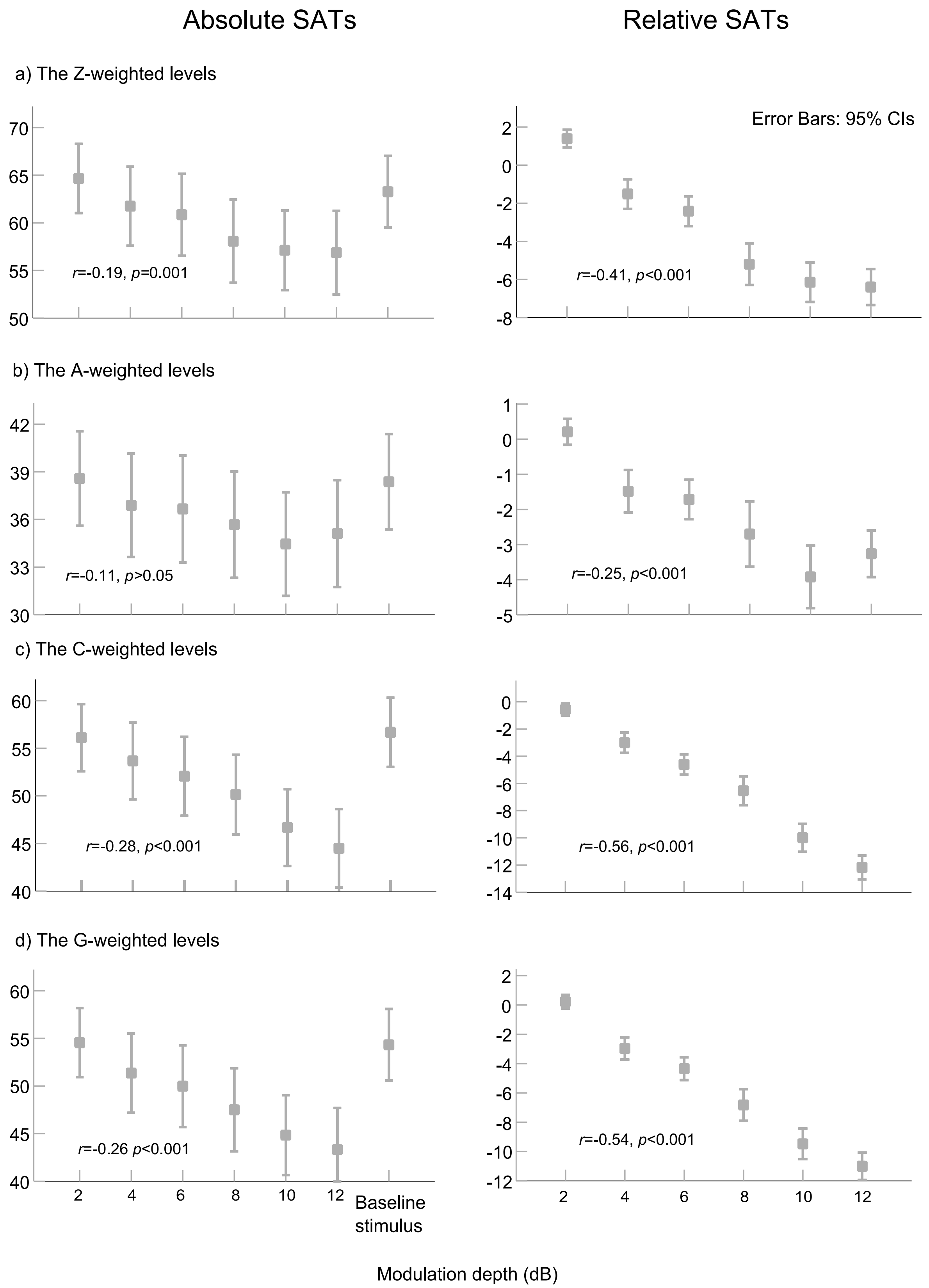


Figure 6.1. Sleep acceptability of swish noise with various degrees of modulation depth using different weightings: a) Z-weighting, b) A-weighting, and c) C-weighting d) G-weighting. All error bars represent 95% CIs. The relative SATs were obtained by subtracting the SATs of the baseline stimulus from the SATs of the other stimuli for each participant. The lower the absolute and relative SATs value, the less acceptable for sleep the noise is as compared to the baseline stimulus.

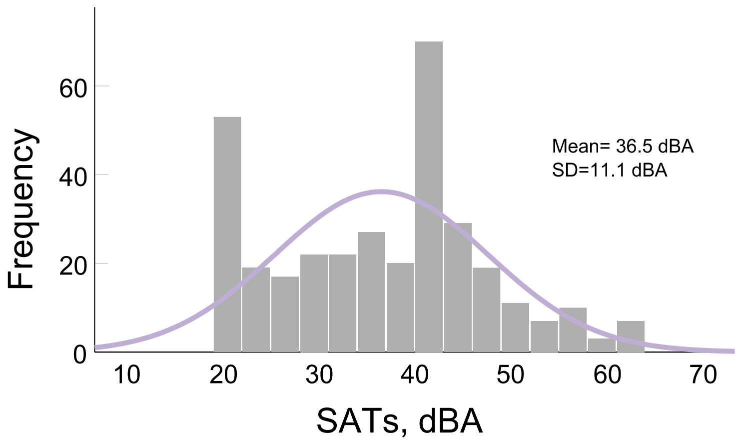


Figure 6.2. Distribution of sleep acceptability threshold levels (SATs).

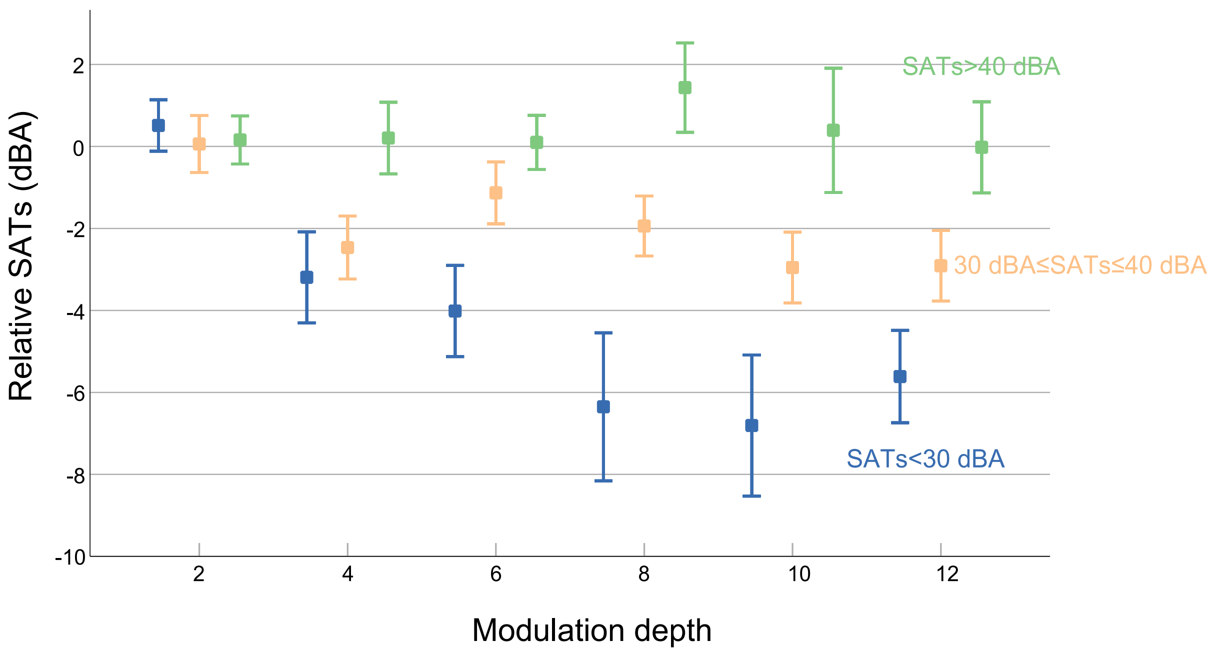


Figure 6.3. Relative sleep acceptability threshold levels (SATs) for three ranges of baseline SATs at different degrees of modulation depth in dBA. Relative SATs were obtained by subtracting the SATs of the baseline stimulus from the SATs of the other stimuli for each participant. The lower the absolute and relative SATs value, the less acceptable for sleep the noise is as compared to the baseline stimulus. The error bars represent 95% CIs.

6.2.1.2. Effects of non-acoustic factors

Effects of non-acoustic factors on absolute and relative SATs are shown in Figure 6.4. Younger participants had higher SATs than older participants. The WFN-sensitive group had the lowest absolute SATs followed by WFN-non-sensitive and quiet rural area groups and RTN sensitive groups, respectively. Groups of sleepiness ratings SATs did not affect absolute SATs. No gender differences were also found on absolute and relative SATs.

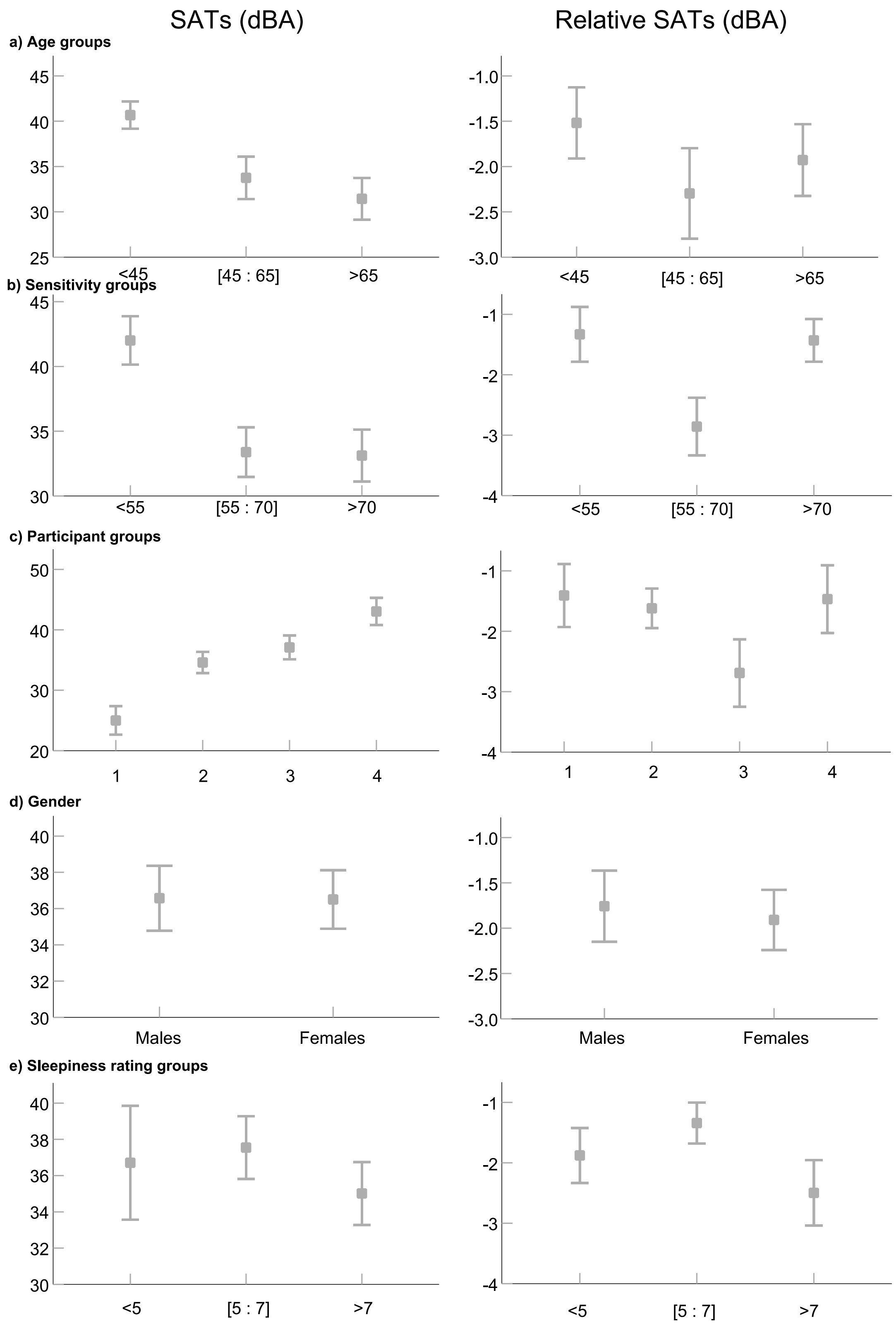


Figure 6.4. The effect of different factors on absolute sleep acceptability threshold levels, SATs (on the left-hand side) and relative SATs (on the right-hand side) is shown for a) Age groups in years b) Sensitivity score groups c) Participants groups (1- WFN-sensitive group 2- WFN-insensitive group 3- Quiet rural area group 4- RTN-sensitive groups) d) gender and e) sleepiness rating groups. Relative SATs were obtained by subtracting the SATs of the baseline stimulus from the SATs of the other stimuli for each participant. The lower the absolute and relative SATs value, the less acceptable for sleep the noise is as compared to the baseline stimulus. The error bars represent 95% CIs.

Appendix F shows detailed differences between absolute and relative SATs for different non-acoustic factor groups at different degrees of modulation depth. The differences between SATs and relative SATs were more evident at high modulation depth and tonal audibility. A maximum average relative SATs of 6.0 ± 1.7 dBA (mean ± CIs) could be obtained for participants with medium sensitivity at a modulation depth of 10 dBA. Absolute SATs also varied significantly between participant groups. For example, they were 24 ± 2.5 dBA (mean ± CIs) for WFN-sensitive participants, while they were 43 ± 2.8 dBA (mean ± CIs) for RTN-sensitive participants.

6.2.2. Machine learning models

6.2.2.1. Absolute SAT models

The performance of absolute and relative SAT models for the swish noise are shown in Tables 6.1 and 6.2, respectively. Neural networks had the best performance as indicated by a high R^2 and low RMS, compared to other machine learning models used for absolute and relative SATs. The contribution of these factors was further investigated by the inclusion and exclusion of factors, respectively. Noise sensitivity was the highest correlated factor with absolute and relative SATs. It was further found that age and noise sensitivity can explain most variability for absolute and relative SATs. Adding more factors either reduced or did not change the accuracy of the models. However, adding the participant group as a factor improved the performance of the neural network model for relative SATs.

Table 6.1. The square of correlation (R^2) and root mean square error (RMSE) of four different machine learning models by including different factors as predictors for the sleep acceptability threshold levels (SATs) of swish noise. Numbers inside the brackets indicate RMSE, while the outside number represents R^2. 10-fold cross-validation was used in all these models. Fully connected neural network models were used with ten neurons in two hidden layers. Gaussian Process Regression with the exponential function was used as it had the best accuracy for predicting SATs.

Model No.	Factors included	Gaussian Process Regression	Boosted trees	Bagged trees	Neural networks
1	Age	0.62(6.84)	0.40(8.66)	0.48(8.00)	0.70(8.66)
2	Noise sensitivity	0.63(6.78)	0.48(8.03)	0.53(7.67)	0.75(7.94)
3	Participant group	0.23(9.77)	0.21(9.88)	0.23(9.79)	0.61(9.22)
4	Age and participant group	0.83(4.65)	0.60(7.03)	0.27(9.52)	0.92(4.42)
5	Age and noise sensitivity	0.87(3.98)	0.75(5.58)	0.70(6.14)	0.96(3.61)
6	Age, noise sensitivity and participant group	0.87(3.98)	0.77(5.30)	0.27(9.55)	0.95(3.67)
7	Age, noise sensitivity and modulation depth	0.38(8.77)	0.70(6.15)	0.43(8.43)	0.60(10.1)
8	Age, noise sensitivity, sleepiness rating	0.78(5.17)	0.72(5.90)	0.28(9.48)	0.84(6.20)
9	Age, noise sensitivity, gender, participant group	0.87(3.98)	0.78(5.26)	0.63(6.78)	0.95(3.81)
10	Age, noise sensitivity, gender, participant group, modulation depth and sleepiness rating	0.66(6.53)	0.74(5.70)	0.27(9.50)	0.86(5.67)

Table 6.2. The square of correlation (R^2) and root mean square error (RMSE) of four different machine learning models by including different factors as predictors for the relative sleep acceptability threshold levels (SATs) of swish noise. Numbers inside the brackets indicate RMSE, while the outside number represents R^2. 10-fold cross-validation was used in all these models. Fully connected neural network models were used with ten neurons in two hidden layers. Gaussian Process Regression with the exponential function was used as it had the best accuracy for predicting relative SATs.

Model No.	Factors included	Gaussian Process Regression	Boosted trees	Bagged trees	Neural networks
1	Age	0.25(4.05)	0.15(4.29)	0.17(4.25)	0.32(4.20)
2	Noise sensitivity	0.11(4.39)	0.10(4.42)	0.11(4.41)	0.47(4.70)
3	Participant group	0.00(4.67)	0.00(4.67)	0.00(4.67)	0.26(4.64)
4	Age and participant group	0.25(4.04)	0.17(4.25)	0.03(4.60)	0.26(6.93)
5	Age and noise sensitivity	0.26(4.02)	0.21(4.16)	0.18(4.23)	0.52(5.29)
6	Age, noise sensitivity and participant group	0.26(4.02)	0.21(4.16)	0.04(4.56)	0.74(3.32)
7	Age, noise sensitivity and modulation depth	0.07(4.50)	0.18(4.21)	0.15(4.29)	0.31(4.12)
8	Age, noise sensitivity, sleepiness rating	0.37(3.70)	0.17(4.24)	0.01(4.76)	0.41(4.58)
9	Age, noise sensitivity, gender, participant group	0.26(4.02)	0.22(4.11)	0.18(4.22)	0.54(3.24)
10	Age, noise sensitivity, gender, participant group, modulation depth and sleepiness rating	0.22(4.12)	0.20(4.16)	0.05(4.54)	0.32(5.51)

6.2.2.2. Relative importance of non-acoustic factors

The ensemble of bagged trees was used to identify the most important parameters affecting these responses, as shown in Figure 6.5. These factors were ordered in terms of their importance to absolute and relative SATs as shown in Table 6.3. Non-acoustic factors such as noise sensitivity, age, sleepiness ratings and participant group were consistently more important than modulation depth for the perception of absolute and relative SATs of broadband AM WFN.

Table 6.3. The importance rank order of acoustic and non-acoustic factors associated with broadband AM WFN on absolute and relative sleep acceptability threshold levels (SATs).

Rank	1	2	3	4	5	6
SATs	noise sensitivity	age	sleepiness ratings	participant group	gender	modulation depth
Relative SATs	noise sensitivity	age	participant group	sleepiness ratings	modulation depth	gender

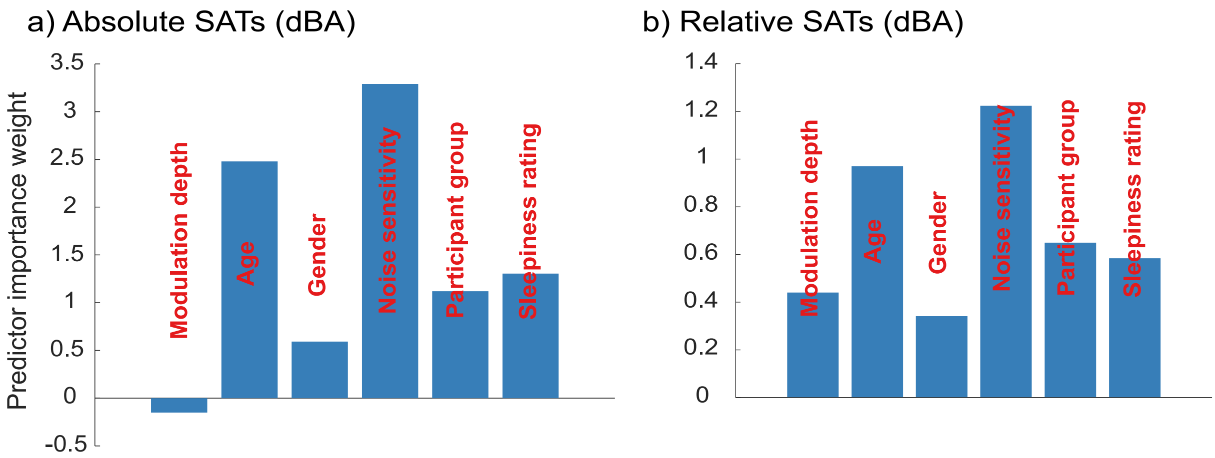


Figure 6.5. The relative importance of acoustic and non-acoustic factors associated with swish noise on a) absolute sleep acceptability threshold levels (SATs) b) relative SATs.

6.3. Discussion

This chapter showed some practical results related to the correlation of different noise weightings with absolute and relative SATs due to the swish noise. It has been argued that the use of A-weighting could be inappropriate for WFN (Hansen et al., 2017 and 2020). In the context of broadband AM WFN, it was shown that other weightings such as the Z-, G-, and C-weighted absolute and relative SATs could be more correlated with modulation depth than the

A-weighted absolute and relative SATs. Therefore, it would be important to use different metrics when assessing the swish noise for compliance.

Relative SATs depended on SAT ranges of the baseline stimulus. The SATs of the baseline stimulus were grouped into three different groups. It was noted that participant responses for baseline SAT levels above 40 dBA showed less variation in relative SATs when broadband AM was present. An explanation of this phenomenon could be that psychoacoustic differences between stimuli can be hardly differentiated at high SPLs. For example, the discrimination of noise intensity decreases at high SPLs (Jesteadt et al., 1977). This is also consistent with the Weber–Fechner law (Fechner et al., 1966). Hünerbein et al. (2013) obtained mean variations from the baseline stimulus of 4.1, 3.6, and 2.2 dBA for stimuli that had an MD between 2 and 12 dBA and modulation frequency of 0.8 Hz played at levels of 30, 35 and 40 dBA, respectively.

Relative SATs of 6.8 ± 1.8 dBA (mean ± CIs) were obtained for the baseline SAT range below 30 dBA. The renewable UK (2013) proposed a penalty scheme for swish noise, which starts at 3 dBA at an MD of 3 dBA with a gradual increase to a maximum value of 5 dBA at an MD of 10 dBA based on the results of Hünerbein et al. (2013) and Ioannidou et al. (2016). Bowdler et al. (2018) suggested that this penalty scheme could be improved by considering the overall SPL of WFN. This study showed that different SATs of WFN baseline stimulus could affect relative SATs. This can encourage developers of noise guidelines to consider other responses such as acceptability for sleep and expected SPLs when assessing WFN.

The results from this project could be used to derive penalties for broadband AM. The method presented by Virjonen et al. (2019) could result in different penalties as these researchers did not consider individual differences between participants.

6.4. Summary

Acoustic and non-acoustic factors were associated with absolute and relative SATs due to broadband amplitude modulated WFN. Modulation depth had significant effects on absolute and relative SATs. In particular, the increase of modulation depth decreased both absolute and relative SATs. Relative SATs varied depending on the baseline SAT range and modulation depth. Different WFN level weightings were also used to show their correlation with absolute and relative SATs. The lowest correlations between the absolute and relative SATs were obtained when using the A-weighted levels, while higher correlations were obtained when using the Z-, C- and G-weighted levels. Therefore, it would be helpful to consider more suitable metrics when assessing broadband amplitude modulated WFN. Age and noise sensitivity explained most variability for absolute and relative SAT models due to the swish noise using all tested machine learning models. Bagged trees were used to identify the most important parameters for absolute and relative SATs. While the type of noise did not affect absolute SATs using A-weighted levels, it affected relative SATs but was less important than non-acoustic factors except for gender.

Chapter 7 Sleep acceptability of wind farm noise in the presence of different masking background noise types

7.1. Introduction

Masking background noise (BGN) can be an important WFN acoustic character affecting perceived responses such as annoyance and sleep acceptability (Bolin, Nilsson, & Khan, 2010). It has been shown that it influences the detection of WFN. Bolin et al. (2010) found that natural masking sounds affected both the detection and perceived loudness of WFN. The detection thresholds of WFN were between signal-to-noise ratios (SNRs) -8 to -12 dBA for natural noise sources such as trees and sea waves. They also found that BGN could affect loudness at SNRs up to 2 dBA. Renterghem et al. (2013) showed that the detection of WFN can contribute to increasing perceived annoyance. They also found that the WFN detection threshold was at an SNR below -20 dBA for highway noise and a lower SNR for road traffic noise (RTN).

Previous surveys have reported that BGN can affect WFN annoyance. Pedersen et al. (2010) surveyed the suitable sites for wind farms and suggested that they should be selected carefully as the visibility and attitudes of wind turbines have a major effect on annoyance. In addition, RTN-exposed sites are preferable for wind farms with lower SPLs, particularly when RTN has SPLs of 20 dB higher than WFN. Therefore, identifying the type of background conditions (e.g., Masking BGN type and level) that makes the perception of WFN less annoying can be beneficial for pre-construction planning of wind farms.

Therefore, this chapter examines the effect of masking BGN on absolute and relative sleep acceptability threshold levels (SATs) using three masking noise types at three different SNRs described in Section 3.1.3.2. A novel non-focused listening test was undertaken with design parameters introduced in Chapter 3 for masking BGN non-focused listening tests.

The relative importance of acoustic and non-acoustic factors to the perception of masking BGN could be another area of interest. It shows the factors that could be targeted for assessment and mitigation strategies. High predictive models can be more reliable in identifying the most important factors as compared to statistical models which could provide the advantages of drawing inferences (underlying mechanisms) (Bzdok et al., 2018). The relative contribution of WFN acoustic characteristics and non-acoustic factors in the presence of masking BGN to absolute and relative WFN SATs is also presented in this chapter.

7.2. Results

7.2.1. Statistical model

7.2.1.1. Effects of acoustic characteristics

The effects of masking BGN type and SNR on absolute and relative SATs are shown in Figure 7.1. Significant differences of absolute SATs were found between noise types only (F (3, 457) = 19, $p < 0.001$). Short-range RTN was significantly different from all other noise types except for RC curve noise (absolute SAT difference in dBA vs long-range RTN 2.03 ± 0.50, $p < 0.001$; vs Pink noise 3.03 ± 0.44, $p < 0.001$; vs RC noise 0.86 ± 0.33, $p = 0.057$).

It can be noticed that there is a variability in the results on the y-axis. This is expected as the outcome is a subjective measure and depends on different factors (i.e., not only the factors on x-axes). Therefore, using one factor only in the x-axis showed high variability for the absolute and relative SATs. The results depend on other factors as can be seen form Figure 7.2 such as the age, sensitivity score, participants groups, gender and sleepiness rating.

Relative SATs were obtained by subtracting the absolute SATs for the WFN baseline stimulus from the absolute SATs for WFN plus masking BGN (i.e., a positive value indicates that the WFN plus masking BGN is more acceptable for sleep than WFN only). Relative SATs showed no significant noise type and SNR interaction effects (F (6, 425) = 0.77, p = 0.597) but significantly increased with increasing SNR (F (2, 234) = 3.1, p = 0.047) and showed statistically significant differences between noise types (F (3, 422) = 21.5, p < 0.001).

Post hoc Bonferroni tests were done to show differences in relative SATs between noise types and their SNRs. WFN masked with short-range RTN was more acceptable for sleep (higher relative SATs) compared to all other noise types (absolute SAT difference in dBA (mean ± SE) vs Long-range RTN 1.67 ± 0.41, p = 0.001; vs Pink noise 3.14 ± 0.42, p < 0.001; vs RC noise 1.15 ± 0.36, p = 0.017). WFN masked with pink noise was less acceptable for sleep (lower relative SATs) compared to all other noise types (absolute difference in dBA vs long-range RTN -1.47 ± 0.36, p = 0.001; vs Short-range RTN -3.14 ± 0.42, p < 0.001; vs RC noise -1.99 ± 0.36, p < 0.001).

The SNR showed statistically significant differences between the relative SATs (p = 0.047). However, a posthoc analysis showed that it was only (SNR=0) that was different from (SNR = -10), where a decrease in the SNR (i.e. WFN is more masked) decreased SATs (i.e., the noise was less acceptable for sleep). The relative difference in dBA between SNRs of -10 and 0 was (mean ± SE) -1.88 ± 0.76, p = 0.043; vs (SNR = 5) -0.82 ± 0.64, p > 0.05).

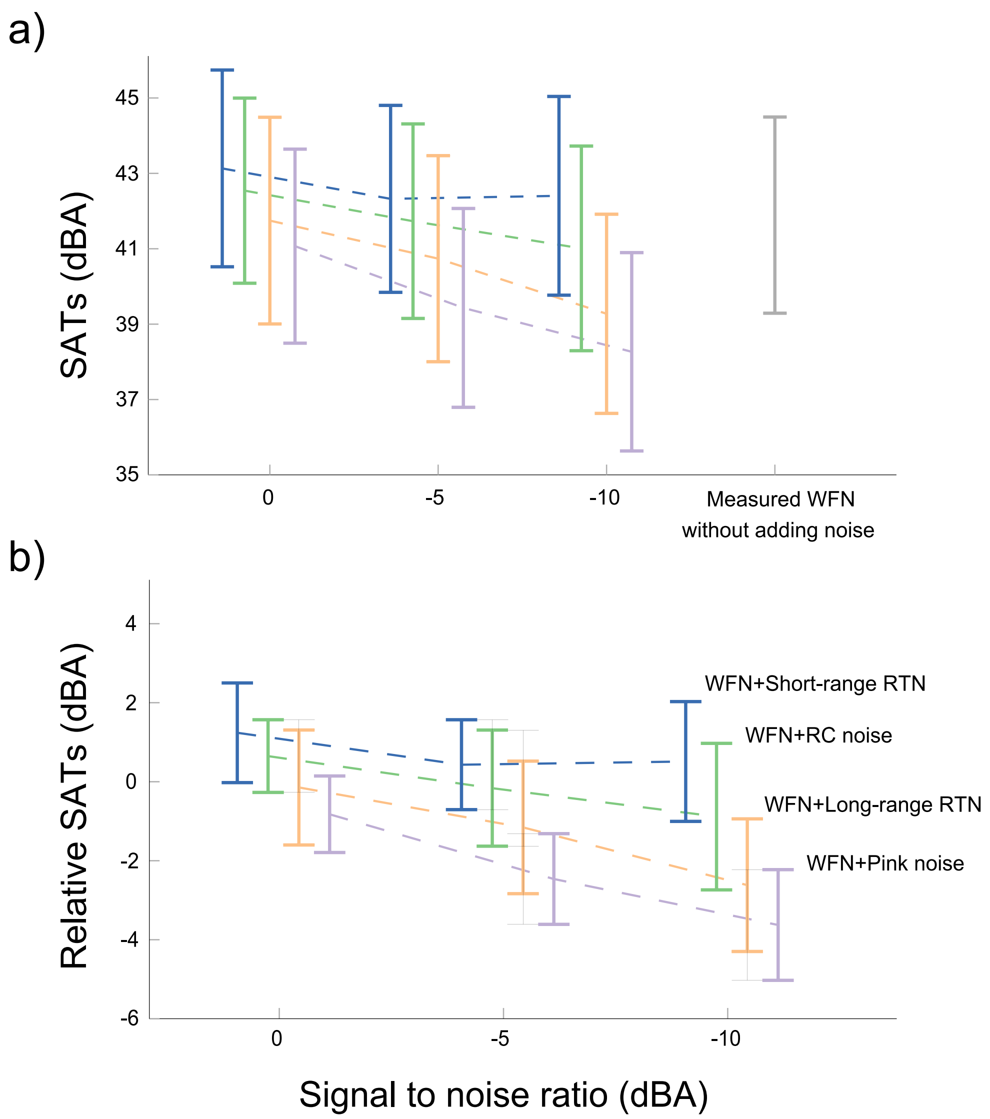


Figure 7.1. Dose-response relationship for absolute and relative sleep acceptability threshold levels (SATs) for different types of masking noise and signal-to-noise ratios (SNRs). Relative SATs were obtained by subtracting the SATs of the baseline stimulus from the SATs of other stimuli for each participant. The lower the absolute and relative SATs value, the less acceptable for sleep the noise is as compared to the baseline stimulus. The error bars represent 95% CIs.

7.2.1.2. Effects of non-acoustic factors

The effects of non-acoustic factors on absolute and relative SATs for WFN in the presence of masking noise are shown in Figure 7.2. The younger participants had higher SATs. Groups of participants also affected responses. Participants with higher sleepiness ratings had lower SATs.

Relative SATs of tonal AM WFN due to non-acoustic factor groups showed that WFN sensitive and insensitive groups found noise less acceptable for sleep compared to tonal AM WFN. Older participants had lower absolute and relative SATs compared to the younger age group.

Appendix G shows differences between absolute and relative SATs for different non-acoustic factor groups at different acoustic characteristics of the masking background noise of WFN (i.e. noise type and SNR). Short-range RTN masking increased WFN sleep acceptability by 3.3 ± 1.5 dBA (mean ± CIs) compared to the baseline stimulus (tonal AM WFN) for noise-sensitive groups (i.e. noise became more acceptable for sleep). Unlike other groups, RTN sensitive group found WFN when masked with short-range RTN was more acceptable for sleep compared to the baseline stimulus at all SNRs. They also had high absolute SATs at these conditions at 47 ± 1.5 dBA.

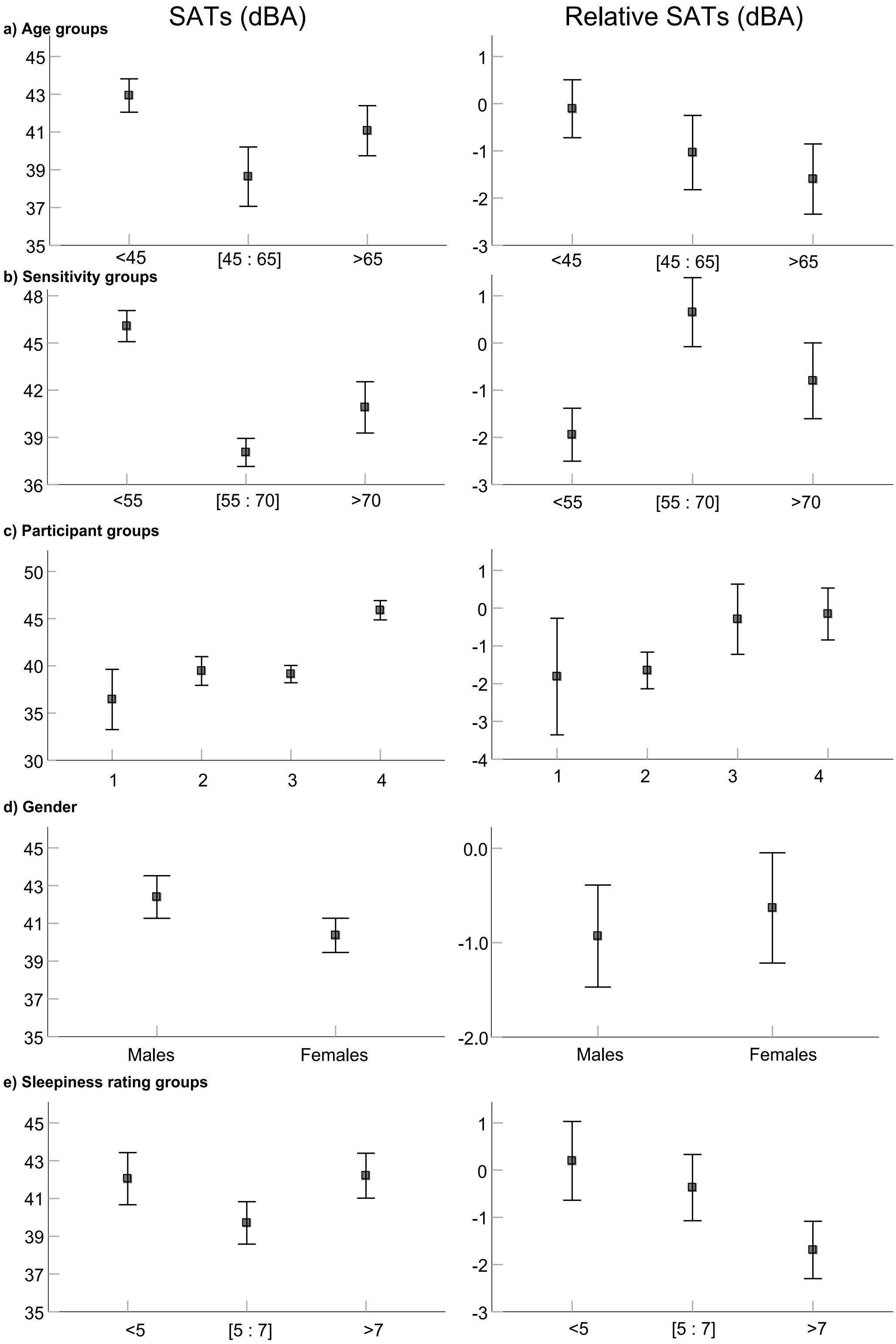


Figure 7.2. The effect of non-acoustic factors on absolute sleep acceptability threshold levels, SATs, in the presence of different masking background noise types and signal to noise ratios (on the left-hand side) and relative SATs (on the right-hand side) is shown for a) Age groups in years b) Sensitivity score groups c) Participants groups (1- WFN-sensitive group 2- WFN-insensitive group 3- Quiet rural area group 4- RTN-sensitive groups) d) gender and e) sleepiness rating groups. Relative SATs were obtained by subtracting the SATs of the baseline stimulus from the SATs of other stimuli for each participant. The lower the absolute and relative SATs value, the less acceptable for sleep the noise is as compared to the baseline stimulus. The error bars represent 95% CIs.

7.2.2. Machine learning models

7.2.2.1. Correlation between factors

The factors included in machine learning models were not correlated with each other. Figure 7.3 shows the correlation matrix for factors included in the models for studying the effects of masking background noise of WFN on both absolute and relative SATs.

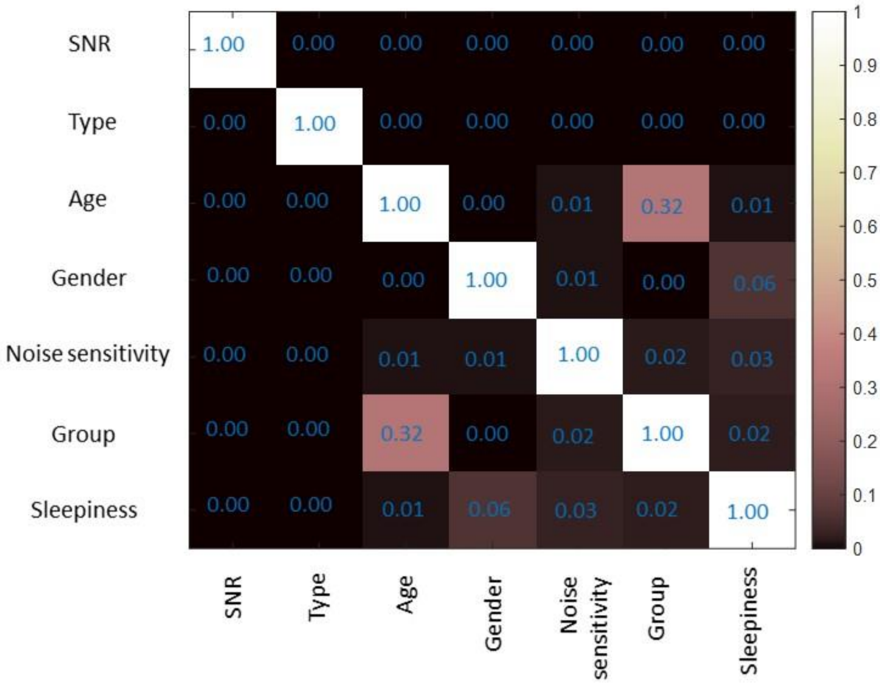


Figure 7.3. Matrix of correlation coefficients (Spearman correlations) for the factors included in machine learning models.

7.2.2.2. Absolute and relative SAT models

For absolute and SAT models, the ensemble of bagged trees had the best accuracy for the regression problem, as shown in Tables 7.1 and 7.2, respectively.

Table 7.1. Comparison between different models of absolute sleep acceptability threshold levels (SATs) of WFN in the presence of masking noise. 10-fold-cross validation was used for all the models. Factors included were the type of background noise (BGN), signal-to-noise ratio (SNR), age, gender, noise sensitivity, participant group and sleepiness ratings. The comparison considered the square of the correlation, mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE).

Model No.	Model	R^2	RMSE	MSE	MAE
1	Linear Regression [linear]	0.23	7.58	57.42	5.97
2	Linear Regression [Interactions linear]	0.38	6.81	46.35	5.26
3	Linear Regression [Robust linear]	0.22	7.62	58.06	5.94
4	Stepwise linear Regression [Stepwise linear]	0.47	6.30	39.66	4.87
5	Tree [Fine tree]	0.83	3.56	12.66	2.72
6	Tree [Medium tree]	0.75	4.29	18.41	3.28
7	Tree [Coarse tree]	0.59	5.56	30.97	4.37
8	Linear SVM	0.20	7.74	59.97	2.91
9	Quadratic SVM	0.39	6.76	45.73	4.96
10	Cubic SVM	0.52	5.97	35.72	4.39
11	Fine Gaussian SVM	0.02	8.55	73.03	6.60
12	Medium Gaussian SVM	0.45	6.40	41.00	4.53
13	Coarse Gaussian SVM	0.19	7.79	60.62	5.88
14	Ensemble [Boosted trees]	0.76	4.25	18.05	3.39
15	**Ensemble [Bagged trees]**	**0.85**	**3.33**	**11.12**	**2.52**
16	Gaussian Process Regression (GPR) [Squared Exponential GPR]	0.56	5.74	32.98	4.17
17	Exponential GPR	0.46	6.36	40.39	4.66
18	Matern 5/2 GPR	0.54	5.87	34.49	4.29
19	Rotational Quadratic GPR	0.56	5.74	32.98	4.17

Table 7.2. Comparison between different models of absolute sleep acceptability threshold levels (SATs) of WFN in the presence of masking noise. 10-fold-cross validation was used for all the models. Factors included were the type of background noise (BGN), signal-to-noise ratio (SNR), age, gender, noise sensitivity, participant group and sleepiness ratings. The comparison considered the square of the correlation, mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE).

Model No.	Model	R^2	RMSE	MSE	MAE
1	Linear Regression [linear]	0.12	4.50	20.22	3.22
2	Linear Regression [Interactions linear]	0.23	4.21	17.69	3.00
3	Linear Regression [Robust linear]	0.09	4.58	20.94	3.19
4	Stepwise linear Regression [Stepwise linear]	0.26	4.11	16.89	2.88
5	Tree [Fine tree]	0.30	4.00	16.03	3.04
6	Tree [Medium tree]	0.20	4.29	18.39	3.18
7	Tree [Coarse tree]	0.06	4.65	21.58	3.524
8	Linear SVM	0.09	4.56	20.79	3.12
9	Quadratic SVM	0.18	4.33	18.73	2.92
10	Cubic SVM	0.17	4.37	19.08	2.99
11	Fine Gaussian SVM	0.04	4.88	23.79	3.43
12	Medium Gaussian SVM	0.22	4.24	17.97	2.84
13	Coarse Gaussian SVM	0.10	4.54	20.58	3.10
14	Ensemble [Boosted trees]	0.50	3.39	11.51	2.43
15	**Ensemble [Bagged trees]**	**0.51**	**3.33**	**11.11**	**2.43**
16	Gaussian Process Regression (GPR) [Squared Exponential GPR]	0.22	4.24	17.96	2.98
17	Exponential GPR	0.20	4.29	18.39	3.00
18	Matern 5/2 GPR	0.22	4.24	17.95	2.98
19	Rotational Quadratic GPR	0.22	4.24	17.96	2.98

7.2.2.3. Relative importance of non-acoustic factors

Given its high performance, the ensemble of bagged trees was used to identify the most important parameters to absolute and relative SATs. The most important parameters for absolute and relative SATs due to WFN in the presence of masking background noise are shown in Figure 7.4. These factors were ordered in terms of their importance to absolute and relative SATs as shown in Table 7.3. Participant group, noise sensitivity and age were consistently the most important factors to absolute and relative sleep acceptability threshold levels (SATs) of WFN in the presence of masking noise, while SNR was the least important factor.

Table 7.3. The importance rank order of acoustic and non-acoustic factors associated with WFN in the presence of masking noise on absolute and relative sleep acceptability threshold levels (SATs).

Rank	1	2	3	4	5	6	7
SATs	participant group	noise sensitivity	age	sleepiness ratings	noise type	gender	SNR
Relative SATs	noise sensitivity	age	participant group	noise type	sleepiness ratings	gender	SNR

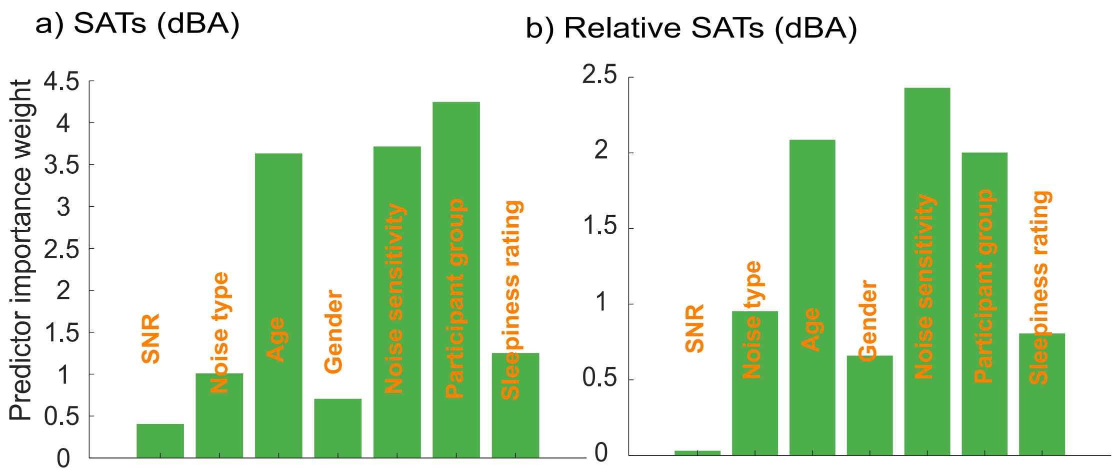


Figure 7.4. The relative importance of acoustic and non-acoustic factors associated with WFN in the presence of masking background noise on a) absolute sleep acceptability threshold levels (SATs) b) relative SATs.

7.3. Discussion

Masking BGN affected the perception of WFN. While some masking noise types increased absolute and relative WFN sleep acceptability, others decreased acceptability. Short-range RTN improved WFN acceptability the most. This effect is consistent with other studies that focused on perceived annoyance rather than sleep acceptability. For example, Pedersen et al. (2010) showed that RTN decreased perceived annoyance responses to WFN with levels (35-40 dBA L_{den}), and only when the SPL of RTN was 20 dBA above that of WFN. While it was initially hypothesised that noise with an RC spectrum shape could be the best masker for WFN, its presence only increased sleep acceptability at SNRs of 0 and -5 dBA.

Low SNRs led to lower relative acceptability for sleep compared to higher SNRs. This could be because WFN may be detected at low SNRs reaching -23 dBA (Van Renterghem et al., 2013). This means that each type of noise becomes more distinct, resulting in dual exposure to two types of noise (Renterghem et al., 2013). Although a reduction in the SNR resulted in a reduced prominence of the low-frequency tonal peak of WFN, this appeared to have a secondary effect on sleep acceptability.

It should be noted that the maximum average of absolute relative SATs obtained in the presence of masking background noise was around 3 dBA. This could be a relatively small difference. However, it should also be noted that some listening test design characteristics could have affected this value. For example, the reproduction method in the current listening test design was through headphones. Some guidelines such as the Nordtest have suggested a psychoacoustic calibration for the headphones (Nordtest, 2002). The headphones should be increased by 3-6 dB to be equally loud with the loudspeaker. However, research to support these level differences is scarce. Nevertheless, the results obtained from current experiments support statistically significant differences between the perception of different noise types.

Wind farms are usually placed in locations with low background noise levels. This could contribute significantly to reduced acceptability of WFN. The results suggest that preferable sites of wind farms could be those already occupied by other sources of noise. Van den Horst (van der Horst, 2007) indicated that the location of wind farms can be an important contributor to the acceptance of wind farms. They found that residents near busy roads were less likely to oppose potential wind farm developments. Wolsink (2007) also suggested that landscapes with natural and cultural values were not preferable for wind farms, while military and industrial areas were suggested as suitable landscapes for wind farms. Torija et al. (2020) found similar results with drone noise, where perceived loudness, annoyance and pleasantness were lower in the presence of short-range RTN at lower SNRs. This study considered the acoustic and non-acoustic factors associated with the perception of WFN in the presence of masking BGN. Short-range RTN improved perception of WFN, while other types of noise decreased its perception.

Besides acoustic characteristics, non-acoustic factors, such as could differentially influence perceived responses to WFN compared to RTN, wind farm visibility, economic factors and noise sensitivity (Michaud et al., 2013). Unpredictability and expectations of noise could also affect perceived responses (Janssen, S. A, 2011). For example, people living near wind farms may expect and be accustomed to lower background noise, compared to residents near RTN and thus have different expectations regarding the surrounding environment.

While the results can be encouraging to have a masking noise of WFN, the type and level of background noise should also be considered. It has been shown that road traffic noise levels above can have negative consequences on human responses (Basner et al., 2011). Further studies are recommended to investigate the safe thresholds of RTN as a masking noise to WFN.

Different masking noise types could be studied to investigate the practicality of other noise environments as masking noise to WFN. Pink noise was included as an example of pleasant noise; however, other types could also be studied. In future studies, other noise types such as natural and industrial noise can be included.

7.4. Summary

Masking background noise types affected sleep acceptability due to wind farm noise (WFN). WFN mixed with short-range road traffic noise (RTN) was suggested to have the highest sleep acceptability threshold levels, SATs (i.e., WFN becomes more acceptable for sleep). Decreasing the signal-to-noise ratio reduced relative SATs, which was an unexpected result. An increase in masking noise was hypothesised to increase acceptability for sleep due to wind farm noise. A possible reason is that the presence of two different noise types reduced acceptability for sleep. Based on previous studies results, it is expected that WFN could still be detected at signal-to-noise ratios down to -10 dBA that were investigated.

Short-range RTN was the most acceptable type of masking BGN in terms of absolute and relative SATs, especially for RTN-sensitive participants. The result that RTN can improve sleep acceptability of wind farm noise at relatively low levels, especially for RTN-sensitive participants, is an important finding that should be considered when siting new wind farm developments. Avoiding rural areas with relatively low ambient noise will help to improve community acceptance of wind farm developments. The type of noise was the most important acoustic parameter to the absolute and relative SATs, followed by signal-to-noise ratios.

Chapter 8 Conclusions and recommendations for future work

8.1. Conclusions

This thesis sought to determine the extent to which subjective responses to WFN are judged differently from road traffic noise (RTN) responses in focused listening tests. This study also included four specifically selected participant groups for the first time to investigate whether the sleep acceptability of WFN characteristics such as tonal and broadband amplitude modulation and masking background noise affect sleep acceptability threshold levels (SATs) in novel non-focused listening tests.

In Chapter 4, WFN acoustic characteristics had significant effects on perceived responses to WFN compared to RTN with different characteristics. For example, tonal AM WFN was perceived as more annoying and yet less loud compared to RTN.

The effects of some WFN acoustic and non-acoustic characteristics on absolute and relative sleep acceptability threshold levels (SATs) were investigated. Studied acoustic characteristics included tonal, and broadband AM and masking noise, while non-acoustic factors included participant groups, noise sensitivity, age, gender and sleepiness ratings. Acoustic and non-acoustic factors affected absolute and relative SATs in a different way.

Different weightings were used for quantifying absolute and relative SATs due to broadband AM WFN in Chapter 5. Modulation depth was not correlated with absolute SATs when using A-weighted SATs as opposed to the significant correlation when using the Z-, G- and C-weightings. Higher and significant correlations were obtained using the Z-, G- and C-weightings for quantifying absolute and relative SATs compared to using the A-weighting.

Modulation depth significantly affected absolute and relative SATs of tonal and broadband AM WFN. Specifically, the increased modulation depth decreased WFN sleep acceptability. This confirmed the significant effects of AM WFN on human responses.

High penalties may be required for tonal and broadband AM WFN at nighttime to reduce potential sleep impacts for residents living near wind farms. Relative SATs of 7.5 ± 2.1 dBA (mean ± CIs) were required for less sleepy participants and high levels of tonal audibility and modulation depth, while they were 6.8 ± 1.8 dBA (mean ± CIs) obtained for broadband amplitude modulated WFN at a modulation depth of 10 dBA and a range below 30 dBA as presented in Chapter 6.

Some non-acoustic factors such as noise sensitivity, age and participant group were consistently more important for sleep acceptability due to tonal, and broadband AM WFN and WFN masking background noise than acoustic characteristics such as modulation depth, modulated frequencies, masking noise type and its signal to noise ratio.

In Chapter 7, the effects of different types and levels of masking background noise on WFN SATs were presented. The type of masking background noise affected WFN SATs. Pink noise masked WFN decreased sleep acceptability more than other masking noise types studied, while short-range RTN as a WFN masking noise increased sleep acceptability for RTN-sensitive groups compared to WFN without masking noise. RTN-sensitive participants also had the highest sleep acceptability for WFN mixed with RTN.

8.2. Implications

Different methods have been proposed to obtain penalties for WFN characteristics but have usually involved comparison with a baseline stimulus. Hongisto et al. (2019) presented WFN penalties by SPLs obtained by finding the difference in SPL required to render the baseline and characteristic of WFN equally annoying. However, using a regression analysis could underestimate the variability of participant ratings. Hünerbein et al. (2013) used pair-wise comparison of WFN stimuli and an adjustment method. This method can be more reliable as it considers variability between participants. This thesis included a baseline stimulus and adjustment method, which is a novel approach that allows absolute and relative acceptable threshold levels to be obtained in one experiment. This is a major advantage because absolute and relative SATs are both valuable responses, as discussed in the previous section.

Relative SATs were presented for different WFN acoustic characteristics as well as non-acoustic factors such as age and noise sensitivity. Developing separate penalties for WFN characteristics could reduce complaints of residents living near wind farms. Low absolute SATs were obtained for older people, WFN groups and highly noise-sensitive groups. However, the percentage of these groups is still unknown (e.g., what is the percentage of noise-sensitive people living near wind farms) and can be investigated by future surveys. Nevertheless, considering the extreme values of absolute and relative SATs potentially helps protect vulnerable people living near wind farms.

People living near wind farms found WFN less acceptable for sleep in terms of low absolute SATs regardless of their acoustic characteristics. It has been shown that these negative subjective responses were associated with long-term sleep disturbance (Beutel et al., 2020). Therefore, those groups are more likely to be affected by WFN.

Some WFN acoustic characteristics such as modulation depth significantly affected absolute and relative SATs. This study also found that masking noise can modify WFN perception. Therefore, better siting of wind farms could be an important future strategy to reduce potential negative impacts of WFN.

Further developments in WFN metrics should be considered. It has been shown that different weightings for swish noise levels had different correlations with SATs. Thus, further investigation of the correlation with participant responses using different level weightings is warranted. Current WFN metrics also do not consider tonal AM but focus exclusively on tonal audibility (Hansen et al., 2017; Hongisto et al., 2018). However, wind farm tones are usually amplitude modulated. Thus, further work is needed to compare metrics of WFN tones at various levels and frequencies, with and without amplitude modulation.

For the first time, this thesis has demonstrated that absolute SATs could be more representative of human responses to WFN and its characteristics compared to relative SATs. For example, people living near wind farms and older participants had lower SATs to tonal AM WFN. However, the relative SATs were not different from the baseline stimulus. An important implication for this is that absolute SATs could be more representative of the feelings of participants. Relative SATs could also be less representative of their overall responses to noise. For example, it has also been shown that there could be no penalty required for WFN-groups; however, their overall SATs of WFN were much less than other groups (i.e., much less acceptance).

While some acoustic factors such as modulation depth have been identified as contributors to decreased sleep acceptability in the presence of WFN characteristics, some non-acoustic factors such as noise sensitivity showed more importance to these responses. It has been shown that some strategies such as community engagement, economic justice and coping strategies can reduce potential negative effects of WFN (Pohl et al., 2018). Information on the relative importance of these factors can also be used to develop targeted interventions aimed at improving the acceptability of WFN for specific groups.

8.3. Recommendations for future work

Further work could investigate how the results would be different if reproduced through a loudspeaker, with accurate calibration to ensure that the reproduced noise is closely matched to the original stimuli. The ecological validity of the reproduction method could also be further investigated in future to check which reproduction method is more ecologically valid. Differences may be found depending on the noise type and level.

Current allowable limits and penalties vary substantially between guidelines. This variation may be because the effect of WFN characteristics is still not well established. More field studies are required to compare the occurrence probability of WFN characteristics. The probability of occurrence of these characteristics can be compared to the responses presented in this thesis to determine whether additional penalties are required.

The effects of sleepiness ratings on relative SATs due to WFN masking noise and broadband AM WFN were different from their impact on relative SATs due to tonal AM WFN. This could be for a couple of reasons. For example, the baseline stimulus included in the masking noise effect test was a tonal AM WFN, so it could still be annoying to some groups. Moreover, the test duration for the swish noise test was around 25 minutes as compared to about 60 minutes for the tonal AM WFN test. Therefore, some factors, such as sleepiness ratings, could have different effects on relative SATs depending on the type of stimuli and test duration. Other listening test designs mentioned in Chapter 2 could have also affected these results. Further experiments may be required to investigate the effects of sleepiness ratings on subjective responses.

References

Alamir, M. A., AlHares, A., Hansen, K. L., & Elamer, A. (2020). The effect of age, gender and noise sensitivity on the liking of food in the presence of background noise. *Food Quality and Preference, 84*. Retrieved from https://doi.org/10.1016/j.foodqual.2020.103950

Alamir, M. A., Hansen, K. L., & Catheside, P. (2021). Penalties applied to wind farm noise: Current allowable limits, influencing factors, and their development. *Journal of Cleaner Production*.

Alamir, M. A., Hansen, K. L., Zajamsek, B., & Catcheside, P. (2019). Subjective responses to wind farm noise: A review of laboratory listening test methods. *Renewable and Sustainable Energy Reviews, 114*, 109317. doi: 10.1016/j.rser.2019.109317

ANSI/ASA-S12.9-2016/Part7. (2017). *Quantities and Procedures for Description and Measurement of Environmental Sound , Part 7 : Measurement of Low-frequency Noise and Infrasound Outdoors and in the Presence of Wind and Indoors in Occupied Spaces*.

Arezes, P. M., Bernardo, C. A., Ribeiro, E., & Dias, H. (2014). Implications of Wind Power Generation: Exposure to Wind Turbine Noise. *Procedia - Social and Behavioral Sciences, 109*, 390–395. doi: 10.1016/j.sbspro.2013.12.478

Ashtiani, P. (2013). Generating a better picture of noise immissions in post construction monitoring using statistical analysis Methodology overview. *5 Th International Conference on Wind Turbine Noise Denver, August*.

Basner, M., Müller, U., & Elmenhorst, E.-M. (2011). Single and combined effects of air, road, and rail traffic noise on sleep and recuperation. *Sleep, 34*, 11–23. doi: Research Support, Non-U.S. Gov't

Bass, J., Cand, M., Coles, D., Davis, R., Irvine, G., Leventhall, G., Levet, T., Miller, S., Sexton, D., & Shelton, J. (2016). *INSTITUTE OF ACOUSTICS IOA Noise Working Group (Wind Turbine Noise) Amplitude Modulation Working Group Final Report A Method for Rating Amplitude Modulation in Wind Turbine Noise Version 1*.

Berglund, B., Lindvall, T., & Schwela, D. H. (1999). *New WHO Guidelines for Community Noise* (Vol. 31, Issue 4). doi: 10.1260/0957456001497535

Betz, A. (1966) Introduction to the Theory of Flow Machines. (D. G. Randall, Trans.) Oxford: Pergamon Press.

Beutel, M. E., Brähler, E., Ernst, M., Klein, E., Reiner, I., Wiltink, J., Michal, M., Wild, P. S., Schulz, A., Münzel, T., Hahad, O., König, J., Lackner, K. J., Pfeiffer, N., & Tibubos, A. N. (2020). Noise annoyance predicts symptoms of depression, anxiety and sleep disturbance 5 years later. Findings from the Gutenberg Health Study. European Journal of Public Health, 30(3), 516–521. doi: 10.1093/eurpub/ckaa015

Bies, D. A., Hansen, C. H., & Howard, C. (2017). Engineering Noise Control. In CRC Press: Vol. Fifth Edit. doi: 10.3397/1.3455051

Blumendeller, E., Kimmig, I., Huber, G., Rettler, P., & Cheng, P. W. (2020). Investigations on low frequency noises of on-shore wind turbines. *Acoustics* (Vol. 2, No. 2, pp. 343-

365). Multidisciplinary Digital Publishing Institute.

Bolin, K., Bluhm, G., & Nilsson, M. E. (2014). Listening test comparing A-weighted and C-weighted sound pressure level as indicator of wind turbine noise annoyance. *Acta Acustica United with Acustica*, *100*(5), 842–847. doi: 10.3813/AAA.918764

Bolin, K., Nilsson, M. E., & Khan, S. (2010). The potential of natural sounds to mask wind turbine noise. *Acta Acustica United with Acustica*, *96*(1), 131–137. doi: 10.3813/AAA.918264

Bowdler, D., Acoustics, D. B., Cand, M., Lea, H., Hayes, M., Partnership, H. M., Irvine, G., & Acoustics, I. (2018). *WIND TURBINE NOISE AMPLITUDE PENALTY CONSIDERATIONS*. *40*, 1–9.

Bray, W., & James, R. (2011). Dynamic measurements of wind turbine acoustic signals, employing sound quality engineering methods considering the time and frequency sensitivities of human perception. *Noise-Con 2011* . Retrieved from https://www.windturbinesyndrome.com/wp-content/uploads/2011/10/Bray-James-NC11-Abstract-76-final-5_20_2011-as-submitted-1.pdf

Brownjohn, J. M. W., & Zheng, X. (2011). *Discussion of human resonant frequency*. 1–6.

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Points of Significance: Statistics versus machine learning. *Nature Methods*, *15*(4), 233–234. doi: 10.1038/nmeth.4642

Caddick, Z. A., Gregory, K., Arsintescu, L., & Flynn-Evans, E. E. (2018). A review of the environmental parameters necessary for an optimal sleep environment. Building and Environment. doi: 10.1016/j.buildenv.2018.01.020

Carlile, S., Davy, J. L., Hillman, D., & Burgemeister, K. (2018). A Review of the Possible Perceptual and Physiological Effects of Wind Turbine Noise. Trends in Hearing, 22(August), 1–10. doi: 10.1177/2331216518789551

Chiu, C. H., Lung, S. C. C., Chen, N., Hwang, J. S., & Tsou, M. C. M. (2021). Effects of low-frequency noise from wind turbines on heart rate variability in healthy individuals. *Scientific reports*, 11(1), 1-12.

Cooper, S. E., & Chan, C. (2017). Subjective assessment of wind turbine noise—The stereo approach. *The Journal of the Acoustical Society of America*, *142*(4), 2700–2700.

Council of Canadian Academies. (2015). *Understanding the evidence: Wind turbine noise* (Issue January).

Cox, R. (2015). *Independent Noise Working Group: Wind Turbine Amplitude Modulation & Planning Control Study Work Package 2.1- Review of Reference Literature* (Issue July).

Crichton, F., & Petrie, K. J. (2015). Health complaints and wind turbines: The efficacy of explaining the nocebo response to reduce symptom reporting. *Environmental Research*, *140*, 449–455. doi: 10.1016/j.envres.2015.04.016

D'Amico, S., Van Renterghem, T., & Botteldooren, D. (2021). Measuring infrasound outdoors with a focus on wind turbines: the benefits of a wind-shielding dome. *Applied Acoustics*, *178*, 108015.

Danish Environmental Protection Agency. (2002). Lab*oratory Evaluation of Annoyance of Low Frequency Noise*. *1*, 66. Retrieved from http://www.miljoindflydelse.dk/udgiv/publications/2002/87-7944-955-7/pdf/87-7944-956-5.pdf

Dittrich, K., & Oberfeld, D. (2009). A comparison of the temporal weighting of annoyance and loudness. *The Journal of the Acoustical Society of America*, *126*(6), 3168–3178.

ETSU-R-97. (1996). *The assessment and rating of noise from wind farms*.

Faller, K. J., Rizzi, S. A., & Aumann, A. R. (2013). *Acoustic performance of a real-time three-dimensional sound-reproduction system*.

Fastl, H., & Zwicker, E. (2001). *Psychoacoustics: Facts and Models*. doi: 10.1007/978-3-642-88163-3

Fechner, G. T., Howes, D. H., & Boring, E. G. (1966). *Elements of psychophysics* (Vol. 1). Holt, Rinehart and Winston New York.

Fields, J. M., De Jong, R. G., Flindell, I. H., Gjestland, T., Job, R. F. S., Kurra, S., Lercher, P., Vallet, M., & Yano, T. (2006). Guidelines for reporting core information from community noise (The International Commission on the Biological Effects of Noise). *Journal of Sound and Vibration*, *206*(June 1997), 685–695.

Fields, J. M., De Jong, R. G., Gjestland, T., Flindell, I. H., Job, R. F. S., Kurra, S., Lercher, P., Vallet, M., Yano, T., Guski, R., Felscher-Suhr, U., & Schumer, R. (2001). Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation. *Journal of Sound and Vibration*, *242*(4), 641–679.

Fowler, K., Koppen, E., & Matthis, K. (2013). International Legislation and Regulations for Wind Turbine Noise. *Proceedings of the 5 Th International Meeting on Wind Turbine Noise. Denver, US., August*, 22.

Frank, T. (1997). ANSI Update: Specification of Audiometers. *American Journal of Audiology*, *6*(3), 29–32.

Fredianelli, L., Carpita, S., & Licitra, G. (2019). A procedure for deriving wind turbine noise limits by taking into account annoyance. *Science of The Total Environment*, *648*, 728–736. doi: 10.1016/j.scitotenv.2018.08.107

Freiberg, A., Schefter, C., Girbig, M., Murta, V. C., & Seidler, A. (2019). *Health e ff ects of wind turbines on humans in residential settings : Results of a scoping review*. *169*(November 2018), 446–463. doi: 10.1016/j.envres.2018.11.032

Gille, L. A., Marquis-Favre, C., & Weber, R. (2016). Noise sensitivity and loudness derivative index for urban road traffic noise annoyance computation. *The Journal of the Acoustical Society of America*, *140*(6), 4307–4317. doi: 10.1121/1.4971329

Gille, L. A., Marquis-Favre, C., & Weber, R. (2017). Aircraft noise annoyance modeling: Consideration of noise sensitivity and of different annoying acoustical characteristics. *Applied Acoustics*, *115*, 139–149. doi: 10.1016/j.apacoust.2016.08.022

Global wind energy production. (2021). Retrieved from https://gwec.net/global-wind-report-2021/

Hafke-Dys, H., Preis, A., Kaczmarek, T., Biniakowski, A., & Kleka, P. (2016). Noise annoyance caused by amplitude modulated sounds resembling the main characteristics of temporal wind turbine noise. *Archives of Acoustics*, *41*(2), 221–232.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*, 23–34.

Hansen, C. H., Doolan, C. J., & Hansen, K. L. (2017). *Wind Farm Noise: Measurement, Assessment and Control* (1st ed.). John Wiley & Sons Ltd.

Hansen, C., & Hansen, K. (2020). Recent Advances in Wind Turbine Noise Research. *Acoustics*, *2*(1), 171–206. doi: 10.3390/acoustics2010013

Hansen, K. L., Zajamsek, B., & Hansen, C. H. (2018). Towards a reasonable penalty for amplitude modulated wind turbine noise. *ACOUSTICS AUSTRALIA*, *46*(1), 21.

Hansen, K., Nguyen, P., Zajamsek, B., Micic, G., & Catcheside, P. (2019). Pilot study on sleep acceptability of low-frequency , amplitude modulated tonal noise. *Proceedings of the 23rd International Congress on Sound and Vibration*, *1*.

Hansen, K., Zajamšek, B., & Hansen, C. (2014). Comparison of the noise levels measured in the vicinity of a wind farm for shutdown and operational conditions. *Internoise 2014*, *5*, 1–11. Retrieved from http://www.acoustics.asn.au/conference_proceedings/INTERNOISE2014/papers/p790.pdf

Hansen, Kristy L., Hansen, C. H., & Zajamšek, B. (2015). Outdoor to indoor reduction of wind farm noise for rural residences. *Building and Environment*, *94*, 764–772.

Hansen, Kristy L., Nguyen, P., Zajamšek, B., Catcheside, P., & Hansen, C. H. (2019). Prevalence of wind farm amplitude modulation at long-range residential locations. *Journal of Sound and Vibration*, *455*, 136–149. doi: 10.1016/j.jsv.2019.05.008

Heutschi, K., Pieren, R., Müller, M., Manyoky, M., Wissen Hayek, U., & Eggenschwiler, K. (2014). Auralization of wind turbine noise: Propagation filtering and vegetation noise synthesis. *Acta Acustica United with Acustica*, *100*(1), 13–24.

Hongisto, V., & Oliva, D. (2018). Annoyance penalty of low-level tonal sounds. *Euronoise*, 1583–1587.

Hongisto, V., Saarinen, P., & Oliva, D. (2019). Annoyance of low-level tonal sounds – A penalty model. *Applied Acoustics*, *145*, 358–361. doi: 10.1016/j.apacoust.2018.09.023

Hünerbein, S. von, King, A., Piper, B., & Cand, M. (2013). *Wind Turbine Amplitude Modulation : Research to Improve Understanding as to its Cause and Effect :Work Package B(2): Development of an AM Dose-Response Relationship Prepared* (Issue December).

Hurtley, C. (2009). *Night noise guidelines for Europe*. World Health Organization(WHO) Europe.

IEC 61400-11:2012 edition 3.0. (2012). Retrieved from https://webstore.iec.ch/publication/5428

Inukai, Y., Nakamura, N., & Taya, H. (2000). Unpleasantness and acceptable limits of low frequency sound. *Jounal of Low Frequency Noise, Vibration and Active Control*, *19*(3), 135–140. doi: 10.1260/0263092001492895

Ioannidou, C., Santurette, S., & Jeong, C.-H. (2016). Effect of modulation depth, frequency, and intermittence on wind turbine noise annoyance. *The Journal of the Acoustical*

Society of America, *139*(3), 1241–1251. doi: 10.1121/1.4944570

Islam, M. R., Mekhilef, S., & Saidur, R. (2013). Progress and recent trends of wind energy technology. *Renewable and Sustainable Energy Reviews*, *21*, 456–468. doi: 10.1016/j.rser.2013.01.007

ISO/TS-15666. (2003). *Acoustics — Assessment of noise annoyance by means of social and socio-acoustic surveys*. *2003*.

ISO 226:2003: Acoustics — Normal equal-loudness-level contours. (2003).

ISO 389-7:2005. (2005).

Jalali, L., Gohari, M., Bigelow, P., McColl, S., & Nezhad-Ahmadi, M. R. (2017). Author response: Letter to the editor: A critical analysis: Why "firm conclusions are not possible." *Environmental Research*, *158*(December 2016), 812–814. doi: 10.1016/j.envres.2017.03.041

Jallouli, J., & Moreau, G. (2009). Virtual Reality as a landscape decision-making tool: The wind turbines' case. *International Journal of Design Sciences and Technology*, *16*(2), 73–88.

Janssen, S. A., Vos, H., Eisses, A. R., & Pedersen, E. (2011). A comparison between exposure-response relationships for wind turbine annoyance and annoyance due to other noise sources. *The Journal of the Acoustical Society of America*, *130*(6), 3746–3753. doi: 10.1121/1.3653984

Jesteadt, W., Wier, C. C., & Green, D. M. (1977). Intensity discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America*, *61*(1), 169–177. doi: 10.1121/1.381278

Jurado, C., Gordillo, D., & Moore, B. C. J. (2019). On the loudness of low-frequency sounds with fluctuating amplitudes. *The Journal of the Acoustical Society of America*, *146*(2), 1142–1149. doi: 10.1121/1.5121700

Kaldellis, J. K., & Zafirakis, D. (2011). The wind energy (r)evolution: A short review of a long history. *Renewable Energy*, *36*(7), 1887–1901. doi: 10.1016/j.renene.2011.01.002

Karthaus, M., Wascher, E., Falkenstein, M., & Getzmann, S. (2019). The ability of young, middle-aged and older drivers to inhibit visual and auditory distraction in a driving simulator task. *Transportation Research Part F: Traffic Psychology and Behaviour*, *68*, 272–284. doi: 10.1016/j.trf.2019.11.007

Kelley, N. D. (1987). A proposed metric for assessing the potential of community annoyance from wind turbine low-frequency noise emissions. *Windpower Conference and Exposition*, 1–11. Retrieved from http://adsabs.harvard.edu/abs/1987wipo.confR....K

Kjellberg, A., Goldstein, M., & Gamberale, F. (1984). An Assesment of dB(A) for Predicting Loudness and Annoyance of Noise Containing Low Frequency Components. *Journal of Low Frequency Noise, Vibration and Active Control*, 10–16.

Krizan, Z., Miller, A., & Hisler, G. (2020). 0276 Does Losing Sleep Unleash Anger? *Sleep*, *43*(Supplement_1), A105–A105. doi: 10.1093/sleep/zsaa056.274

Lee, S., Kim, K., Choi, W., & Lee, S. (2011). Annoyance caused by amplitude modulation of wind turbine noise. *Noise Control Engineering Journal*, *59*(1), 38. doi:

10.3397/1.3531797

Leiva, A., Andrés, P., Servera, M., Verbruggen, F., & Parmentier, F. B. R. (2016). The role of age, working memory, and response inhibition in deviance distraction: A cross-sectional study. *Developmental Psychology*, *52*(9), 1381–1393. doi: 10.1037/dev0000163

Leung, D. Y. C., & Yang, Y. (2012). Wind energy development and its environmental impact: A review. *Renewable and Sustainable Energy Reviews*, *16*(1), 1031–1039. doi: 10.1016/j.rser.2011.09.024

Leventhall, G. (2003). A Review of Published Research on Low Frequency Noise and its Effects. *DEFRA Report*, *May*, 88. Retrieved from http://westminsterresearch.wmin.ac.uk/4141/1/Benton_2003.pdf

Leventhall, G. (2010). Review: Low Frequency Noise. What we know, what we do not know, and what we would like to know. *Noise Notes*, *8*(4), 3–28. doi: 10.1260/1475-4738.8.4.3

Maffei, L., Iachini, T., Masullo, M., Aletta, F., Sorrentino, F., Senese, V. P., & Ruotolo, F. (2013). The effects of vision-related aspects on noise perception of wind turbines in quiet areas. *International Journal of Environmental Research and Public Health*, *10*(5), 1681–1697. doi: 10.3390/ijerph10051681

Maffei, L., Masullo, M., Di Gabriele, M., Votsi, N. E. P., Pantis, J. D., & Senese, V. P. (2015). Auditory recognition of familiar and unfamiliar subjects with wind turbine noise. *International Journal of Environmental Research and Public Health*, *12*(4), 4306–4320. doi: 10.3390/ijerph120404306

Manyoky, M., Wissen Hayek, U., Pieren, R., Heutschi, K., & Grêt-Regamey, A. (2016). Evaluating a visual-acoustic simulation for wind park assessment. *Landscape and Urban Planning*, *153*, 180–197. doi: 10.1016/j.landurbplan.2016.03.013

Manyoky, Madeleine, Wissen Hayek, U., Heutschi, K., Pieren, R., & Grêt-Regamey, A. (2014). Developing a GIS-Based Visual-Acoustic 3D Simulation for Wind Farm Assessment. *ISPRS International Journal of Geo-Information*, *3*(1), 29–48. doi: 10.3390/ijgi3010029

Michaud, D. S., Feder, K., Keith, S. E., Voicescu, S. A., Marro, L., Than, J., Guay, M., Denning, A., Murray, B. J., Frcp, C. D., Weiss, S. K., Villeneuve, P. J., & Berg, F. Van Den. (2013). *Effects of Wind Turbine Noise on Self-Reported and Objective Measures of Sleep*.

Micic, G., Zajamsek, B., Lack, L., Hansen, K., Doolan, C., Hansen, C., Vakulin, A., Lovato, N., Bruck, D., Ching, ·, Chai-Coetzer, L., Mercer, J., & Catcheside, P. (2018). A Review of the Potential Impacts of Wind Farm Noise on Sleep. *Acoustics Australia*. doi: 10.1007/s40857-017-0120-9

Moller, H. (1987). Annoyance of audible infrasound. *Journal of Low Frequency Noise and Vibration*, *6*(1), 1–17. doi: 10.1177/026309238700600101

Møller, H., & Pedersen, C. S. (2011). Low-frequency noise from large wind turbines. *The Journal of the Acoustical Society of America*, *129*(6), 3727–3744. doi: 10.1121/1.3543957

Moorhouse, A., Waddington, D., & Adams, M. (2005). *Proposed criteria for the assessment of low frequency noise disturbance*. *February*.

Morocho-Cayamcela, M. E., Lee, H., & Lim, W. (2019). Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions. *IEEE Access*, *7*(October), 137184–137206. doi: 10.1109/ACCESS.2019.2942390

Nguyen, D. P., Hansen, K., Zajamsek, B., & Catcheside, P. (2020). Evaluation of wind farm noise amplitude modulation synthesis quality. *Applied Acoustics*, *166*, 107349. doi: 10.1016/j.apacoust.2020.107349

NHMRC. (2015). *NHMRC Statement : Evidence on Wind Farms and Human Health*. *February*, 2015. Retrieved from https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/eh57_nhmrc_statement_wind_farms_human_health_0.pdf

Nilsson, M. E. (2007). A-weighted sound pressure level as an indicator of short-term loudness or annoyance of road-traffic sound. *Journal of Sound and Vibration*, *302*(1–2), 197–207. doi: 10.1016/j.jsv.2006.11.010

Nordtest. (2002). Acoustics: Human Sound Perception - Guidelines for Listening Tests. In NT ACOU 111.

NZS 6808 Acoustics – wind farm noise. Standards New Zealand. (2010).

Oerlemans, S. (2013). *Wind Turbine Amplitude Modulation: Research to Improve Understanding as to its Cause and Effect: Work Package A(1): An explanation for enhanced amplitude modulation of wind turbine noise*.

Oliva, D., Hongisto, V., & Haapakangas, A. (2017). Annoyance of low-level tonal sounds – Factors affecting the penalty. *Building and Environment*, *123*, 404–414.

Pawlaczyk-Łuszczyńska, M., Zaborowski, K., Dudarewicz, A., Zamojska-Daniszewska, M., & Waszkowska, M. (2018). Response to Noise Emitted by Wind Farms in People Living in Nearby Areas. *International Journal of Environmental Research and Public Health*, *15*(8), 1575. doi: 10.3390/ijerph15081575

Pedersen, E., van den Berg, F., Bakker, R., & Bouma, J. (2009). Response to noise from modern wind farms in The Netherlands. *The Journal of the Acoustical Society of America*, *126*(2), 634–643. doi: 10.1121/1.3160293

Pedersen, E., van den Berg, F., Bakker, R., & Bouma, J. (2010). Can road traffic mask sound from wind turbines? Response to wind turbine sound at different levels of road traffic sound. *Energy Policy*, *38*(5), 2520–2527. doi: 10.1016/j.enpol.2010.01.001

Perkins, R. A., & Lotinga, M. J. (2015). *A Method To Control Amplitude Modulation in Wind Turbine Noise Within the Uk Planning Regime*. 1–8.

Persson, K., Bjorkman, M., & Rylander, R. (1985). An experimental evaluation of annoyance due to low frequency noise. *Journal of Low Frequency Noise Vibration and Active Control*, *4*(4), 145–153.

Persson Waye, K., Bengtsson, J., Kjellberg, A., & Benton, S. (2001). Low frequency noise “pollution” interferes with performance. *Noise & Health*, *4*(13), 33–49. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12678934

Pieren, R., Heutschi, K., Müller, M., Manyoky, M., & Eggenschwiler, K. (2014). Auralization of wind turbine noise: Emission synthesis. *Acta Acustica United with Acustica*, *100*(1), 25–33. doi: 10.3813/AAA.918683

Pohl, J., Gabriel, J., & Hübner, G. (2018). Understanding stress effects of wind turbine noise – The integrated approach. *Energy Policy*, *112*(October 2017), 119–128.

Poulsen, T. (1991). Influence of session length on judged annoyance. *Journal of Sound and Vibration*, *145*(2), 217–224. doi: 10.1016/0022-460X(91)90588-B

Radun, J., Hongisto, V., & Suokas, M. (2019). Variables associated with wind turbine noise annoyance and sleep disturbance. *Building and Environment*, *150*, 339–348.

Renewable UK (Issue December). (2013).

Ruotolo, F., Senese, V. P., Ruggiero, G., Maffei, L., Masullo, M., & Iachini, T. (2012). Individual reactions to a multisensory immersive virtual environment: The impact of a wind farm on individuals. *Cognitive Processing*, *13*(1 SUPPL).

Sakamoto, S., Yokoyama, S., Yano, H., & Tachibana, H. (2014). Experimental study on hearing thresholds for low-frequency pure tones. In Acoustical Science and Technology (Vol. 35, Issue 4). doi: 10.1250/ast.35.213

Schäffer, B., Pieren, R., Schlittmeier, S. J., & Brink, M. (2018). Effects of different spectral shapes and amplitude modulation of broadband noise on annoyance reactions in a controlled listening experiment. *International Journal of Environmental Research and Public Health*, *15*(5). doi: 10.3390/ijerph15051029

Schäffer, B., Pieren, R., Wissen Hayek, U., Biver, N., & Grêt-Regamey, A. (2019). Influence of visibility of wind farms on noise annoyance – A laboratory experiment with audio-visual simulations. *Landscape and Urban Planning*, *186*(October 2018), 67–78.

Schäffer, B., Schlittmeier, S. J., Pieren, R., Heutschi, K., Brink, M., Graf, R., & Hellbrück, J. (2016). Short-term annoyance reactions to stationary and time-varying wind turbine and road traffic noise: A laboratory study. *The Journal of the Acoustical Society of America*, *139*(5), 2949–2963. doi: 10.1121/1.4949566

Schmidt, J. H., & Klokker, M. (2014). Health effects related to wind turbine noise exposure: A systematic review. *PLoS ONE*, *9*(12), 1–28. doi: 10.1371/journal.pone.0114183

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2016). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, *104*(1), 148–175. doi: 10.1109/JPROC.2015.2494218

Shaughnessy, J. ., Zechmeister, E. ., & Jeanne, Z. (2011). *Research methods in psychology*. McGraw-Hill.

Shinohara, R. T., Sweeney, E. M., Goldsmith, J., Shiee, N., Mateen, F. J., Calabresi, P. A., Jarso, S., Pham, D. L., Reich, D. S., & Crainiceanu, C. M. (2014). Statistical normalization techniques for magnetic resonance imaging. *NeuroImage: Clinical*, *6*, 9–19.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. doi: 10.1098/rstb.2009.0138

Smith, M. G., Ögren, M., Thorsson, P., Hussain-Alkhateeb, L., Pedersen, E., Forssén, J., Morsing, J. A., & Waye, K. P. (2020). A laboratory study on the effects of wind turbine noise on sleep: Results of the polysomnographic WiTNES study. *Sleep*, 1–11. doi: 10.1093/ofid/ofy092/4987343

Søndergaard, B. (2013). Low Frequency Noise from Wind Turbines: Do the Danish Regulations Have Any Impact? *5th International Conference on Wind Turbine Noise, August*, 1–14.

Søndergaard, L. S., & Pedersen, T. H. (2013). Tonality in wind turbine noise . IEC 61400-11 ver . 2 . 1 and 3 . 0 and the Danish / Joint Nordic method compared with listening tests. *5th Int. Meet. Wind Turbine Noise, August*, 2–10.

Stigwood, D., & Bingham, D. (2015). *Independent Noise Working Group: Wind Turbine Amplitude Modulation & Planning Control Study Work Package 5 - Towards a draft AM condition*. *November*.

Subedi, J. K., Yamaguchi, H., Matsumoto, Y., & Ishihara, M. (2009). Annoyance of low frequency tones and objective evaluation methods. *Journal of Low Frequency Noise, Vibration and Active Control*, *28*(2), 81–96. doi: 10.1260/0263092054531000

Szychowska, M., Hafke-Dys, H., Preis, A., Kociński, J., & Kleka, P. (2018). The influence of audio-visual interactions on the annoyance ratings for wind turbines. *Applied Acoustics*, *129*, 190–203. doi: 10.1016/j.apacoust.2017.08.003

Tabe-Bordbar, S., Emad, A., Zhao, S. D., & Sinha, S. (2018). A closer look at cross-validation for assessing the accuracy of gene regulatory networks and models. *Scientific Reports*, *8*(1), 1–11. doi: 10.1038/s41598-018-24937-4

Taghipour, A., Pieren, R., & Schäffer, B. (2019). Short-term annoyance reactions to civil helicopter and propeller-driven aircraft noise: A laboratory experiment. *The Journal of the Acoustical Society of America*, *145*(2), 956–967. doi: 10.1121/1.5090500

TheWorldHealthOrganisation(WHO). (2018). *Environmental noise guidelines for the European region*.

Thorsson, P., Persson Waye, K., Smith, M., Ögren, M., Pedersen, E., & Forssén, J. (2018). Low-frequency outdoor–indoor noise level difference for wind turbine assessment. *The Journal of the Acoustical Society of America*, *143*(3), EL206–EL211. doi: 10.1121/1.5027018

Thorsson, P., Waye, K. P., Ögren, M., Smith, M., Pedersen, E., & Forssén, J. (2019). Creating sound immission mimicking real-life characteristics from a single wind turbine. *Applied Acoustics*, *143*, 66–73. doi: 10.1016/j.apacoust.2018.08.015

Tonin, R., Brett, J., & Colagiuri, B. (2016). The effect of infrasound and negative expectations to adverse pathological symptoms from wind farms. *Journal of Low Frequency Noise Vibration and Active Control*, *35*(1), 77–90.

Torija, A. J., Li, Z., & Self, R. H. (2020). Effects of a hovering unmanned aerial vehicle on urban soundscapes perception. *Transportation Research Part D: Transport and Environment*, *78*, 102195. doi: 10.1016/j.trd.2019.11.024

Vallin, P.-A., Marquis-Favre, C., Bleuse, J., & Gille, L.-A. (2018). Railway noise annoyance modeling: Accounting for noise sensitivity and different acoustical features. *The Journal of the Acoustical Society of America*, *144*(6), 3381–3390. doi: 10.1121/1.5082296

van den Berg, G. P. (2005). The Beat is Getting Stronger: The Effect of Atmospheric Stability on Low Frequency Modulated Sound of Wind Turbines. *Journal of Low Frequency Noise, Vibration and Active Control*, *24*(1), 1–23.

van der Horst, D. (2007). NIMBY or not? Exploring the relevance of location and the politics of voiced opinions in renewable energy siting controversies. *Energy Policy*, *35*(5), 2705–2714. doi: 10.1016/j.enpol.2006.12.012

Van Renterghem, T., Bockstael, A., De Weirt, V., & Botteldooren, D. (2013). Annoyance, detection and recognition of wind turbine noise. *Science of the Total Environment*, *456–457*, 333–345. doi: 10.1016/j.scitotenv.2013.03.095

Virjonen, P., Radun, J., & Hongisto, V. (2019). Annoyance penalty of periodically amplitude-modulated wide-band sound. *The Journal of the Acoustical Society of America*, *In print*. doi: 10.1121/1.5133478

Waye, K. P., & Öhrström, E. (2002). Psycho-acoustic characters of relevance for annoyance of wind turbine noise. *Journal of Sound and Vibration*, *250*(1), 65–73. doi: 10.1006/jsvi.2001.3905

Weinstein, N. D. (1978). Individual differences in reactions to noise: A longitudinal study in a college dormitory. *Journal of Applied Psychology*, *63*(4), 458–466.

Wolsink, M. (2007). Planning of renewables schemes: Deliberative and fair decision-making on landscape issues instead of reproachful accusations of non-cooperation. *Energy Policy*, *35*(5), 2692–2704. doi: 10.1016/j.enpol.2006.12.002

Yeowart, N. S., & Evans, M. J. (1974). Thresholds of audibility for very low frequency pure tones. *The Journal of the Acoustical Society of America*, *55*(4), 814–818.

Yokoyama, S., Kobayashi, T., & Tachibana, H. (2016). Perception of tonal components contained in wind turbine noise. *Inter-Noise 2016. Hamburg*, 2628:2639.

Yokoyama, S. S., & Tachibana, H. (2014). *Perception of low frequency components in wind turbine noise*.

Yokoyama, S., Sakamoto, S., & Tachibana, H. (2013a). Perception of low frequency components contained in wind turbine noise. *5th International Meeting on Wind Turbine Noise*, *August*.

Yokoyama, S., Sakamoto, S., & Tachibana, H. (2013b). Study on the amplitude modulation of wind turbine noise: part 2- Auditory experiments. *Inter-Noise 2013. Austria*, 1–10. Retrieved from http://rethinkpylons.org/wp-content/uploads/2007/03/Internoise-2013-Innsbruck-Austria-15-18-Sept-2013.pdf

Yoon, K., Gwak, D. Y., Seong, Y., Lee, S., Hong, J., & Lee, S. (2016). Effects of amplitude modulation on perception of wind turbine noise. *Journal of Mechanical Science and Technology*, *30*(10), 4503–4509. doi: 10.1007/s12206-016-0918-7

Zagubień, A., & Wolniewicz, K. (2019). The impact of supporting tower on wind turbine noise emission. *Applied Acoustics, 155*, 260-270.

Zajamsek, B., Hansen, K., & Hansen, C. (2014). Investigation of the time dependent nature of infrasound measured near a wind farm. *Internoise 2014*, *7*(9), 1–10.

Zajamšek, B., Hansen, K. L., Doolan, C. J., & Hansen, C. H. (2015). Characterisation of wind farm infrasound and low-frequency noise. *Journal of Sound and Vibration*, 1–15.

Appendices

Appendix A. The verbal instructions before and after the listening test.

Before the listening test:

Hello and thank you for agreeing to take part in this listening test. This test aims to investigate acceptable levels of noise during sleep. Please make yourself comfortable and imagine that you are in your own bed trying to fall asleep.

During this test, you will adjust the volume of the noise by turning the hand-held volume controller. After each adjustment, please give yourself time to try to fall asleep. Then, take the time to decide if you need more adjustments to reach the loudest level that is acceptable for sleep. Imagine that the noise would be present for the whole night that you were trying to fall asleep. Feel free to make as many adjustments as necessary.

We will begin with a practice listening test so that you have a chance to become familiar with the testing procedure and the volume controller. Whilst we would like you to imagine that you are trying to fall asleep, please try your best to remain awake so that you can attend to the noise. After each 3-minute noise sample, there will be an alarm that sounds like this: *Alarm*. You will then record your sleepiness on the touch screen in front of you. When you have finished, please press next and the next noise sample will begin.

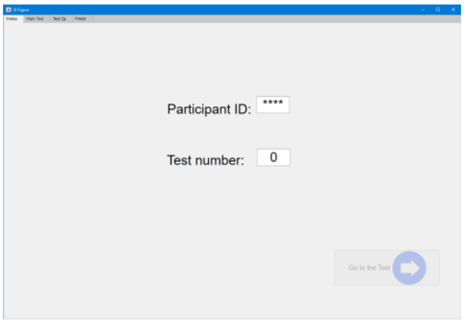
The experimenter will stay in the room with you during the practice test so feel free to ask them any questions at any time. After the practice test, you will do the main listening test which includes between six and seventeen 3-minute samples. We will now begin.

After the listening test:

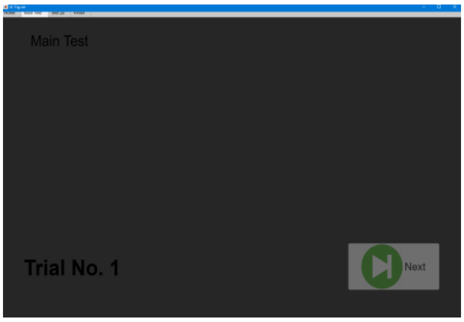
The listening test has now finished. Thank you very much for your time, we really appreciate it. The experimenter can see on the computer in the control room that you have provided responses to all of the noise samples and have finished this task. Please relax for a few moments and the experimenter will return to your bedroom shortly.

Appendix B. The graphical user interface (GUI) used in these experiments.

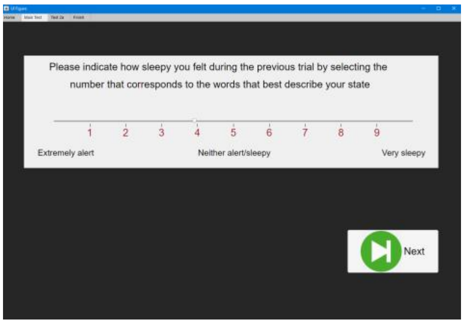
a) Beginning the test



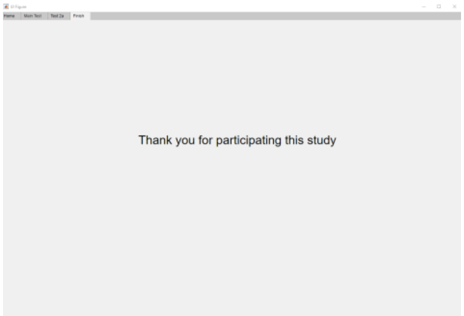
b) During the test



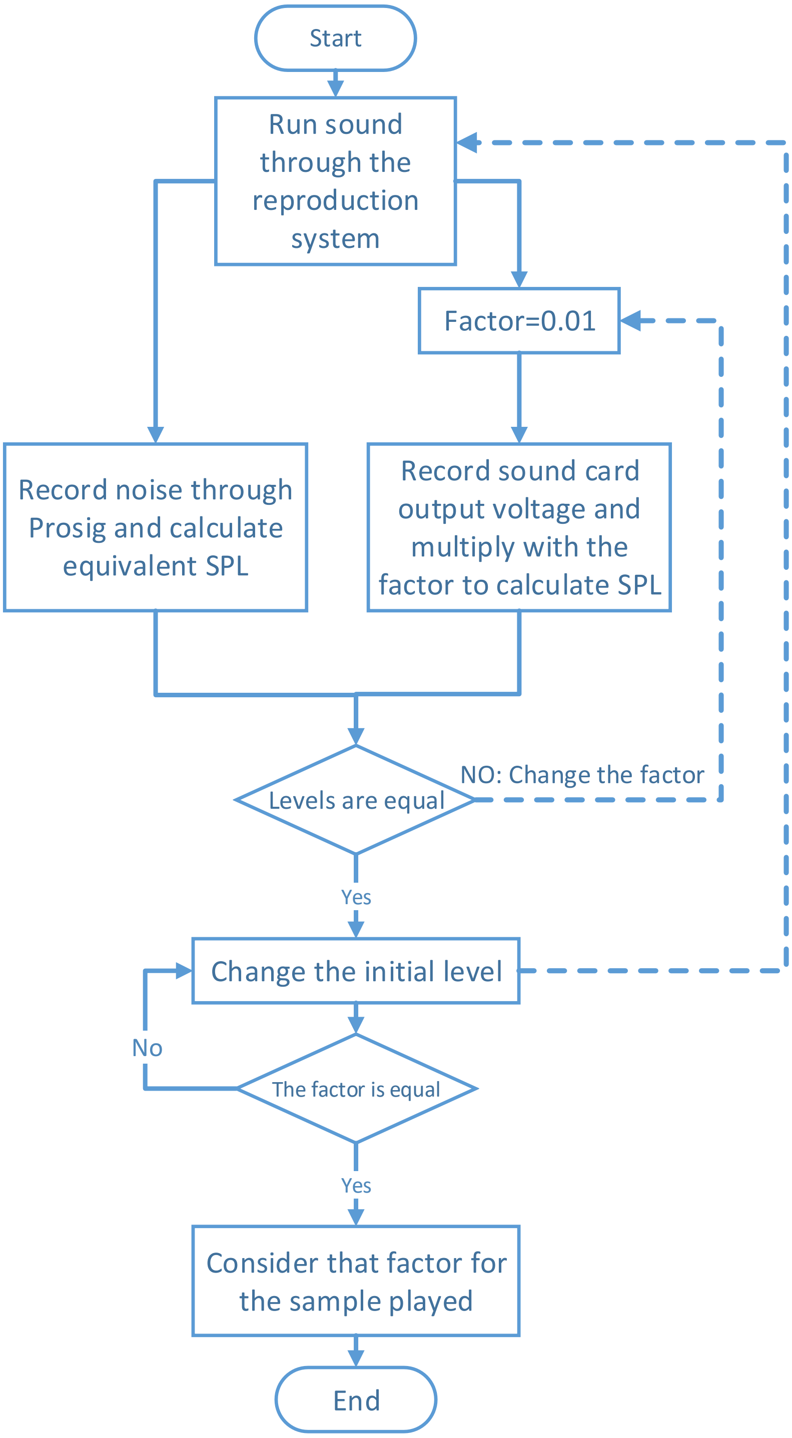
c) After each trial



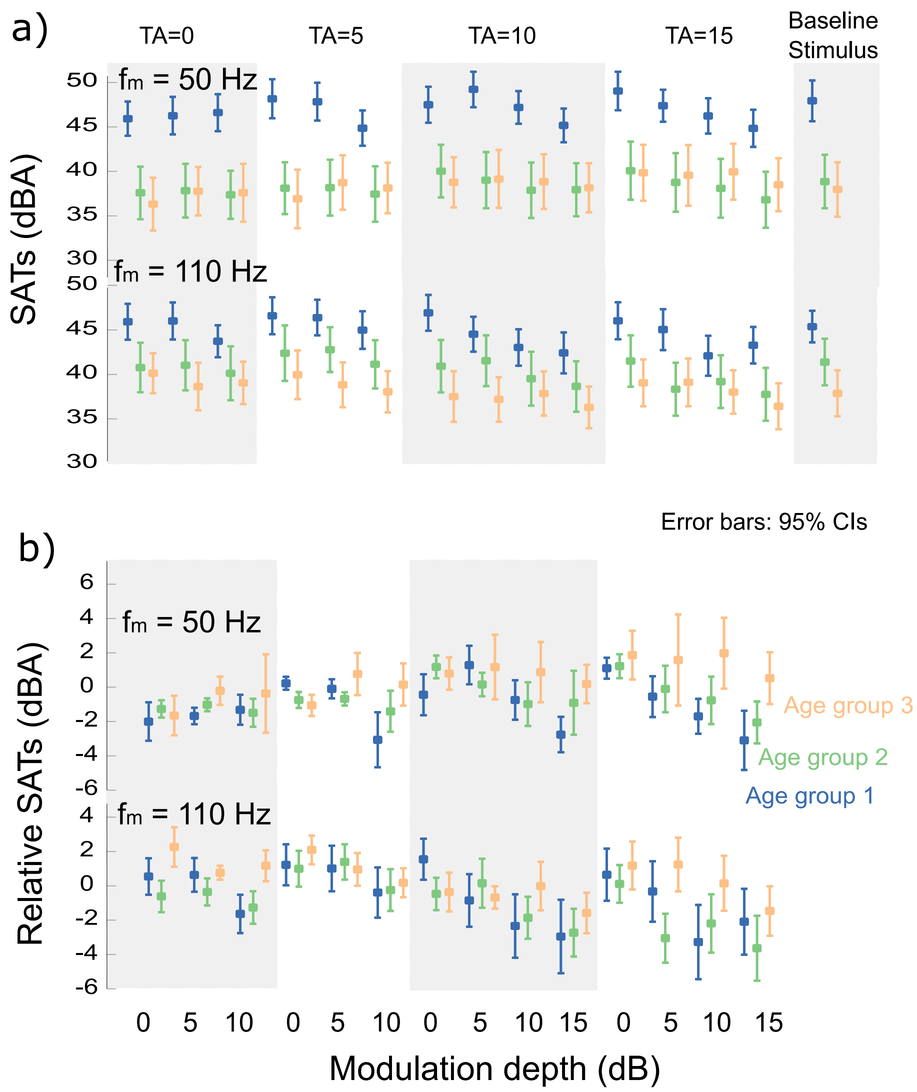
d) The end of the test



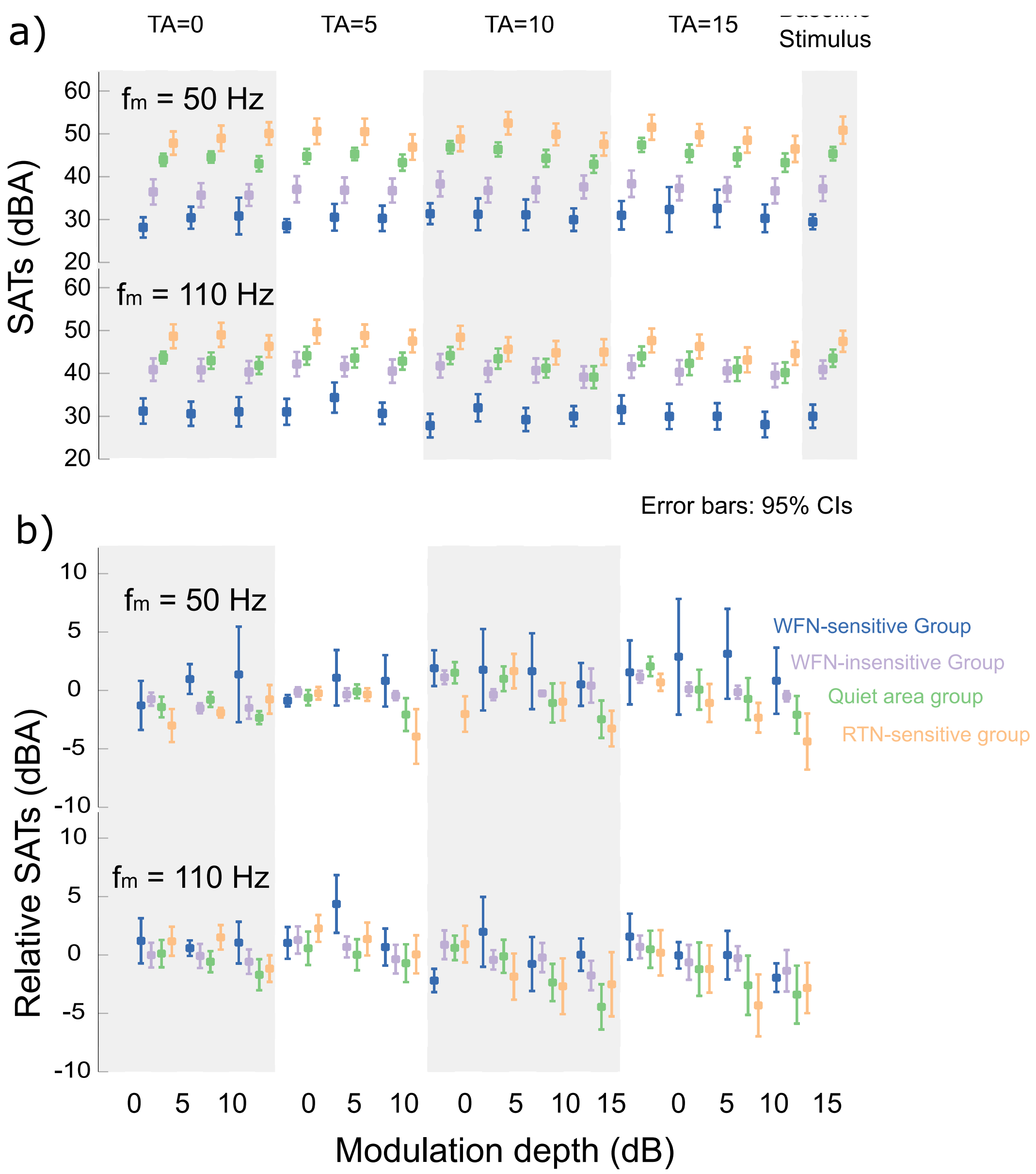
Appendix C. Flow chart showing the calibration procedures for the headphones.



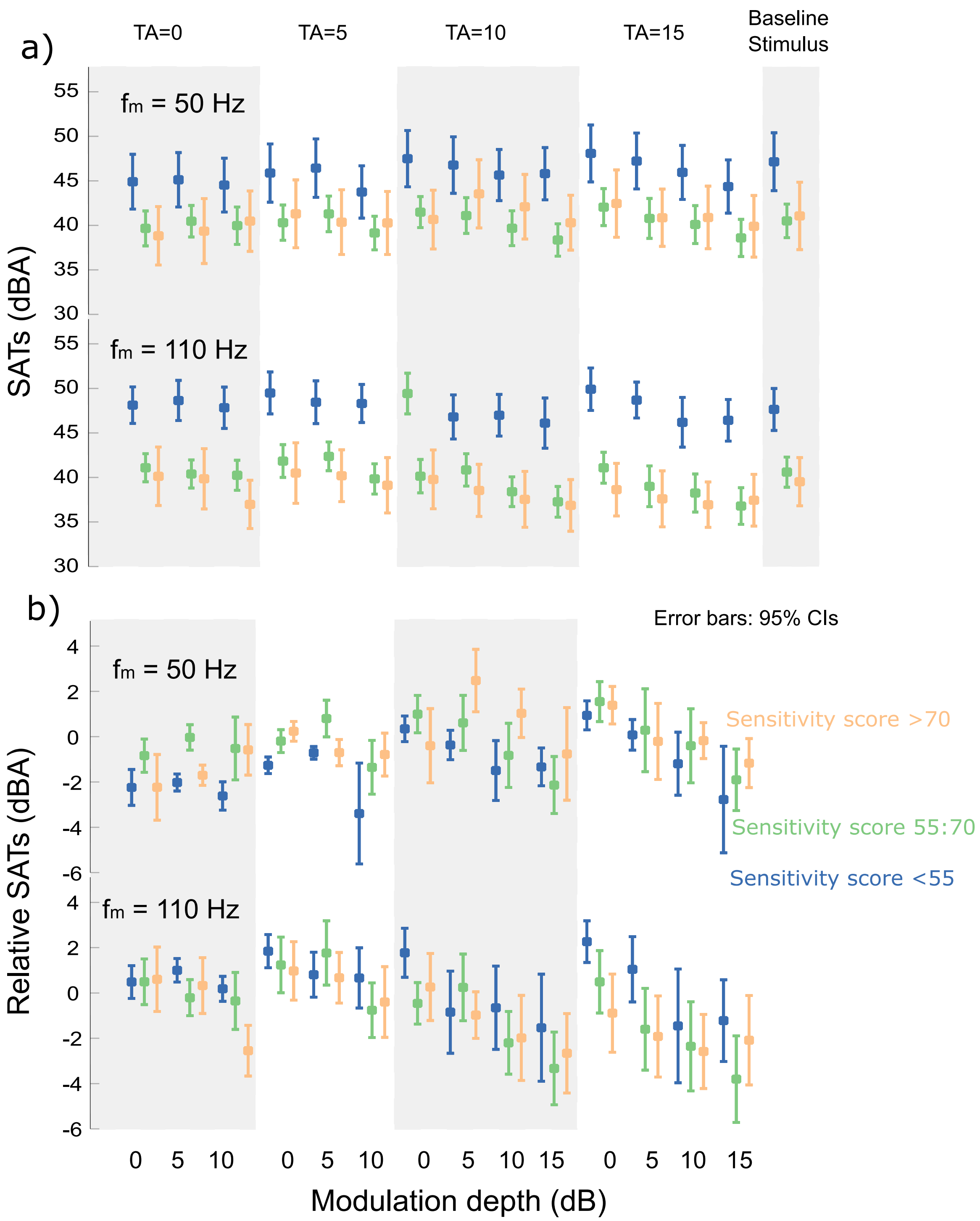
Appendix D. The effect of non-acoustic factor groups at different tonal AM characteristics on a) sleep acceptability threshold levels (SATs) b) relative SATs. AM characteristics included tonal audibility (TA) and modulation depth in dB and modulated frequency (fm) in Hz. The relative SATs were obtained by subtracting the SATs of the baseline stimulus from the SATs of the other stimuli for each participant. The lower the absolute and relative SATs value are, the less acceptable for sleep the noise is as compared to the baseline stimulus.



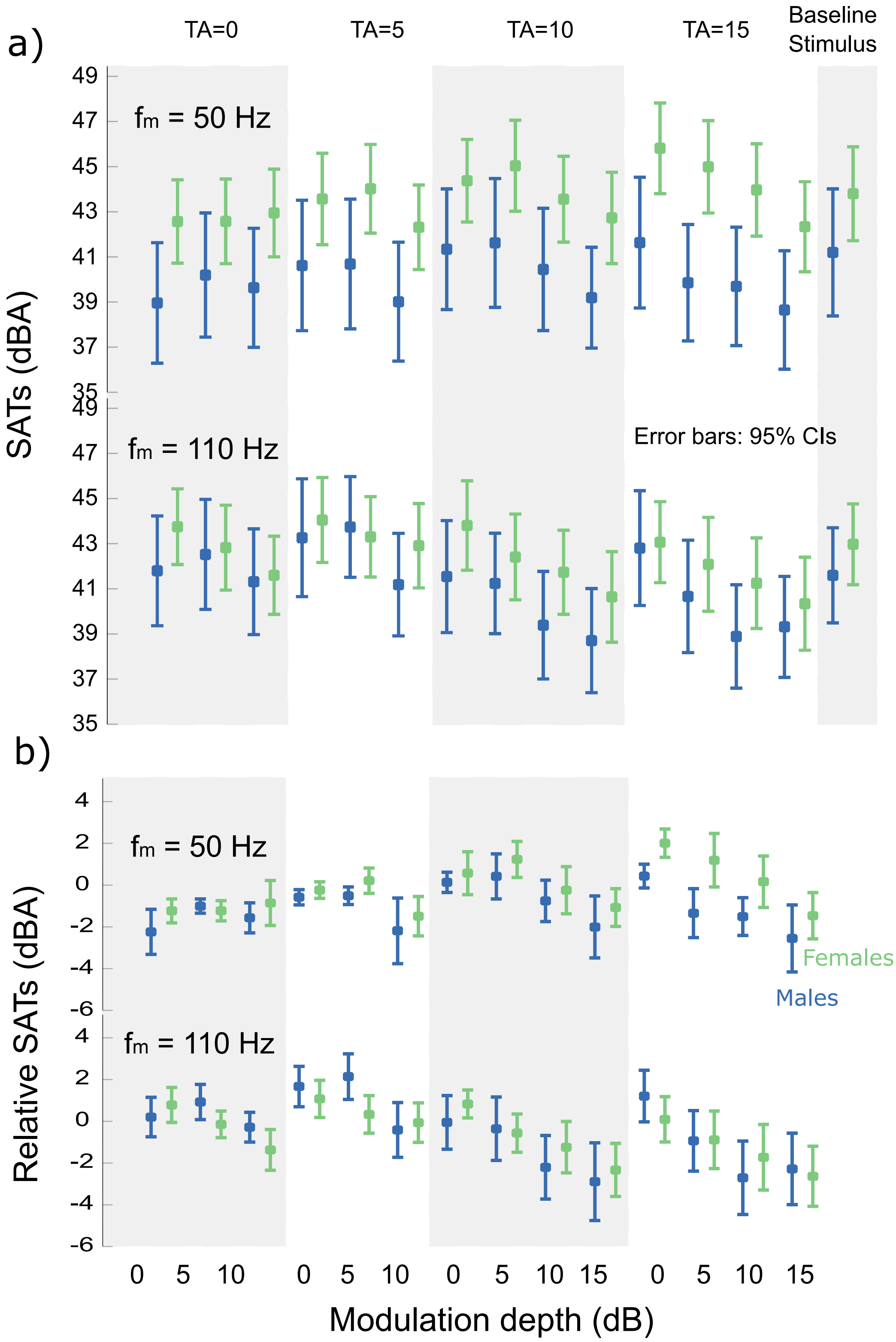
a) Age groups



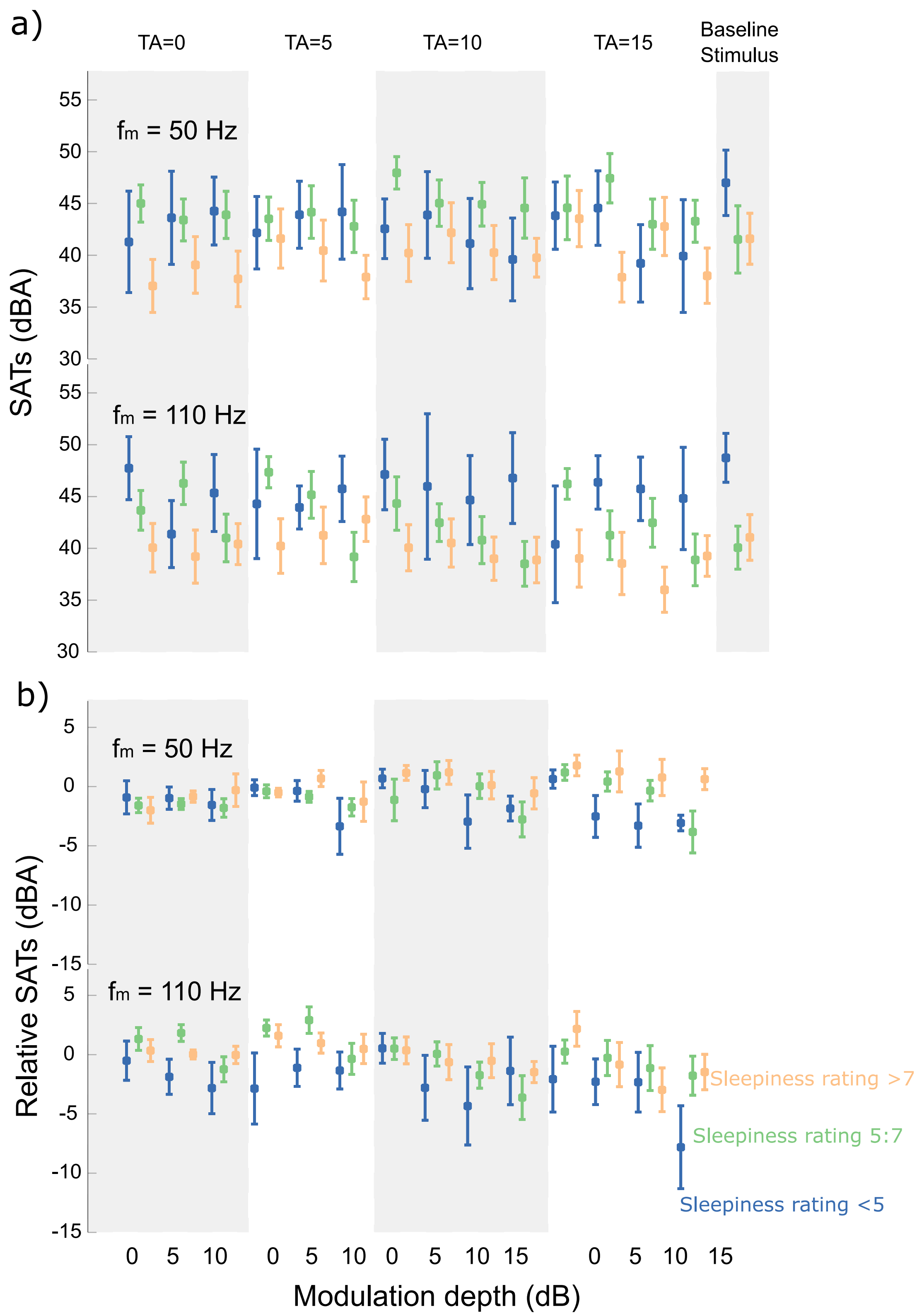
b) Participant groups



c) Self-reported noise sensitivity groups

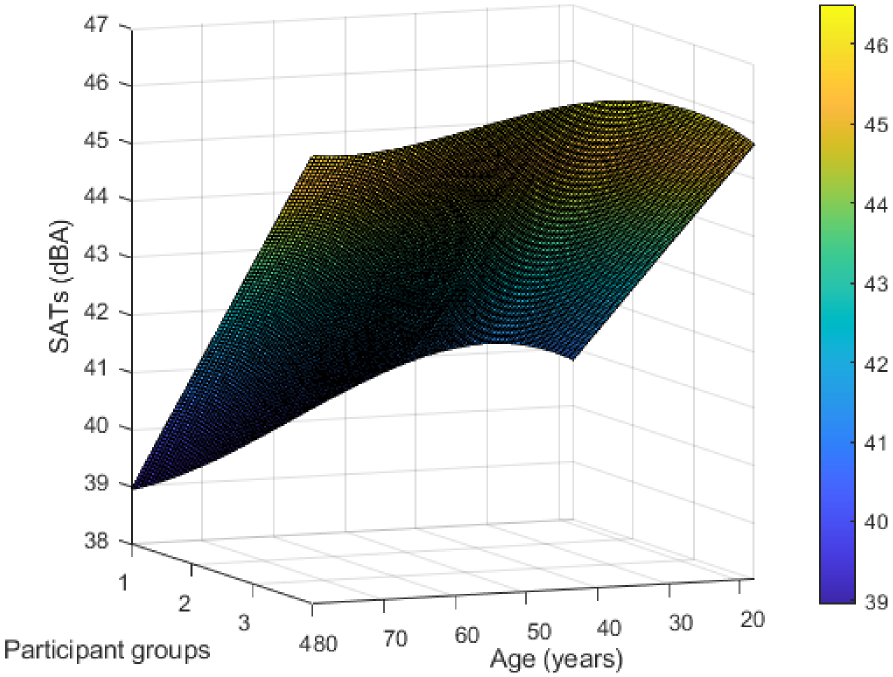


d) Gender groups

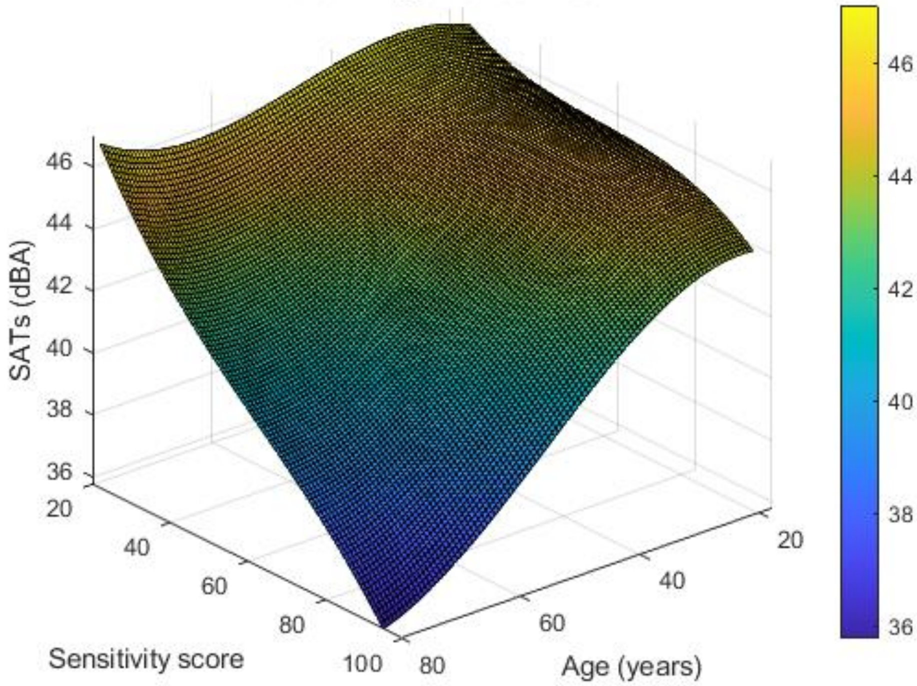


e) Sleepiness rating groups

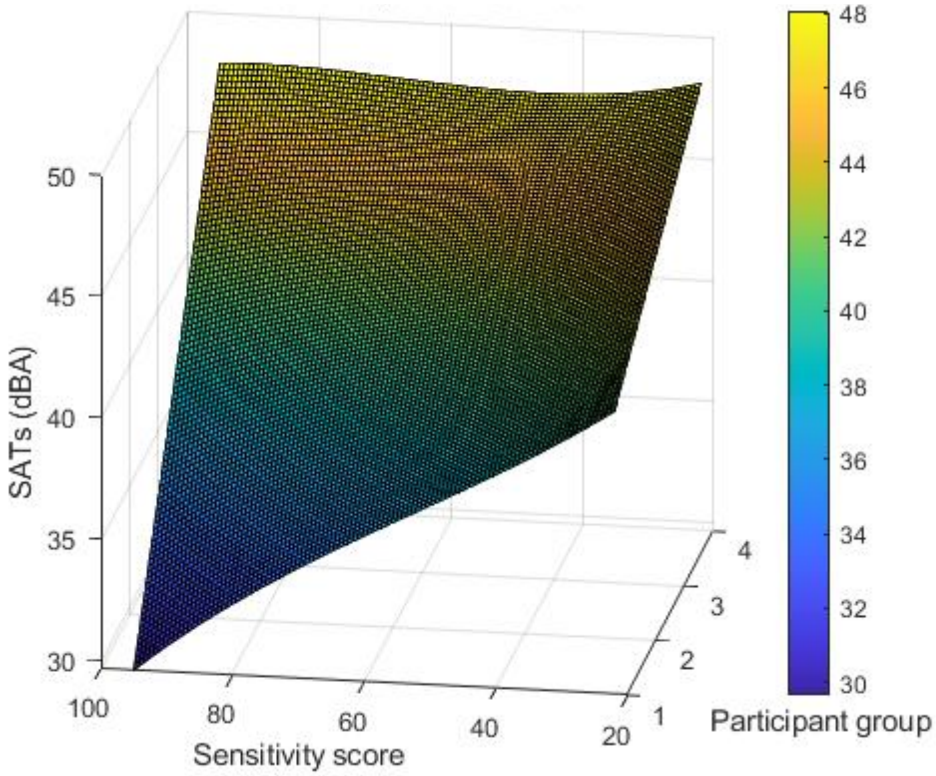
Appendix E. Partial dependence plots of tonal amplitude modulated WFN sleep acceptability threshold levels (SATs) for the interaction between a) Age and participant group b) Age and noise sensitivity c) Participant group and noise sensitivity d) Participant group and modulation depth. For participant groups: 1. WFN sensitive groups 2. WFN insensitive groups 3. Quiet rural area groups 4. RTN sensitive groups. Participant group was a categorical variable; however, a continuous line was used between participant groups for easier visualisation.



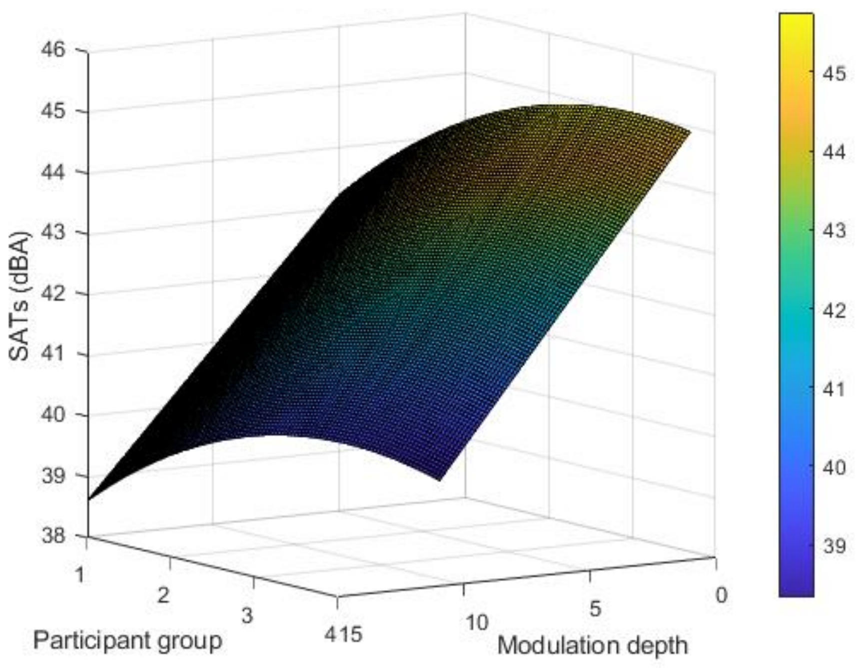
a) Age and participant group



b) Age and noise sensitivity

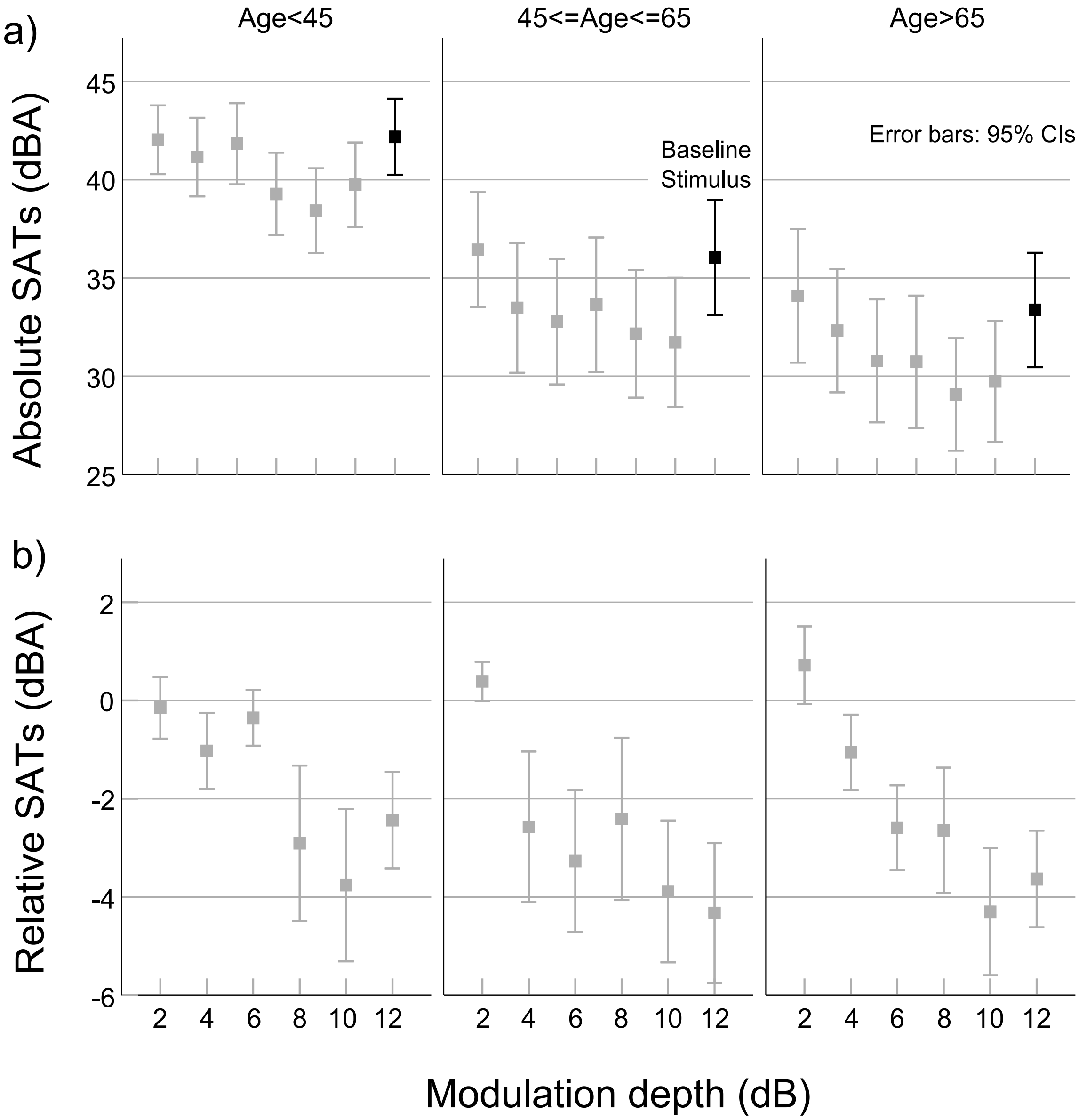


c) Participant group and noise sensitivity

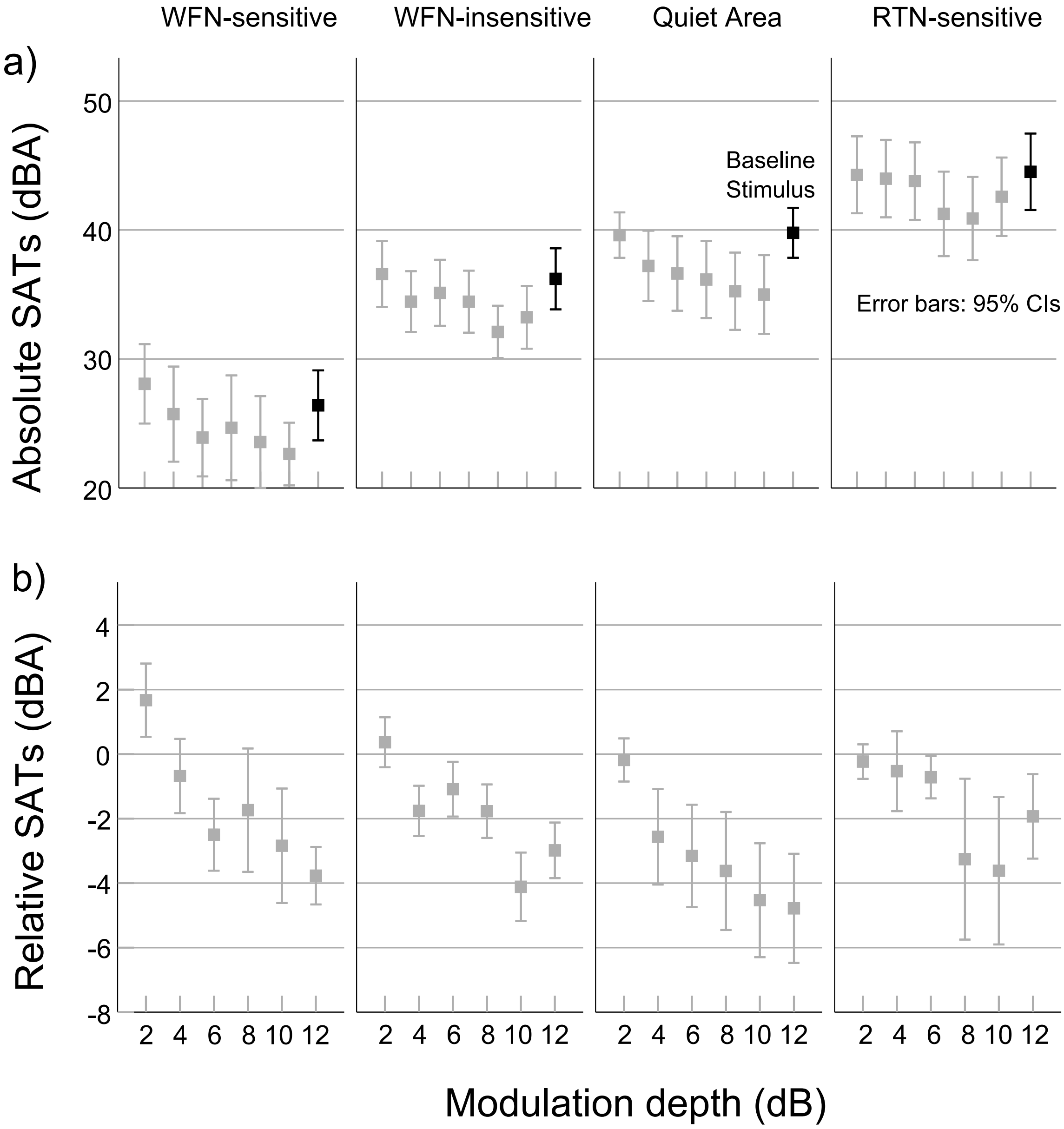


d) Participant group and modulation depth

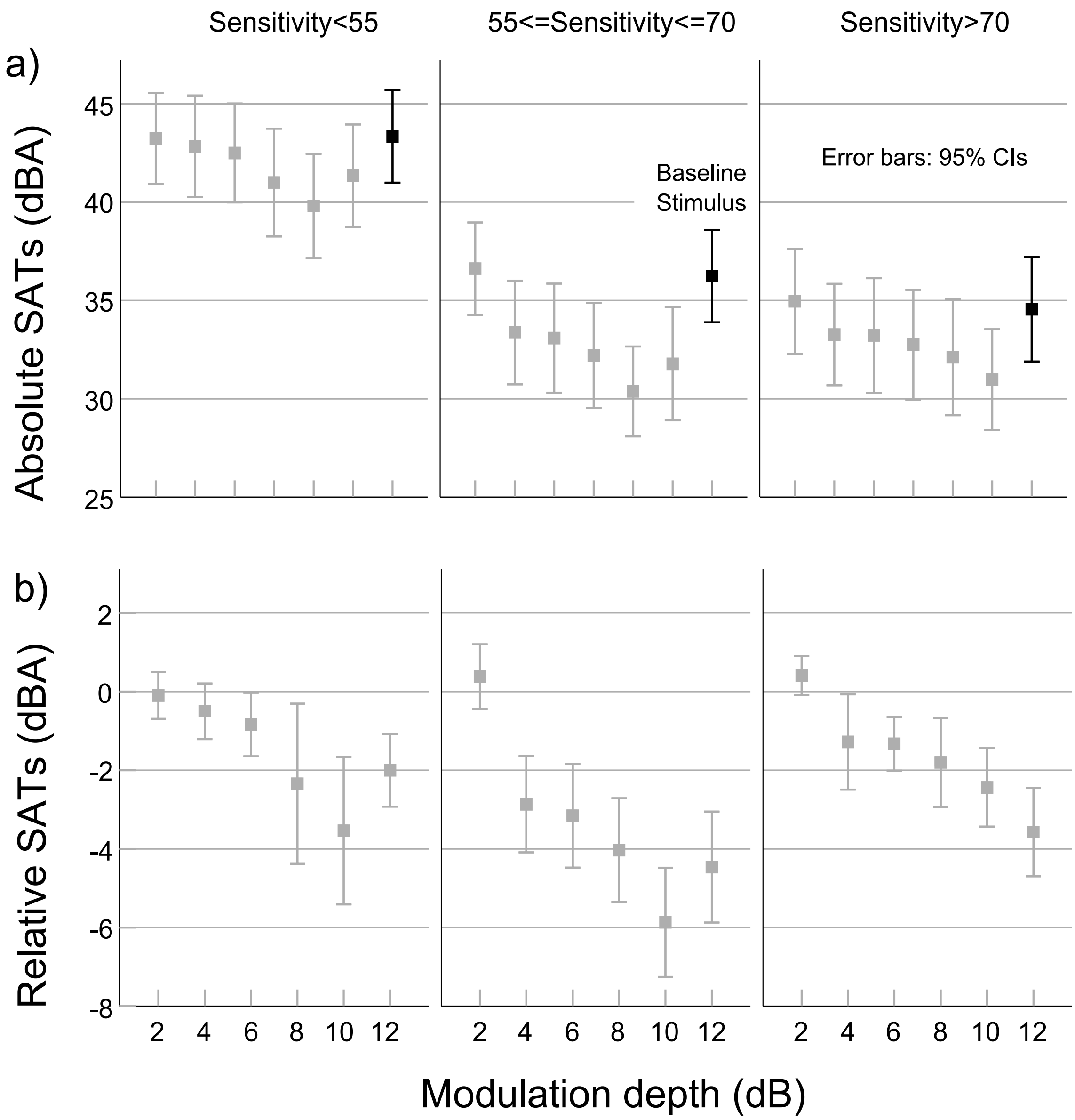
Appendix F. The effect of non-acoustic factor groups at different degrees of the swish noise modulation depth on a) sleep acceptability threshold levels (SATs) b) relative SATs. Relative SATs were obtained by subtracting the SATs of the baseline stimulus from the SATs of other stimuli for each participant. The lower the absolute and relative SATs value are, the less acceptable for sleep the noise is as compared to the baseline stimulus.



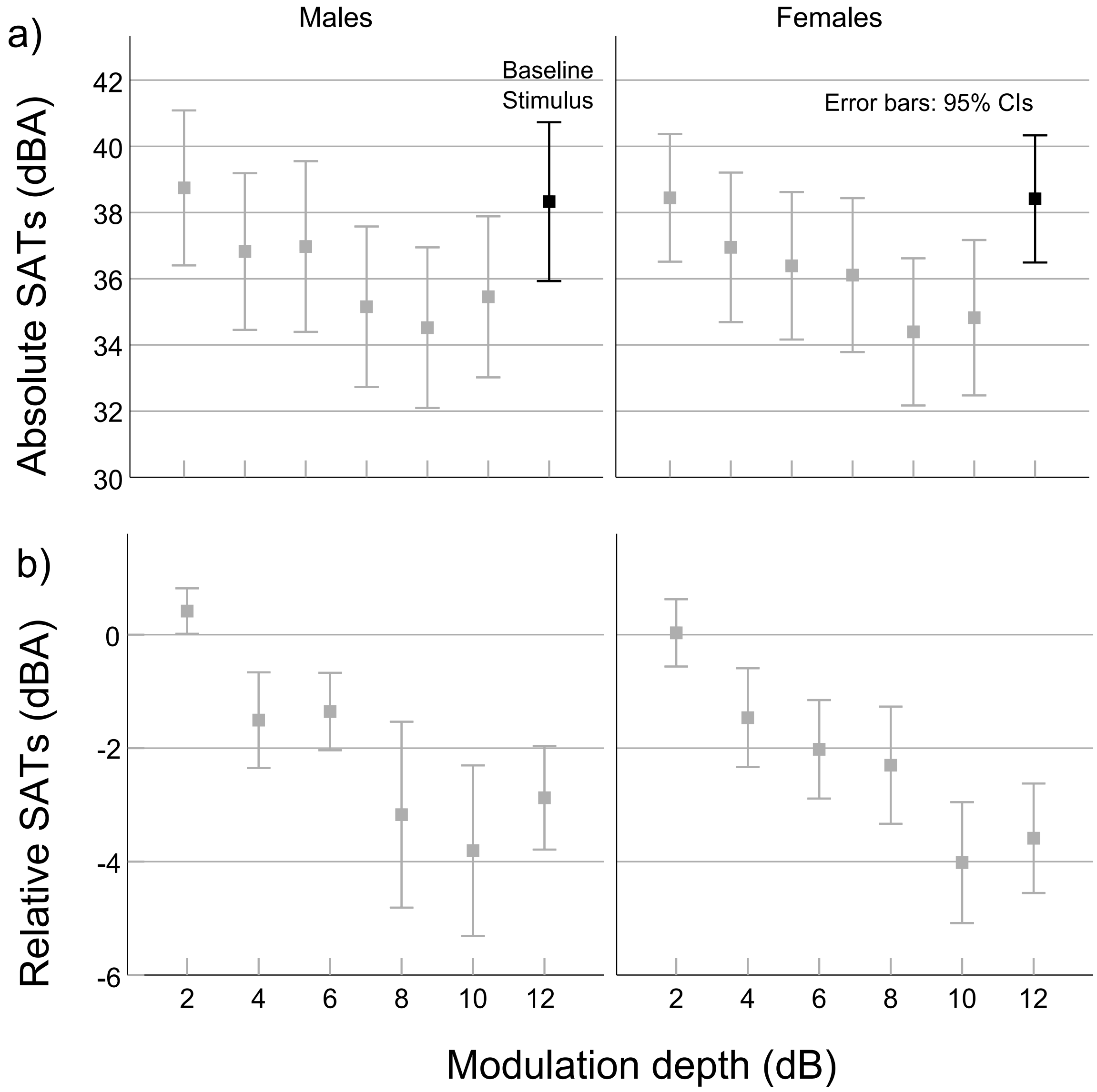
a) Age groups



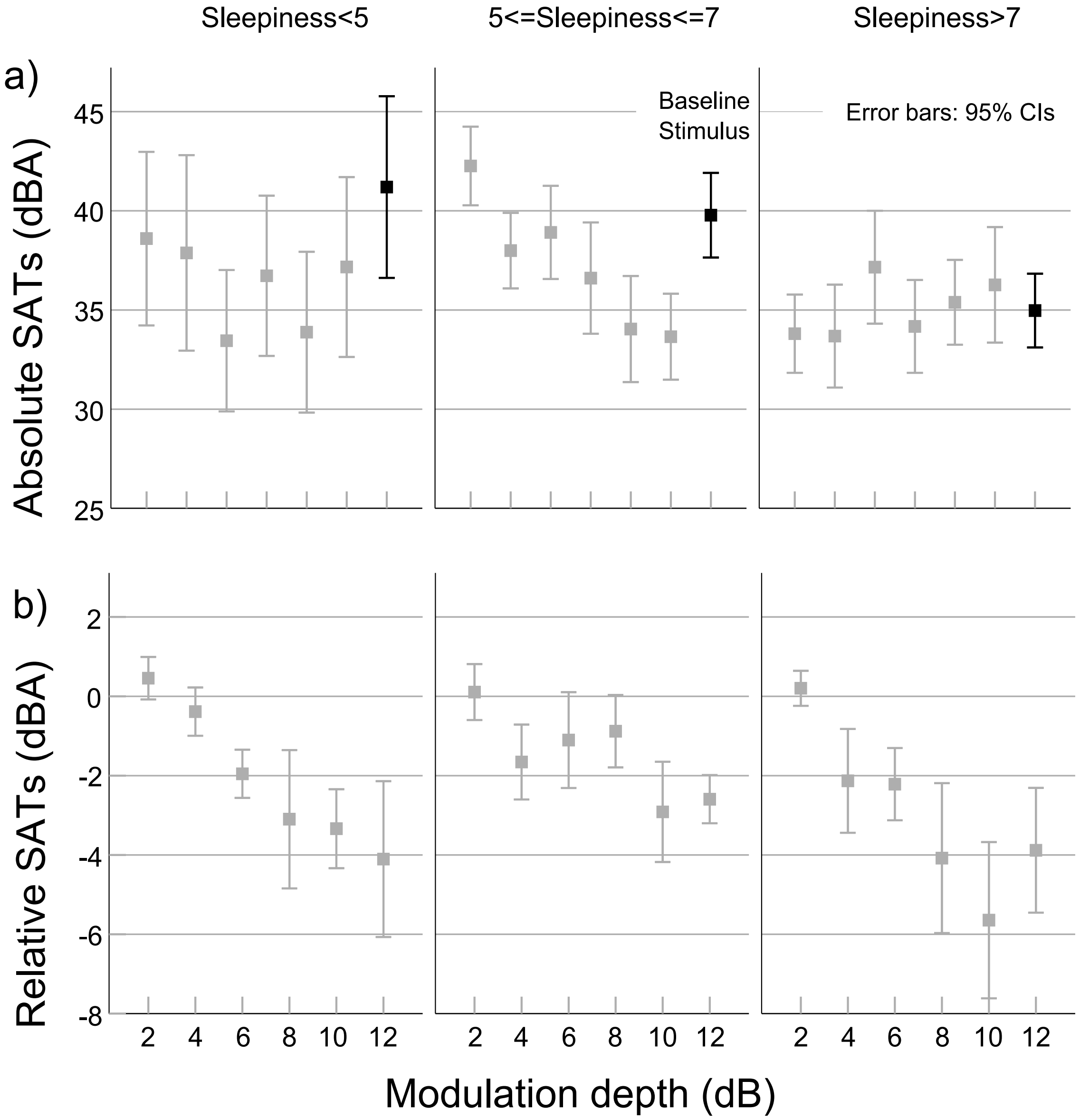
b) Participant groups



c) Self-reported noise sensitivity groups

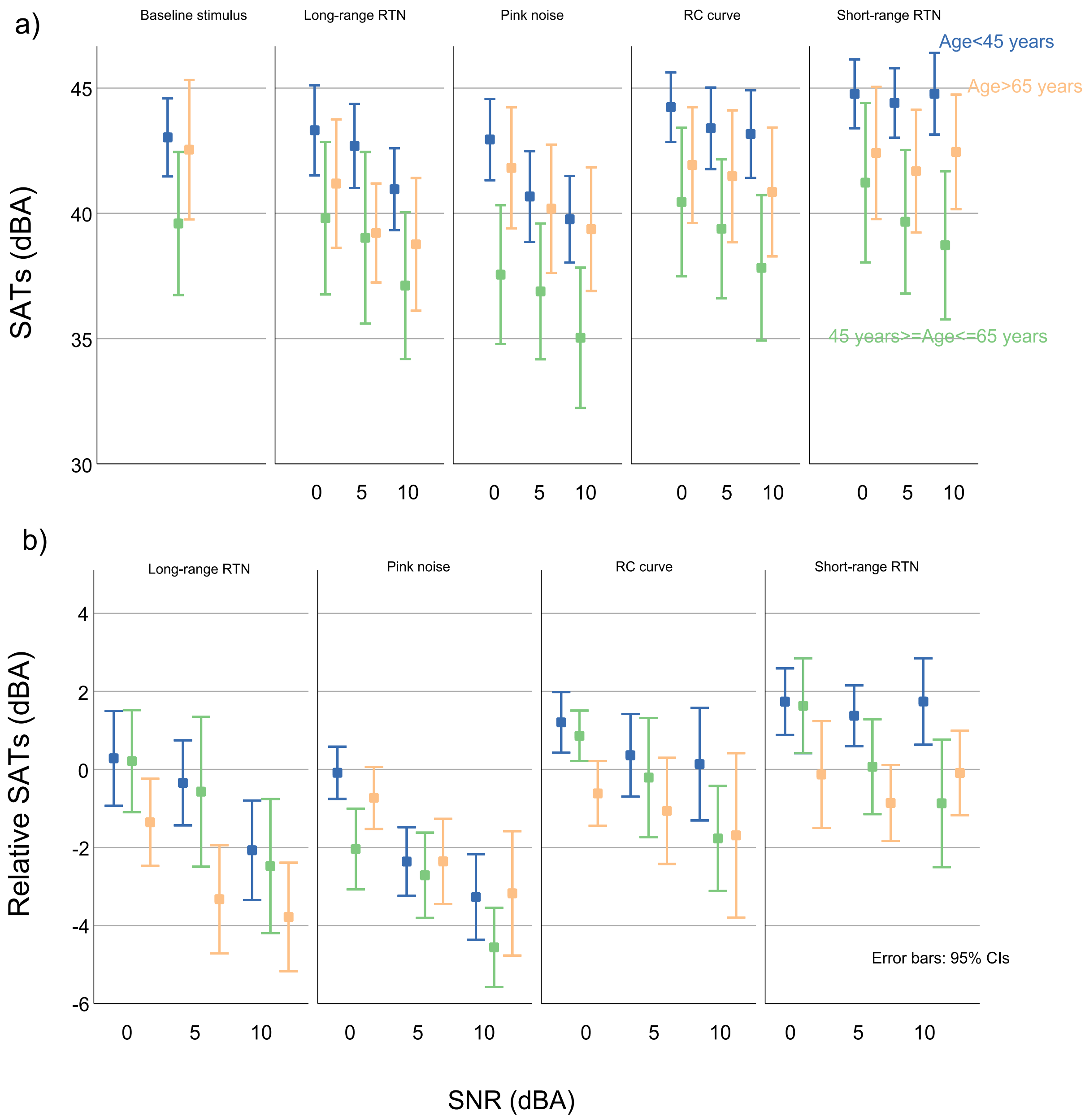


d) Gender groups

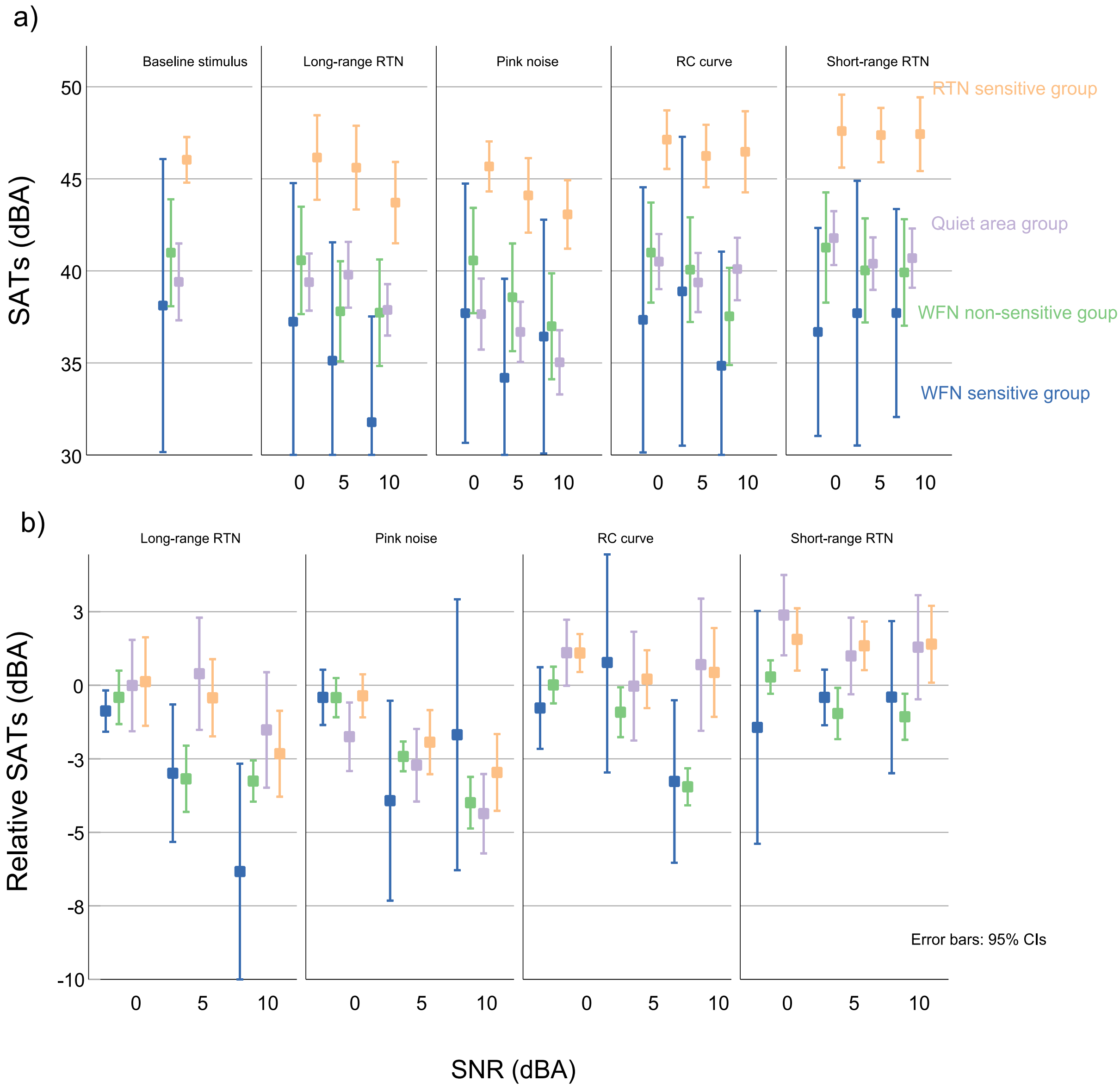


e) Sleepiness ratinggroups

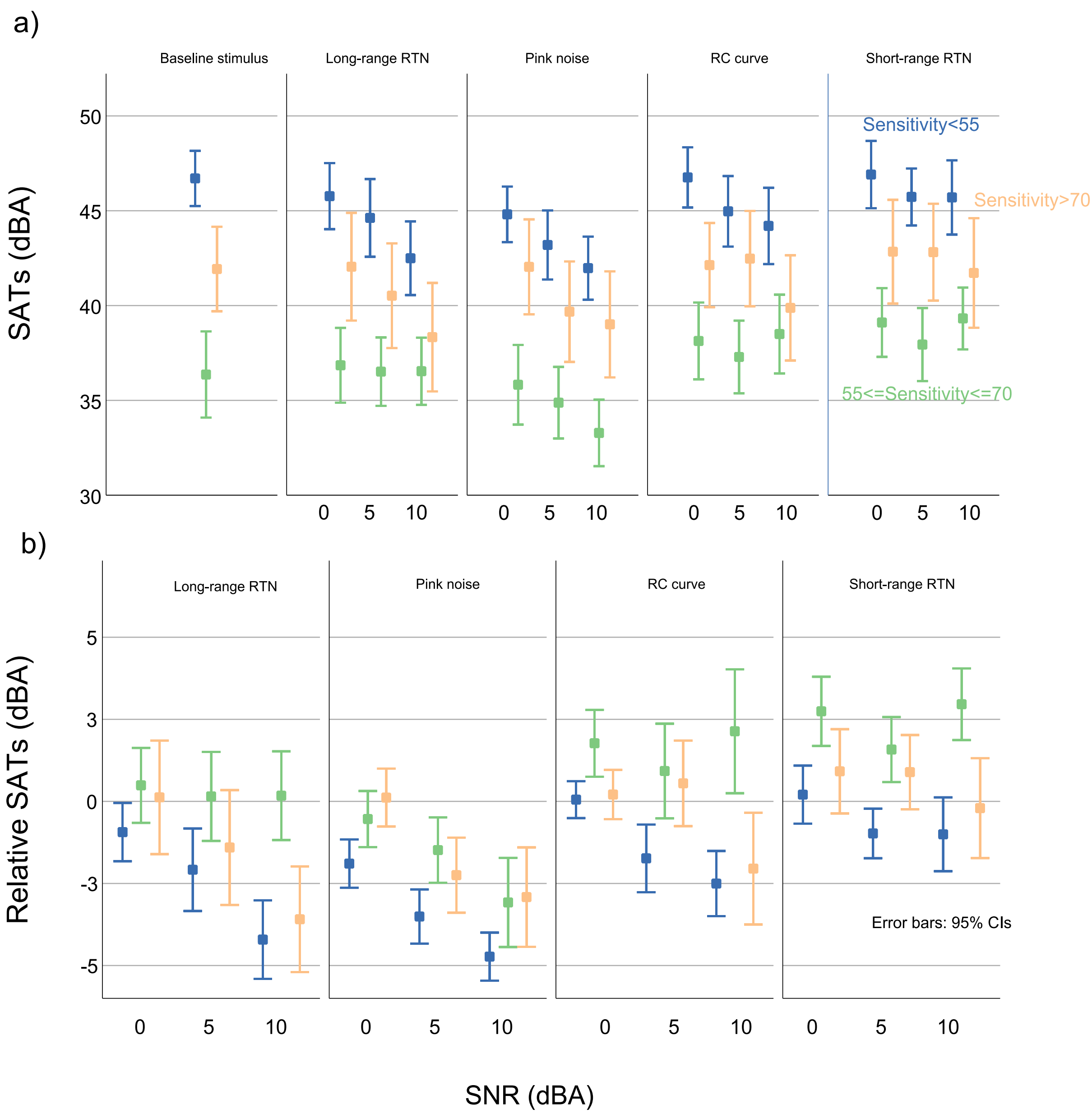
Appendix G. The effect of non-acoustic factor groups at different acoustic characteristics of the masking background noise of WFN (i.e. noise type and signal-to-noise ratio (SNR)) on a) sleep acceptability threshold levels (SATs) b) relative SATs. Relative SATs were obtained by subtracting the SATs of the baseline stimulus from the SATs of other stimuli for each participant. The lower the absolute and relative SATs value are, the less acceptable for sleep the noise is as compared to the baseline stimulus.



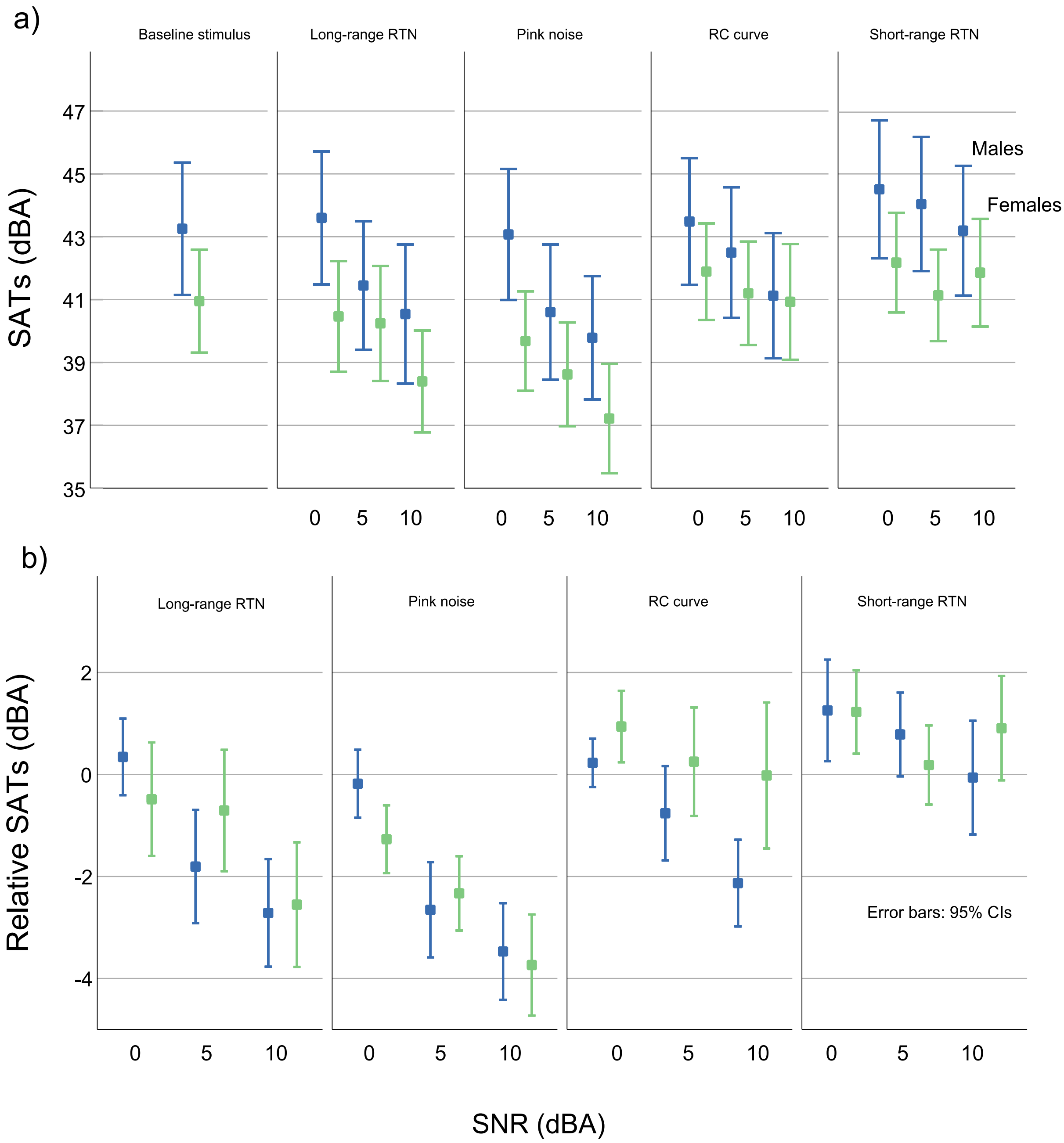
a) Age groups



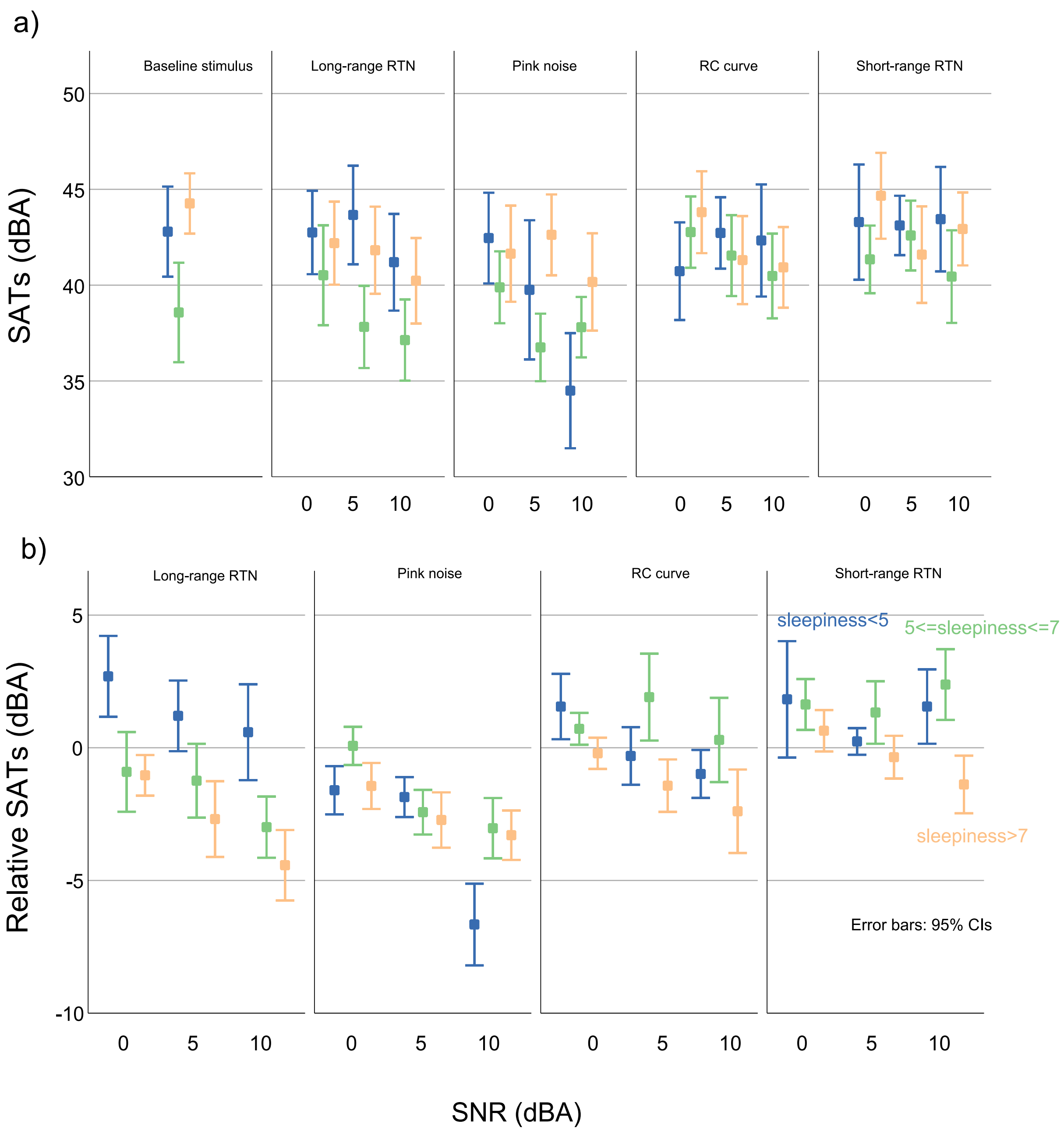
b) Participant groups



c) Self-reported noise sensitivity groups



d) Gender groups



e) Sleepiness rating groups