Psychoacoustic Evaluation of Modelled Wind Turbine Noise

Pockelé, J.S.; Merino Martinez, R.

Publication date
2024

Document Version
Final published version

Published in
Proceedings of the 30th International Conference on Sound and Vibration

30th International Congress on Sound and Vibration

PSYCHOACOUSTIC EVALUATION OF MODELLED WIND TURBINE NOISE

Josephine Siebert Pockelé and Roberto Merino-Martínez
Delft University of Technology, Delft, the Netherlands.
E-mail: j.s.pockele@tudelft.nl

Current validation of wind turbine noise models primarily focuses on sound levels averaged over time, typically expressed in metrics such as the equivalent A-weighted sound pressure level ($L_{\mathrm{A,eq}}$). Whereas valid for regulatory purposes, these methods are not sufficient for psychoacoustic research, as time-averaged levels alone do not fully explain the measured noise annoyance. Therefore, this research aims to establish whether psychoacoustic sound quality metrics (SQMs) provide additional value when analysing wind turbine noise models. This work employs the Horizontal Axis Wind turbine simulation Code 2 (HAWC2) to generate noise source spectrograms, which are propagated through the atmosphere with a Gaussian beam-tracing approach. The final sound signals are retrieved through an inverse Short-Time Fourier Transform (iSTFT). This methodology is applied to a case study featuring a stall-controlled, horizontal-axis, Nordtank NTK 500/41 wind turbine. The results are evaluated against measurements by considering $L_{\mathrm{A,eq}}$, SQMs, and a comparative listening experiment. The $L_{\mathrm{A,eq}}$ metric shows a consistent underprediction of the simulations with respect to measurements, which is partly explained by a ground reflection modelling error. In the high-frequency range, stall noise is known to be significantly underpredicted by the aero-acoustic simulation model. This usually translates in increasing discrepancies between measurements and models as the wind speed increases. The comparative listening experiment confirms that participants experience the simulations and the measurements as significantly different. The difference ratings show a good agreement with the differences in the psychoacoustic annoyance and loudness metrics. It is more difficult to relate the results from the listening experiment to $L_{\mathrm{A,eq}}$. These findings confirm that an evaluation with psychoacoustic metrics next to conventional methods provides additional value in validating wind turbine noise models for human perception research.

Keywords: psychoacoustics, wind turbine noise, listening experiments, aeroacoustics, auralisation

1. Introduction

Worldwide climate goals are driving a continuous increase in the installed capacity of wind turbines. While the offshore industry is growing, onshore wind still accounts for most of the installed capacity in Europe [1]. One of the main permitting challenges facing onshore wind is noise emissions in neighbouring residential areas. Given the complex nature of noise issues surrounding wind energy, recent publications suggest tackling it from a socio-technical perspective [2]. Therefore, accounting for the human perception in the analysis of wind turbine noise will become increasingly important. Conventional evaluation of wind turbine noise primarily focuses on time-averaged sound pressure levels, which do not fully explain the experienced annoyance. Therefore, a psychoacoustic method of determining wind turbine noise annoyance is suggested as a solution to this issue [3].

The suggested framework of [3] requires a method to generate wind turbine noise. When it comes to future wind turbine types and installations, only numerical methods can be applied. A new method of validating these numerical models through psychoacoustic sound quality metrics is proposed to overcome the limitations of time-averaged sound pressure levels in capturing human perception. The goal of this study is to evaluate numerically generated wind turbine noise signals with conventional sound pressure levels, psychoacoustic sound quality metrics, and a comparative listening experiment, to prove the additional value of sound quality metrics to validate wind turbine noise models in human perception research.

Section 2 presents the four methodologies: (1) noise modelling and auralisation, (2) the noise measurement setup, (3) the sound quality metrics (SQMs) evaluation, and (4) the listening experiment setup. The spectral differences between the modelled and measured noise are characterised in Section 3. The results from the different analyses are presented to show the shortcomings of sound pressure levels in Section 4.1 and to prove the additional value provided by SQMs in Section 4.2. The paper closes with some concluding remarks.

2. Methodology

2.1 Noise modelling and auralisation

The numerical noise generation consists of three modules: (1) source simulations with the Horizontal Axis Wind turbine simulation Code, 2nd generation (*HAWC2*), (2) propagation with Gaussian beam tracing and (3) the retrieval of audible sound files from the spectrograms. These three modules are combined in the Python-based auralisation code *WinTAur v.1.0.0* [4].

At the source level, noise is generated with the HAWC2 aeroacoustics module from the Technical University of Denmark. The code [5] models three aerodynamic noise sources relevant in wind turbines: (1) turbulent inflow noise, (2) turbulent boundary layer trailing edge noise, and (3) stall noise. Turbulent inflow noise is determined using Amiet's theory. The TNO model is combined with the BPM model directivity pattern to calculate the trailing edge noise. A model based on Amiet's theory calculates the stall noise [5].

Using the methodology in [6], the wind turbine noise is calculated on a uniformly spaced, spherical grid, which acts as the input for the Gaussian beam tracing propagation model [6]. The governing ray equations are solved with a forward Euler scheme in time. Atmospheric and ground absorption are applied to the Gaussian beams, based on the weather, wind, and ground conditions defined by the simulation cases shown in Table 1 [6]. At the receiver position(s), the Gaussian beam tracing defines an acoustic energy distribution depending on the perpendicular distance between the central ray of the beam, and a receiver. The beam width is defined by the angular spacing of the spherical source grid [6].

To retrieve the signal in the time domain from its frequency domain representation, a windowed inverse short-time Fourier transform (iSTFT) method is applied with a uniformly distributed random phase spectrum [6]. The time resolution of the final spectrograms is $\Delta t = 0.01$ s. The iSTFT uses a Hanning window with size $N_{FFT} = 8192$, with a two-sided reconstruction overlap of 750% and no zero-padding. This results in a final wind turbine noise signal with a sampling frequency of 51.2 kHz.

2.2 Noise measurement setup

This work uses the acoustic field measurements of a *NordTank NTK 500/41* wind turbine from [5] and [7]. They are applied to the simulation setup and in the psychoacoustic evaluation of the auralisation

Table 1: Summary of wind turbine data selected for this study (Adapted from Table 9.1 in [6]).

Case nr.	Mean Wind Speed (m/s)	Mean Rotor RPM	Mean Power (kW)	Turbulence intensity (%)	Temperature (°C)	Pressure (hPa)	Humidity (%)	Mean Wind Direction (°)	Mean yaw (°)
1	6.46	26.8	83	3	12.09	1018	78	291	295
2	8.85	26.9	237	8	12.24	1018	75	289	298

tool. The three-bladed, stall-regulated, horizontal axis wind turbine was located at the *DTU Risø* campus. The wind and weather measurement setup consists of a 36 m tall meteorological mast at a distance of $2.5D_r$ to the west of the turbine tower, where the rotor diameter is $D_r = 41$ m. Instruments at various heights measure wind speed and direction, as well as air temperature, pressure and humidity. The turbine is instrumented for measurements of yaw, rotor speed, electrical output, and status information. Wind speed and direction are also measured with a nacelle-mounted LIDAR. These parameters are all recorded in 16-bit quantities at 35 Hz, through a PC-based analogue-to-digital data acquisition unit. Eight microphones, manufactured by *BSWA Technology Co. (ref. MPA 261 combining a 1/2" microphone and pre-amplifier)*, are mounted according to IEC 61400-11:2012 [8]. They are placed 45 m from the tower, with an approximately even azimuthal spacing.

This work analyses two 2-minute time series of the measurement campaign, which are selected based on the A-weighted signal-to-noise ratio and the lack of audible disturbances, such as birds, aircraft, or traffic. These cases are summarised in Table 1 [6]. One should note the low turbulence intensity and mean wind speeds, as these may be less representative of typical operating conditions. The parameters in Table 1 are used as the inputs for the simulations in *WinTAur*.

2.3 Sound quality metric evaluation

Sound Quality Metrics (SQMs) describe the subjective perception of sound by human hearing. This is unlike $L_{\mathrm{A,eq}}$, which mostly characterizes noise from a physical magnitude point of view. Hence, SQMs are expected to better capture the auditory behaviour of the human ear compared to conventional sound metrics, as typically employed in noise assessments. The five most commonly-used SQMs [9] are:

- Loudness (N): Subjective perception of sound magnitude corresponding to the overall sound intensity.
- Tonality (K): Measurement of the perceived strength of unmasked tonal energy within a complex sound.
- Sharpness (S): Representation of the high-frequency sound content.
- Roughness (R): Hearing sensation caused by sounds with modulation frequencies between 15 Hz and 300 Hz.
- Fluctuation strength (FS): Assessment of slow fluctuations in loudness with modulation frequencies up to 20 Hz, with maximum sensitivity for modulation frequencies around 4 Hz.

These five SQMs are calculated for each noise signal and combined into a single global psychoacoustic annoyance (PA) metric. All the SQMs and the PA metric are computed using the open-source MATLAB toolbox SQAT (Sound Quality Analysis Toolbox) v1.1 [10]. The Zwicker model [11] is used to determine the PA metric, since the tonality metric is very low for both the measured and simulated sound signals [6].

2.4 Listening experiment setup

Listening experiments were performed in the Psychoacoustic Listening Laboratory (PALILA) at the Delft University of Technology, Faculty of Aerospace Engineering. PALILA is a box-in-box concept, soundproof booth, which serves to isolate participants from external noise and other disturbances during the experiment [12]. Participants interact with a Python-based graphical user interface (GUI) on a laptop. Audio is reproduced with a pair of *Sennheiser HD 559* open-back headphones connected through a universal 3.5 mm audio jack [6].

The experiment section from [6] relevant to this work consists of 20 comparative questions, where participants were asked to rate the difference between a simulated and a recorded sample of wind turbine noise. Both samples had a 7 s duration, were separated by 2 s of silence, and could only be played once. Every combination of two samples was presented twice in opposite order. Participants rated the difference on a scale from 0 to 4 in integer steps, and were asked to base their rating on the overall annoyance and feeling related to the samples [6].

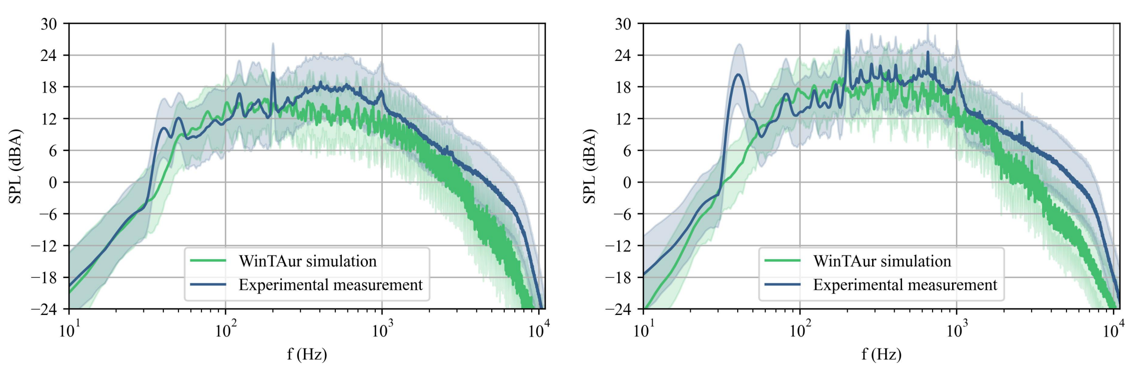
Since participants had a different baseline for measuring differences between two samples, the difference ratings (R_Δ) are normalised by subtracting the mean per participant: $\tilde{R}_\Delta = R_\Delta - \overline{R_\Delta}$. The per-participant statistics support this normalisation, as the standard deviations of the responses are very consistent between participants.

3. Characterisation of Spectral Differences

Figure 1 presents the A-weighted, narrowband sound pressure level spectra of the simulations and measurements for both cases. Whereas there are differences between the microphones, these two spectra represent the overall observations from [6].

The spectra of the measured wind turbine noise show three tones: one between 30 and 60 Hz, one at 200 Hz, and one at 1000 Hz. The lowest frequency tone is attributed to the gearbox and generator, as it matches the expected high-speed shaft frequency of the turbine. The latter tones are attributed to damage or imperfections on a single blade, as they repeat at the the rotor's rotational frequency. These two tones are highly directional, occurring most strongly in the measurements under the downward stroke of the rotor (microphones 1 and 2). The tonal noise components are not present in the simulations, as they are not included in the modelling [5].

From 200 Hz on, the simulated noise has lower levels than the measured noise. This is partly at-



(a) Case 1, 6.46 m/s, microphone 2.

(b) Case 2, 8.85 m/s, microphone 5.

Figure 1: Mean A-weighted, narrowband sound pressure level spectra of two representative simulations and their respective experimental measurements. Shaded areas show the standard deviations.

tributed to a difference in ground reflection coefficient [6]. Above 2 kHz, the discrepancies are partly attributed to the known underprediction of stall noise in *HAWC2*, where the discrepancies increase with wind speed [6]. The spectral fluctuations in the high-frequency range are artefacts of the signal reconstruction process, as they are not present in the spectrograms from which the output signal is retrieved. The standard deviation of the sound pressure levels over time matches well between the simulations and measurements. The overall findings are in line with the validation results of *HAWC2* [5] (not shown here), indicating that the included engineering noise models are the main source of error in *WinTAur*.

4. Results

This section compares the different analysis methods used in the validation of auralisation using wind turbine noise models. All parameters in this section quantify the difference between the simulations and their corresponding measurements. The polar plots in this section show the results in the orientation around the turbine corresponding to the microphone positioning. In the centre of each polar, the orientation of the turbine and the incoming wind is shown.

For the listening experiment results, the mean of the normalised difference ratings are presented ($\tilde{R}_{\Delta,\text{mean}} = \frac{1}{N}\sum_p \tilde{R}_\Delta$). The SQMs are presented as the difference of the mean psychoacoustic annoyance metric between the simulations and measurements ($\Delta\overline{PA} = \overline{PA}_{\text{measured}} - \overline{PA}_{\text{simulated}}$). The sound pressure level results are shown as the difference between the equivalent, A-weighted sound pressure levels ($\Delta L_{\text{A,eq}} = L_{\text{A,eq,measured}} - L_{\text{A,eq,simulated}}$). Each plot presents two of these parameters overlapped with scaled axes to represent a linear fit between these parameters. The axis scaling in the plots is a compromise between the best fit and readability.

4.1 Shortcomings of sound pressure levels

This subsection elaborates on where the results of $\Delta L_{\text{A,eq}}$ have shortcomings regarding their representation of human perception, as commonly argued in literature [3]. To this end, the results of the listening experiment and the SQMs are compared to the results from the sound pressure level analysis.

Firstly, the relation between the listening experiment and $\Delta L_{\text{A,eq}}$ is investigated. Figure 2 shows $\tilde{R}_{\Delta,\text{mean}}$ compared to $\Delta L_{\text{A,eq}}$. The Pearson correlation coefficient is $\rho = 0.501$ with a p-value $p = 0.98 \times 10^{-3}$, showing some correlation, which is considered statistically significant.

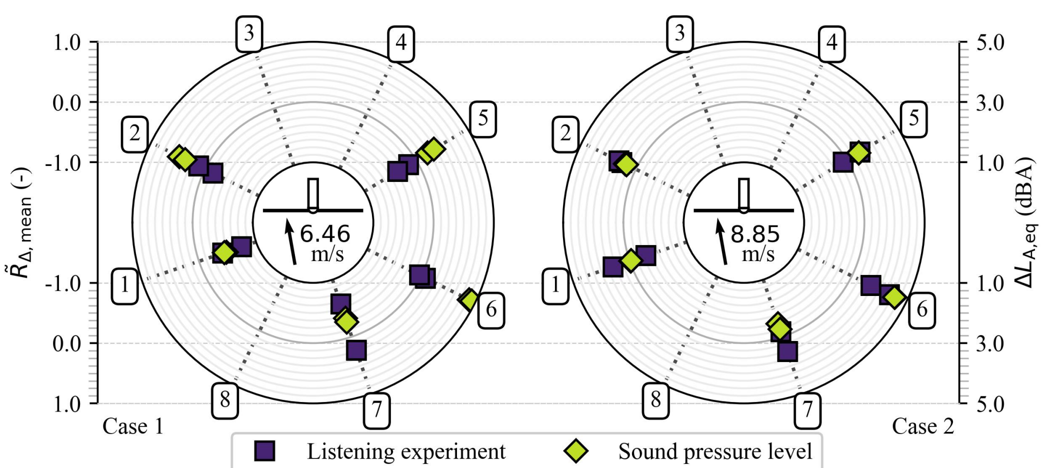


Figure 2: Listening experiment difference ratings against the differences in A-weighted equivalent sound pressure levels. Microphones 3, 4, and 8 were not included in the listening experiment.

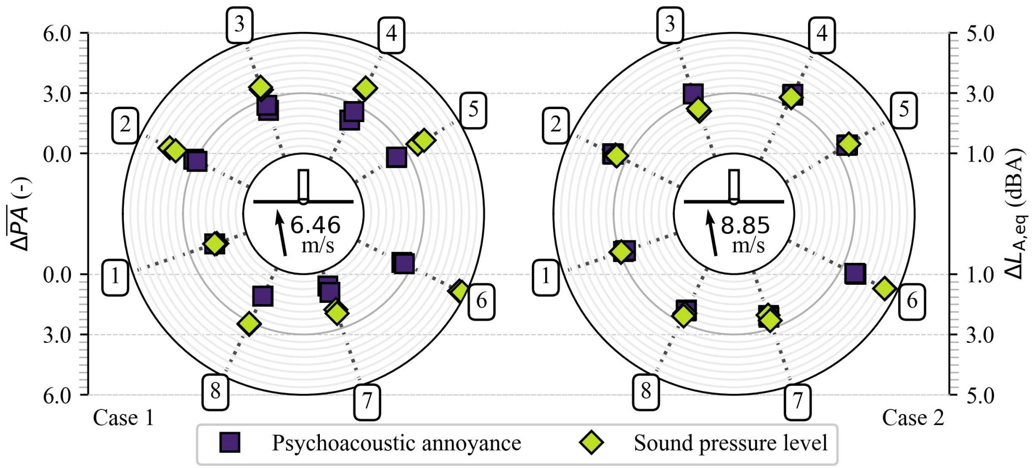


Figure 3: Differences in psychoacoustic annoyance against the differences in A-weighted equivalent sound pressure levels.

In Fig. 2, the results match only for some microphone positions. Microphone 6 presents a significant outlier, which is not a desirable behaviour for validation. The lower difference rating $\tilde{R}_{\Delta,\mathrm{mean}}$ in case 1 is not reflected in $\Delta L_{\mathrm{A,eq}}$.

The increase in perceptual difference with wind speed is an important finding of [6]. Therefore, the lack of its evidence in the sound pressure levels is a significant shortcoming of this analysis method. The driver of this perceptual difference is found in the high-frequency range of the sound spectra. Otherwise, the sound pressure levels present the directional aspect of the perceptual differences in a sufficiently accurate manner.

Figure 3 shows the results from the sound pressure level analysis in relation to the psychoacoustic annoyance results. In this case, the Pearson correlation coefficient is $\rho = 0.480$ with a p-value $p = 5.39 \times 10^{-3}$, indicating a weaker correlation than in the previous case, although still with statistical significance. The same trend is found as in Fig. 2, where some microphones present a better correspondence than others. Another similarity with Fig. 2 is the perceptual variation with wind speed that is reflected in $\Delta\overline{PA}$ and not in $\Delta L_{\mathrm{A,\,eq}}$. At the downstream microphones (2 to 5), which are not well represented in the listening experiment, this lack of wind speed variation is further remarked by the fact that the perceptual differences at these positions are largest.

4.2 Validation using sound quality metrics

This subsection shows the additional value of including SQMs in the validation of noise models for research on human perception. The data in Fig. 4 shows that the psychoacoustic annoyance metric presents a good match with the listening experiment results. The Pearson correlation coefficient in this case is $\rho = 0.623$ with a p-value $p = 17.9 \times 10^{-6}$. Clearly, the correlation is better than with sound pressure levels, while also having a greater statistical significance.

In terms of observable patterns in Fig. 4, there is a good correspondence between the SQMs and the listening experiment when it comes to the increased difference with wind speed. The directional differences are also very similar, without the significant outliers found in the sound pressure level analysis. Table 2 shows that the main contributor to the correlation is the loudness N, which has a larger correlation coefficient than PA. Differences in the other SQMs are less correlated to the results of the listening experiment. Some SQMs show little to no statistically significant correlation. The latter can be explained by the relative magnitudes of these SQMs, as found in [6].

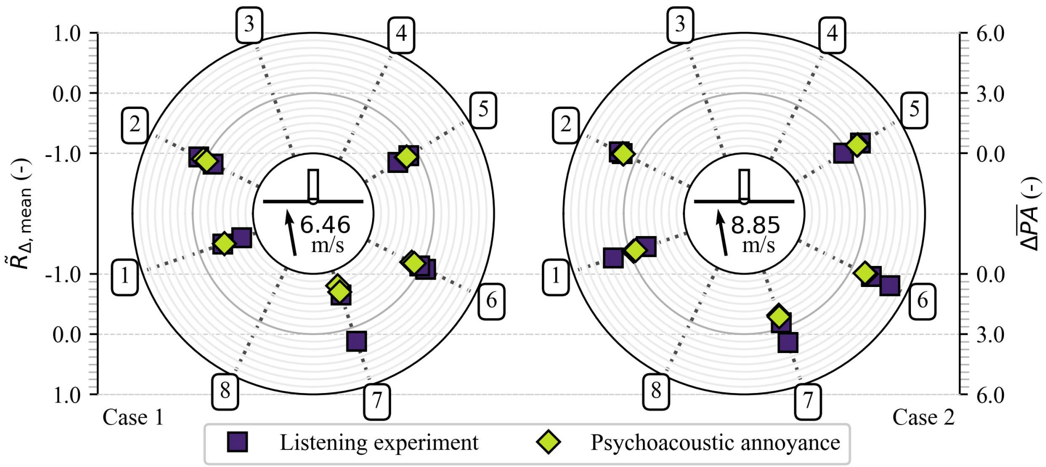


Figure 4: Listening experiment difference ratings against the differences in psychoacoustic annoyance. Microphones 3, 4, and 8 were not included in the listening experiment.

Table 2: Pearson correlation coefficients and corresponding p-values, between the indicated metrics, and the listening experiment difference ratings $\tilde{R}_{\Delta,\mathrm{mean}}$.

	ΔPA	ΔN	ΔS	ΔR	ΔFS	$\Delta L_{\mathrm{A,eq}}$
ρ	0.623	0.654	0.092	0.359	0.289	0.501
p-value	17.9×10^{-6}	4.72×10^{-6}	0.574	0.023	0.070	0.98×10^{-3}

5. Conclusions

This paper presented a methodology to auralise wind turbine noise from aeroacoustic models for human perception research. This methodology was evaluated against experimental measurements in three ways: (1) with A-weighted equivalent sound pressure levels, (2) psychoacoustic sound quality metrics, and (3) through a subjective listening experiment. These three methods were compared in their performance representing the human perception of the differences between the auralised and the measured noise. The analysis showed that the psychoacoustic annoyance sound quality metric has a better correlation with results found in the listening experiment than $L_{\mathrm{A,eq}}$. This improved correlation comes with an increased statistical significance, proving the additional value of SQMs in validating wind turbine noise models, for use in human perception research.

Having shown the additional value of the analysis and validation of wind turbine noise models with sound quality metrics, a sensitivity analysis will be conducted. The goal is to investigate how different operating conditions can influence human perception. Based on those findings, *WinTAur* will be developed further with different modelling approaches.

Acknowledgments

The authors thank Franck Bertagnolio and Andreas Fischer from the *Technical University of Denmark*, and Daniele Ragni from *Delft University of Technology* for their co-supervision of the Master's Thesis work from which this paper is derived [6]. The participants of the listening experiment are also recognised for their contribution to this research.

This publication is part of the project *Listen to the future* (with project number 20247) of the research programme Veni 2022 (Domain Applied and Engineering Sciences), awarded to Roberto Merino-Martínez, which is (partly) financed by the Dutch Research Council (NWO).

REFERENCES

1. Costanzo, G. and Brindley, G. Wind Europe, Wind energy in Europe - 2023 Statistics and the outlook for 2024-2030. Brussels, Belgium, (2024).

2. Kirkegaard, J. K., Rudolph, D. P., Nyborg, S., Solman, H., Gill, E., Cronin, T. and Hallisey, M. Tackling grand challenges in wind energy through a socio-technical perspective, *Nature Energy*, **8** (7), 655–664, (2023).

3. Merino-Martínez, R., Pieren, R., Schäffer, B. and Simons, D. G. Psychoacoustic model for predicting wind turbine noise annoyance, *Proceedings of the 24th International Congress on Acoustics*, Gyeongju, Korea, Oct., vol. A11: Psychoacoustics, (2022).

4. Pockelé, J. S., (2023), *WinTAur - Wind Turbine Auralisation tool*. V.1.0.0. Zenodo. DOI: 10.5281/ZENODO.10118451.

5. Bertagnolio, F., Madsen, H. A. and Fischer, A. A combined aeroelastic-aeroacoustic model for wind turbine noise: verification and analysis of field measurements, *Wind Energy*, **20** (8), 1331–1348, (2017).

6. Pockelé, J. S., *Auralisation of Modelled Wind Turbine Noise for Psychoacoustic Listening Experiments*, Master of Science Thesis, Delft University of Technology and Technical University of Denmark, Delft, Netherlands, (2023).

7. Bertagnolio, F., Madsen, H. A., Fischer, A. and Bak, C. Validation of an Aero-Acoustic Wind Turbine Noise Model Using Advanced Noise Source Measurements of a 500kW Turbine, *Proceedings of the 16th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery*, Apr., (2016).

8. International Electrotechnical Commission (IEC). IEC 61400-11:2012, Wind turbines - Part 11: Acoustic noise measurement techniques, (2012).

9. Greco, G. F., Merino-Martinez, R., Osses, A. and Langer, S. SQAT: a MATLAB-based toolbox for quantitative sound quality analysis, *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Chiba, Tokyo, Japan, Aug., pp. 7172–7183, (2023).

10. Greco, G. F., Merino-Martínez, R. and Osses, A., (2024), *SQAT: a sound quality analysis toolbox for MATLAB*. V1.1. Zenodo. DOI: 10.5281/ZENODO.10580337.

11. Fastl, H. and Zwicker, E., *Psychoacoustics: facts and models*, no. 22 in Springer series in information sciences, Springer, Berlin , Germany; New York, NY, USA, 3rd edn. (2007).

12. Merino-Martínez, R., von den Hoff, B. and Simons, D. G. Design and Acoustic Characterization of a Psycho-acoustic listening facility, Carletti, E. (Ed.), *Proceedings of the 29th international congress on sound and vibration*, Prague, Czech Republic, Jul., pp. 274–281, IIAV CZECH s.r.o., (2023).