Evaluation of wind farm noise amplitude modulation synthesis quality

Duc Phuc Nguyen[1], Kristy Hansen[1*], Branko Zajamsek[2] and Peter Catcheside[2]

[1] College of Science and Engineering, Flinders University, Adelaide, Australia;
[2] Adelaide Institute for Sleep Health, Flinders University, Adelaide, Australia;
* Correspondence: kristy.hansen@flinders.edu.au; Tonsley (4.21); GPO Box 2100, Adelaide 5001, Australia.

Abstract: Wind farm noise amplitude modulation (WFN_{AM}) is a major contributor to annoyance and could cause sleep disturbance. In laboratory listening experiments assessing its annoyance and sleep disturbance potential, WFN_{AM} stimuli are commonly synthesised and can thus suffer from a lack of ecological validity. Here, five stimuli synthesis methods were compared with measured noise in terms of their perceived similarity. An ABX discrimination listening test and one-third octave band spectra were used for evaluation of the aural and visual similarity, respectively, between the synthesised and measured noise spectra. The results showed that synthesising WFN_{AM} using a simple method can be ecologically valid as listeners could not accurately differentiate between measured and synthesised WFN_{AM}. However, time varying features of WFN_{AM} do play a small but significant role in human perception and therefore hearing test evaluation of synthesis is recommended for obtaining the most ecologically valid synthesised WFN_{AM}.

Keywords: Wind farm noise; Amplitude modulation; ABX listening test; Synthesised noise; Tonal noise.

---

1. Introduction

Amplitude modulation (AM) is a major feature of wind farm noise (WFN), can be annoying [1-5] and has the potential to disturb sleep [6]. Wind farm noise amplitude modulation (WFN_{AM}) can be measured up to a few kilometres from a wind farm, be audible and have a characteristic amplitude modulated 46.5 Hz tone [7]. In physical terms, WFN_{AM} is a slow periodic variation in the level of the noise characterised by modulation depth at the blade-pass frequency around 0.8 Hz [8, 9]. This noise is intermittent and can be masked by the noise from road traffic, wind and agricultural activities [8]. WFN_{AM} is most easily detectable during late evening, night and early morning hours due to favourable meteorological conditions for noise propagation, minimal wind-induced noise and low agricultural activity noise [7].

Previous short term listening tests assessing WFN_{AM} annoyance used synthesised noise to systematically study the response to certain AM parameters [1-3, 9]. Each study used a unique synthesis method and the quality of the synthesised WFN_{AM} was inferred from the comparison of measured and synthesised spectra. As annoyance depends on both spectral and temporal characteristics, judging stimuli quality based only on spectral comparison and neglecting subtle-temporal AM characteristics, might not be sufficient. This study thus presents five WFN_{AM} synthesis methods and their quality evaluation using an ABX discrimination listening test [10] and spectral comparison with ten participants.

2. Materials and Methods

2.1. WFN_{AM} synthesis methods

For the purpose of signal synthesis, WFN_{AM} was assumed to be a combination of WFN signals with AM signals, as follows:

$$\mathrm{WFN_{AM}} = \beta \times (\mathrm{AM} + \alpha \times \mathrm{WFN}), \tag{1}$$

where β is a constant for controlling the overall sound pressure level (SPL), and α is a constant for controlling the level of WFN. To model WFN signals, this study used two methods which are based on the power-law spectrum and measured WFN spectrum.

For the first method, the power-law $(1/f^{\gamma})$ with a linear relationship between the log power spectral density and the log frequency [11] was used. The parameter γ can take values between 0 and 2 with characteristic values of 0, 1 and 2 for white, pink and brown noise, respectively. In the present study, pink noise (1/f), a clear outlier in terms of perception with respect to WFN, was used as a response bias control sample [12]. The response bias could in this experiment occur due to the lack of attention, misunderstanding of instructions or hearing impairments since hearing acuity was not checked. Regarding WFN synthesis, Yokoyama et al. [5] used a power law spectrum with a slope of -4 dB/octave (for comparison, pink noise has a slope of -3 dB/octave). The technique for synthesising pink noise is accessible to many acoustic practitioners as it is simple and available in many basic software packages. A general process to synthesise pink noise is shown in **Figure 1**a (upper panel).

The second method for synthesising WFN was based on the measured recordings [1], as shown in **Figure 1**a (lower panel) and Appendix A (pseudocode). According to this method, the measured sample was transformed into the frequency domain after moving average filtering, providing the smooth spectrum. The spectrum was then multiplied with the white noise spectrum, and the product was transformed back into the time domain using the inverse fast Fourier transform (IFFT).

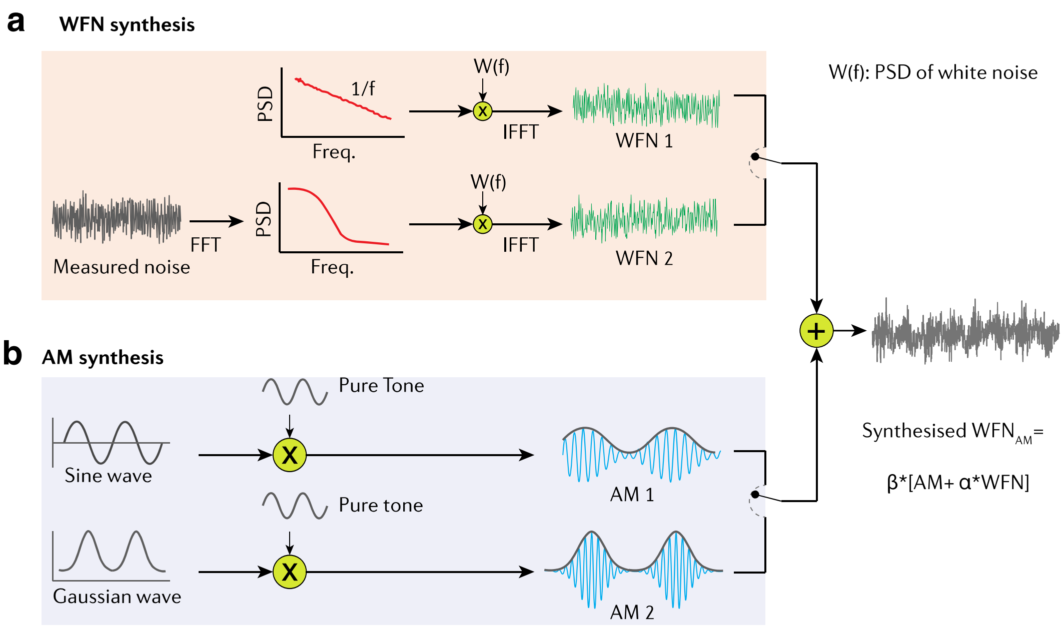

Figure 1| WFN$_{AM}$ synthesis methods. (a) Two methods for synthesising WFN signals using pink-noise (upper panel) and measured recordings (lower panel). (b) Two methods for synthesising AM signals using sine wave (upper panel) and Gaussian pulse train (lower-panel). Modulating signals are multiplied with a 46.5 Hz tone to create tonal AM signals which are then added to a WFN to form a characteristic WFN$_{AM}$.

For the synthesis of AM signals, a sine wave and Gaussian pulse train were used as shown in **Figure 1**b. The Gaussian pulse train more accurately represents AM [9] as it allows for greater control over the pulse shape and spacing between the pulses. This is important because the AM pulse can be asymmetrical, depending on the position of the receiver, and the spacing can vary due to propagation effects and changes in the blade rotational frequency. On the other hand, while the sine wave did not allow for such detailed tuning, it allegedly provided a satisfactory approximation of the AM [4]. Both methods have been used in previous studies [9, 13].

The final WFN_{AM} synthesised noise was created by combing WFN signals and AM signals into 5 unique combinations as shown in **Table 1**. All synthesised samples had a 0.8 Hz modulation frequency, carrier centre frequency of 46.5 Hz, modulation depth of 8 ± 0.5 dB quantified using the IOA 'reference method' [14], and tonal audibility of 10 dB estimated using IEC 611400-11 standard [15]. However, to further increase ecological validity, a random difference between peaks and troughs after each period was introduced, termed as random amplitude in **Table 1** (method 4 and 5). The random amplitude modulation was introduced via random amplitude modifications of the Gaussian wave and gave rise to the ± 0.5 dB variation in AM depth. A key focus of this study was on the quality of the AM noise synthesis and hence the focus on only the AM tone at 46.5 Hz as narrowband analysis of data measured at nine residences located between 1.3 and 8.8 km from a South Australian wind farm revealed that the most prominent amplitude modulation occurs at 46.5 Hz [7, 16]. Residents living near this wind farm have complained of a 'thumping' and/or 'rumbling' noise [7] and current work is underway to investigate the annoyance potential of this tonal AM.

Table 1. WFN_{AM} synthesis methods.

Method	WFN_{AM}	AM signals
1	WFN 1 + AM 1	Constant amplitude, sine wave
2	WFN 2 + AM 1	Constant amplitude, sine wave
3	WFN 2 + AM 2	Constant amplitude, symmetric Gaussian shape
4	WFN 2 + AM 2	Random amplitude, symmetric Gaussian wave
5	WFN 2 + AM 2	Random amplitude, asymmetric Gaussian wave

2.2. Wind farm noise spectrum

A comparison between the different WFN spectra used in the synthesis of WFN_{AM} is shown in **Figure 2**. The normalized SPL in each 1/3-octave band was calculated by subtracting the overall unweighted SPL from the SPL in each 1/3-octave band. WFN 2 spectra had a good approximation of the averaged measured background noise from 1562 10-minute recordings inside a residence 2.5 km away from a wind farm. The WFN 2 used in the present study was based on a single 10-minute recording with the amplitude modulated 46.5 Hz tone removed and a moving average filter applied for smother frequency response.

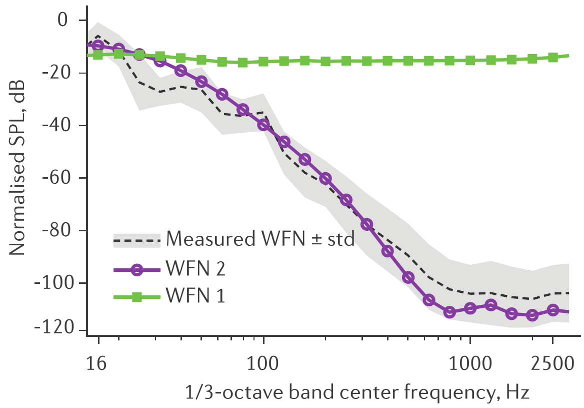

Figure 2| Background noise comparison.

2.3. Participants

Following approval from the Social and Behavioural Research Ethics Committee (SBREC) at Flinders University under project number 7536, ten participants (5 males) age from 21 to 50 years old were recruited for the listening test. Eight participants (5 acoustic engineers and 3 psychologists) were familiar with WFN. All participants had normal self-reported hearing.

2.4. Testing room, instrumentation and WFN stimuli

The listening test was conducted in a bedroom at the Adelaide Institute for Sleep Health (AISH), Flinders University where the daytime background noise level is below 21 dBA. The noise reproduction system consisted of an RME Babyface Pro sound card, Lab Gruppen C 16:4 power amplifier and Krix Harmonix MK2 loudspeaker. The SPL at participants' ears was 50 dBA and noise samples were smoothly ramped up and down using a 0.5 s, raised-cosine function. The loudspeaker was positioned in front of the participants, and the loudspeaker centre was aligned with the participant's ear level (**Figure 3**b). The listening test was delivered via a MATLAB GUI on a tablet PC with touch control, as shown in **Figure 7** in the Appendix. **Figure 3**a shows the measured WFN_{AM}

spectrum on which the synthesis of WFN 2 is based, with detail showing the region around the 46.5 Hz amplitude modulated tone.

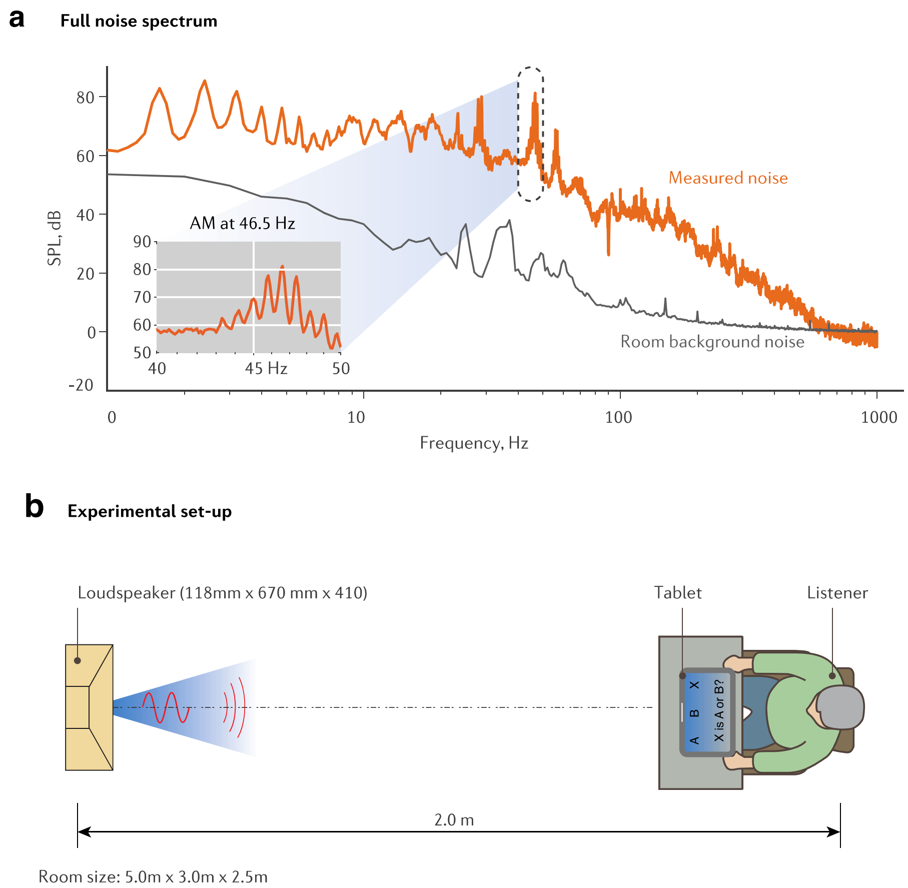

Figure 3| Stimuli and experimental design. (**a**), The spectrum of measured WFN$_{AM}$ and testing room background noise. The dashed line window shows a magnified view of the spectrum between 40 and 50 Hz where the 46.5 Hz tonal AM occurs. (**b**) Listening test set-up.

2.5. Experimental design

An ABX listening test [10] was used to evaluate how well listeners could distinguish between synthesised and measured WFN$_{AM}$ stimuli. During the test, participants were instructed to listen carefully to paired noise samples (i.e., sound A and B), each lasting 10 seconds prior to hearing and rating a third noise sample (sound X). Sound A was either measured WFN$_{AM}$ or synthesised WFN$_{AM}$ using one of the 5 methods summarised in **Table 1**. Sound B was the corresponding alternative, in randomised order of measured vs synthesised WFN$_{AM}$. For example, if sound A was a measured sample, sound B was a synthesised sample. After hearing sounds A and B, the participants were presented with a sound X which was identical to either sound A or B, determined randomly. The participants had to decide whether sound X was sound A or B.

Each participant underwent five trials corresponding to the five pairs of noise samples. There are 10 participants who rated each method once, resulting 50 ratings in total which is 10 ratings per method. Participants were asked to remain focussed on the task to facilitate task concentration and auditory memory, the experiment was kept as short as possible [10, 18] and took less than 20 minutes.

Apart from the ABX task, the participant was also asked to rate *"How confident are you about your choice?"* and *"How likely it is that sound A and B belong to the same recording?"* on an 11-point discrete scale from 0 (Not at all) to 10 (Extremely). The extreme alternatives were labelled as "Not at all" and "Extremely".

2.6. Data and Statistical analysis

To analyse ABX data, a response matrix was constructed using signal detection theory [10, 19]. In the response matrix, correctly recognising "X matches A" is termed a *Hit* and failing to recognise it as a *Miss*. Mistakenly recognising "X = A" as "X = B" is a *false alarm* and responding "B" to "X = B" is a *correct rejection*. The number of correct answers is the sum of the *hits* and *correct rejections*. From the response matrix a hit rate (HR), $R = \frac{Hit}{Hit+Miss}$, false alarm rate (FAR), $FAR = \frac{False\ alarm}{False\ alarm+Correct\ rejection}$, and sensitivity measure $d' = z(HR) - z(FAR)$, where z stands for z-transform, are calculated.

One-tailed binomial exact tests were used to assess probability of correct identification by chance. Pearson's correlation was used to examine the strength of relationships between measured and synthesised noise spectra. Effects of noise synthesis method and presentation order on rating scores were assess using linear-mixed model analysis (lmerTest package in R) using noise synthesis method and presentation order as fixed effects and with subject as a random effect, each with their own intercept, and the degree of freedom for F-tests was approximated using Satterthwaite's method. In the case of significant mixed model effects, relevant post-hoc contrasts were examined using Holm adjustment for multiple comparisons. All statistical analyses were performed using R (http://www.r-project.org/) using a significance threshold of $P = 0.05$.

3. Results

3.1. Measured and synchronised noise comparison

The similarity between 1/3-octave band SPL spectra of synthesised and measured noise (**Figure 4a**) was assessed using Pearson's correlation analysis and the corresponding results are shown in **Figure 4**b. The 1/3-octave band spectra from synthesised method 1 correlated poorly with the real measured WFNAM spectra (Pearson's correlation coefficient = 0.52, 95% CI [- 0.04 to 0.83], p = 0.07). In contrast, there was strong agreement between methods 2-5 and real WFNAM (Pearson's correlation coefficient = 0.98, 95% CI [0.92 to 0.99], p < 0.001), as shown in **Figure 4a** and **b**. Visual comparison using 1/3-octave band spectrum was consistent with correlation analysis, as shown in **Figure 4**a and b. On the other hand, differences between signals were less visible in the time domain, as shown in **Figure 4c**. According to these results, methods 2-5 would be judged as equally good since there is no discernable differences between them.

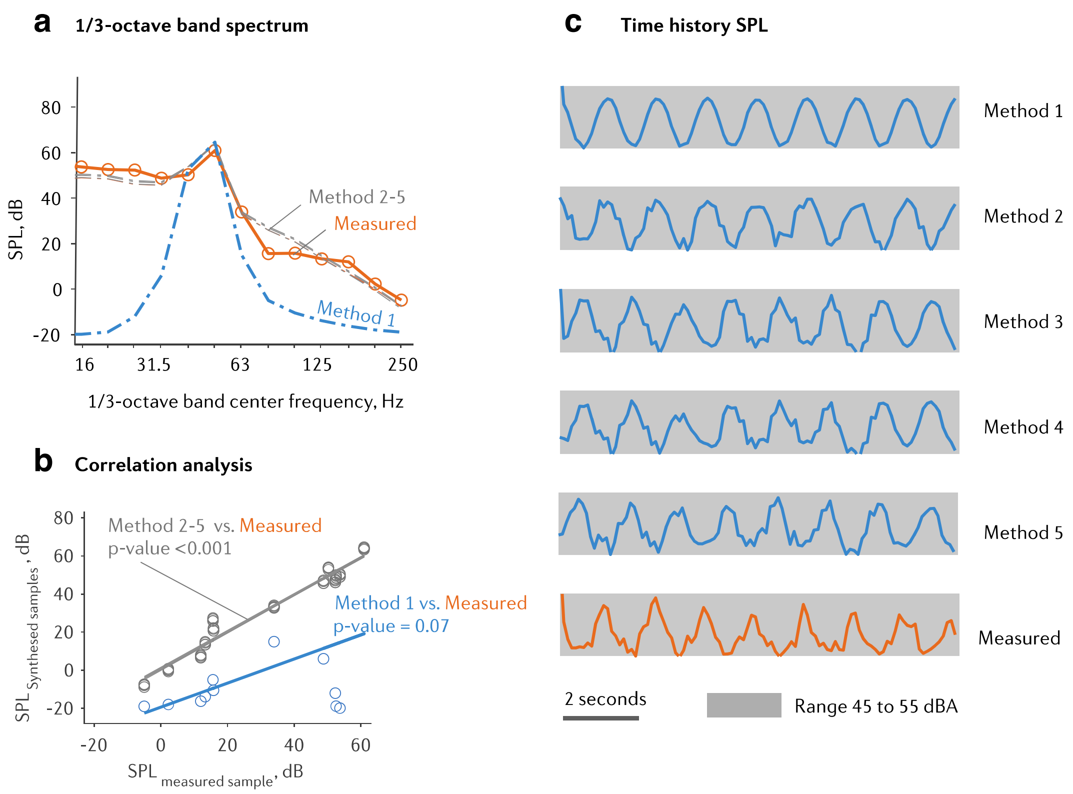

Figure 4 | Comparison between synthesised and measured noise. (**a**), Comparison in 1/3- octave band spectrum. (**b**), Correlation analysis. The levels at each 1/3-octave frequency from the synthesised WFNAM is plotted against the corresponding levels from the measured spectrum. The linear best fit

line also is plotted. (**c**). The time history of overall SPL. Gray shaded region shows the range of SPL between 45 and 55 dBA.

3.2. ABX test

An exemplary ABX test response matrix for method 5 is shown in **Table 2**. The number of correct answers is the sum of the *hits* and *correct rejections.* The number of trials is the sum of all cells in the matrix, and listeners have 50% probability of correct identification by chance. The number of correct answers for each method is shown in **Figure 5a**, and the results of binomial test are shown in **Figure 5b**. Only samples from method 1 could be distinguished (p = 0.01) by a significant margin, while samples from other methods were indistinguishable (**Figure 5**b).

Table 2. A typical ABX test response matrix.

Sample sequence	Response, X =	
	A	B
X = A	Hit (2)	Miss (3)
X = B	False alarm (2)	Correct rejection (3)

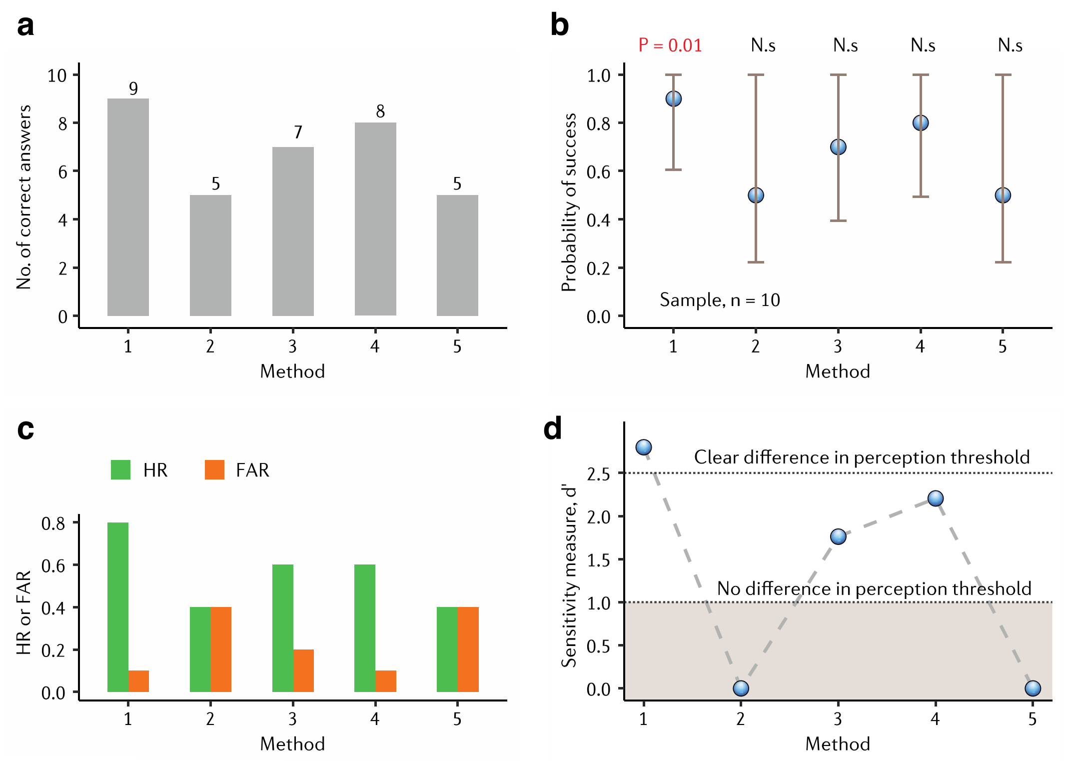

Figure 5| ABX test results (n = 10) for all 5 methods. (**a**) Number of correct answers out of 10. (**b**) Probability of success with one-tailed binomial exact test results where N.s stands for non-significance and error bars indicate 95% CI. (**c**) Hit rates (HR) and false alarm rates (FAR) according to signal detection theory (**d**) Sensitivity measure, d′, according to signal detection theory.

A binomial test for testing the null hypothesis of ABX results is not optimal, although valid, due to the small sample size for which signal detection theory is more suitable [19]. The signal detection theory results in **Figure 5**c and d reinforced the findings from using binomial test in **Figure 5**a and b, yet they appeared to be more sensitive, especially so for methods 2-5. Samples from methods 2 and 5 where the hardest to separate with HR = FAR (**Figure 5**c) meaning that participants were randomly guessing which was further confirmed by $d' = 0$ (**Figure 5**d). Generally, the value d' lies between 0 and 4.65, indicating no and maximal difference between stimuli, respectively [10]. Furthermore, d' larger than 2.5 represents a clearly perceivable difference whereas d' of 1 is considered a threshold value below which a participant cannot distinguish between the two types of noise [20]. The difference between measured and synthesised noise was thus clearly perceived for method 1 with $d' \approx 3$ while for methods 2 and 5 $d' = 0$ indicating no perceptible difference (**Figure 5**c). Methods 3 and 4 were somehow ambiguous without clear difference between measured and synthesised samples. The participant ability to correctly versus incorrectly discriminate between measured versus synthesised noise was not affected by presentation order (whether X = A or X = B; Odds ratio = 0.32, 95% CI = 0.09 to 1.19, Fisher's exact test $p = 0.13$).

3.3. Confidence and similarity rating results

The participants were very confident in distinguishing between the measured and synthesised noise from method 1, as reflected by high confidence ratings (**Figure 6**a). There was a significant difference between confidence ratings for the various methods ($p = 0.005$), with large differences between method 1 and all the other methods except pair 5-1 (**Figure 6**c). Similar trends were apparent for the similarity rating in **Figure 6**d, with a significant difference between the methods ($p < 0.001$) where method 1 was clearly different to all the other methods. The potential bias regarding the presentation order (measured or synthesised noise played first) was found no significance effects on the rating scores (p=0.5).

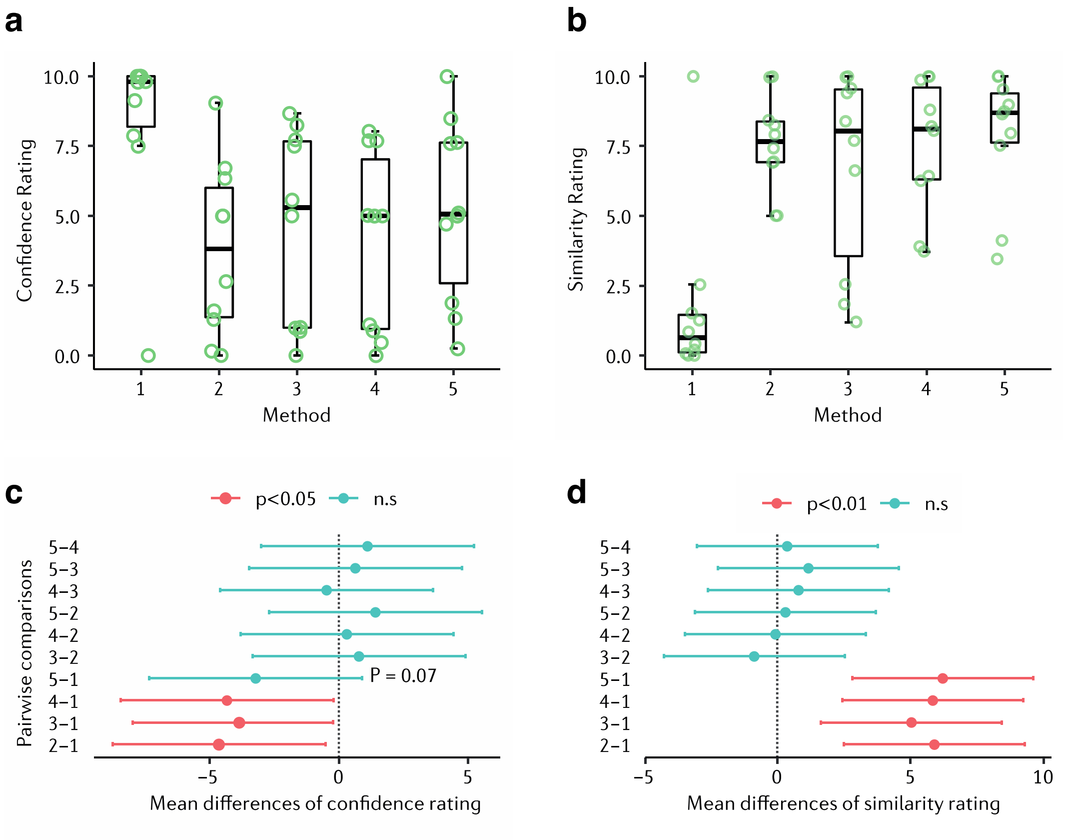

Figure 6| (a) Similarity and (b) confidence rating results. Multiple comparisons of mean (Tukey contrasts) for (c) confidence and (d) similarity rating results with error bars indicating 95% CI.

4. Discussion

The ABX results revealed that synthesised WFNAM noise can be representative of measured WFNAM and that an increased complexity of the synthesised method does not necessarily mean higher authenticity. Amplitude modulation modelled as a constant amplitude sine wave produced equally realistic WFNAM as a physically more correct and elaborate Gaussian pulse train with an asymmetric pulse shape and random amplitude. In contrast, the synthesised WFN spectra played an important role in WFNAM perception where only spectra based on measured WFN produces synthesised WFNAM indistinguishable from measured noise.

This study showed that the quality of synthesis cannot be fully captured by a visual spectral comparison which is most likely due to the loss of subtle and important AM clues during transformation to the frequency domain. However, if the measured and synthesised spectra obviously disagree then the perceptual difference is also likely to be large. When the spectral differences are small, the audio perceptual differences can help in identifying the most ecologically valid samples.

A limitation of the present study is that participants were selected based on their normal self-reported hearing instead of audiometry testing results. As a result, hearing acuity may have affected the participant's ability to detect differences between the two sounds. However, given that noises were played at a relatively high SPL (50 dBA), participants with self-reported normal hearing would be expected to clearly hear and to be able to demonstrate relatively normal discrimination between noise characteristics. A further limitation is that these results of the present study are based on a small sample size (n = 10), although listeners familiar with the WFN_{AM} were included, and thus the results should be interpreted with caution and listening tests with a larger sample size are warranted.

5. Conclusion

Five methods for synthesising WFN_{AM} were evaluated using an ABX listening test. The three main outcomes of the study are that the synthesis of WFN based on measured WFN_{AM} should be used for producing ecologically valid synthesised WFN_{AM}, that visual comparisons of spectra are not sufficient for identifying the most authentic synthesised noise and that relatively simple synthesis methods are sufficient for good synthesis. Some synthesised WFN_{AM} were found indistinguishable from measured WFN_{AM}, which shows that synthesised noise can have great ecological validity together with complete control over its parameters, which makes it ideal for laboratory experiments.

Acknowledgments: The authors gratefully acknowledge financial support from the Australian Research Council, Projects DP120102185 and DE180100022 and fellowship FT120100510 and the National Health and Medical Research Council, Project 1113571. We would also like to thank our colleagues who took part in this testing.

Appendix A

Algorithm 1 WFN synthesis

```
Procedure WFN(y(t))
Prepare a measured noise sample, y(t), which has N samples
Estimate PSD of the measured sample, Y(f), which has N/2 + 1 samples
```

4: Apply a moving average filter to the PSD spectrum at step 3 to obtain a general spectrum of the noise, $Y_{av}(f)$
5: Create a white noise signal, $w(t)$, and transform to frequency domain $W(f)$ (N samples)
6: Multiply the general spectrum at step 3, $Y_{av}(f)$, with N/2 + 1 first samples of the white noise in step 4: $X(f) = Y_{av}(f).\ W(f)$
7: Create a conjunction complex number of the derived product at step 6 to obtain N points complex number: $X(f) = [X(f); conj(X(f))]$
8: Invert Fourier transform $X(f)$ to obtain synthesised noise $x(t)$.

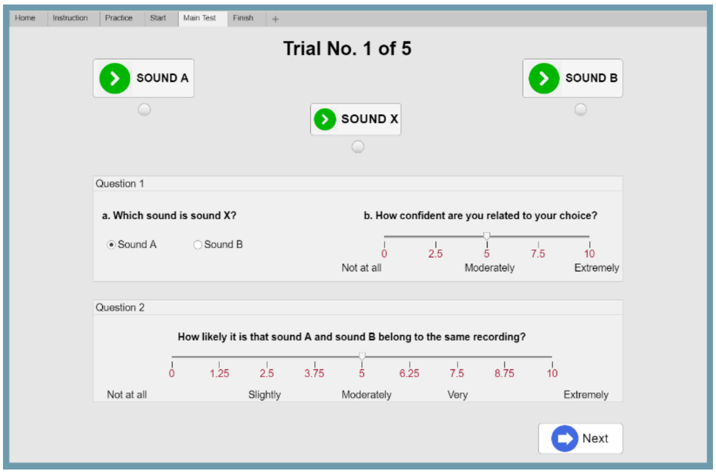

Figure 7| GUI MATLAB using for the experiment.

References

1. Lee, S., et al., *Annoyance caused by amplitude modulation of wind turbine noise.* Noise Control Engineering Journal, 2011. **59**(1): p. 38-46.
2. Ioannidou, C., S. Santurette, and C.-H. Jeong, *Effect of modulation depth, frequency, and intermittence on wind turbine noise annoyance.* The Journal of the Acoustical Society of America, 2016. **139**(3): p. 1241-1251.
3. Schäffer, B., et al., *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, 2016. **139**(5): p. 2949-2963.
4. Von Hünerbein, S., et al., *Wind turbine amplitude modulation: research to improve understanding as to its cause & effect.* Work package B (2): development of an AM dose-response relationship. RenewableUK, 2013.
5. Yokoyama, S., S. Sakamoto, and H. Tachibana. *Study on the amplitude modulation of wind turbine noise: part 2-Auditory experiments*. in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. 2013. Institute of Noise Control Engineering.
6. Micic, G., et al., *A review of the potential impacts of wind farm noise on sleep.* Acoustics Australia, 2018. **46**(1): p. 87-97.
7. Hansen, K.L., et al., *Prevalence of wind farm amplitude modulation at long-range residential locations.* Journal of Sound and Vibration, 2019. **455**: p. 136-149.
8. Hansen, C.H., C.J. Doolan, and K.L. Hansen, *Wind farm noise: measurement, assessment, and control*. 2017: John Wiley & Sons.

9. UK, R., *Wind turbine amplitude modulation: research to improve understanding as to its cause & effect*. 2013, Renewable UK.

10. Macmillan, N.A. and C.D. Creelman, *Detection theory: A user's guide*. 2004: Psychology press.

11. Halley, J.M. and W.E. Kunin, *Extinction risk and the 1/f family of noise models.* Theoretical Population Biology, 1999. **56**(3): p. 215-230.

12. Kirk, R.E.J.H.o.P., Second Edition, *Experimental design.* 2012. **2**.

13. Zwicker, E. and H. Fastl, *Psychoacoustics: Facts and models*. Vol. 22. 2013: Springer Science & Business Media.

14. AMWG, *A Method for Rating Amplitude Modulation in Wind Turbine Noise*. 2016, UK Institute of Acoustics.

15. International Electrotechnical Commission Geneva, S.I.E.C., *IEC 61400-11 Wind Turbine Generator Systems—Part 11: Acoustic Noise Measurement Techniques.* 2012.

16. Hansen, K., B. Zajamsek, and C. Hansen. *Comparison of the noise levels measured in the vicinity of a wind farm for shutdown and operational conditions*. in *INTER-NOISE and NOISE-CON congress and conference proceedings*. 2014. Institute of Noise Control Engineering.

17. Thorsson, P., et al., *Creating sound immission mimicking real-life characteristics from a single wind turbine.* Applied Acoustics, 2019. **143**: p. 66-73.

18. Greenaway, R.E., *ABX Discrimination Task*, in *Discrimination Testing in Sensory Science*. 2017, Elsevier. p. 267-288.

19. Boley, J. and M. Lester. *Statistical analysis of ABX results using signal detection theory*. in *Audio Engineering Society Convention 127*. 2009. Audio Engineering Society.

20. Lawless, H.T. and H. Heymann, *Sensory evaluation of food: principles and practices*. 2013: Springer Science & Business Media.

21. ISO389-7, *Acoustics: Reference Zero for the Calibration of Audiometric Equipment Part 7: Reference Threshold of Hearing under Free-Field and Diffuse-Field Listening Conditions*. 2005.