Aircraft noise-induced annoyance analysis using psychoacoustic listening experiments

Merino Martinez, R.; Buzețelu, V.S.

Publication date
2025

Document Version
Final published version

Published in
Proceedings of the 11th Convention of the European Acoustics Association (Forum Acusticum - Euronoise

Citation (APA)
Merino Martinez, R., & Buzețelu, V. S. (2025). Aircraft noise-induced annoyance analysis using psychoacoustic listening experiments. In *Proceedings of the 11th Convention of the European Acoustics Association (Forum Acusticum - Euronoise*

AIRCRAFT NOISE-INDUCED ANNOYANCE ANALYSIS USING PSYCHOACOUSTIC LISTENING EXPERIMENTS

Roberto Merino-Martínez[1*] **Vlad Ştefan Buzeţelu**[1]
[1] Faculty of Aerospace Engineering, Delft University of Technology, The Netherlands

ABSTRACT

Aircraft noise annoyance is inherently subjective, and its accurate quantification represents a challenging task. There is a lack of consensus in the scientific community regarding which metrics are best for effectively representing this type of annoyance. The present study aims at relating various sound metrics to noise annoyance ratings measured in listening experiments featuring 60 aircraft flyover recordings (30 landings and 30 take-offs). This is done by considering different sound quality metrics (SQMs), psychoacoustic annoyance models, and more conventional noise certification metrics, such as the effective perceived noise level (EPNL) or the sound exposure level. A correlation analysis was subsequently performed on a large pool of sound metrics considering both linear and non-linear functions. The results show that, in general, metrics derived from psychoacoustic annoyance models (especially those proposed by Zwicker and Di *et al.*) present considerably better correlations compared to conventional metrics and most individual SQMs. The metrics of loudness, EPNL, and maximum perceived noise level (PNL) also exhibit strong correlations and capacity to predict a substantial portion of the variance observed in the reported annoyance ratings. Moreover, considering non-linear functions (e.g. logarithmic or hyperbolic tangent power) further improves the prediction performance.

Keywords: *psychoacoustic annoyance, sound quality metrics, aircraft noise annoyance, psychoacoustic listening experiments*

*__Corresponding author__: *r.merinomartinez@tudelft.nl.*

1. INTRODUCTION

Aircraft noise is the main source of annoyance for communities living in the vicinity of airports and has experienced an increasing trend in the past decades, which is in line with the prospects of continued growth of air traffic. Previous studies have shown that environmental noise exposure is correlated with severe health risks of strokes, coronary heart disease, and cardiovascular diseases, but also with psychosocial health concerns [1]. Furthermore, it seems that aircraft-induced noise generates higher degrees of annoyance compared to road and rail noise [2], which makes the need to address this issue particularly important. Despite the various technologies that have been implemented in aircraft in the past decades, which have mostly provided reductions in the engine noise levels [3], the aircraft airframe is also responsible for a large portion of the produced noise during landing. In the work of Merino-Martínez *et al.* [4], the strong tonal components arising from the Airbus A320 aircraft family's nose landing gear system were investigated, and it was found that they were strongly correlated with the velocity of the aircraft. Despite being perceived as highly annoying, such tonal components are typically ignored in the aircraft noise prediction models [5, 6].

Since noise annoyance is inherently subjective, there is still a lack of consensus in the scientific community regarding which sound metrics best capture the variance in annoyance responses. In some cases, some metrics are preferred over others with better performance due to their simplicity in implementation [7]. Traditional energy-based metrics, such as the Sound Pressure Level (SPL) have been augmented via more complex metrics, such as the Effective Perceived Noise Level (EPNL) and the A-weighted sound level, which take into account, to some extent, the human perception of noise through spectral irregularities, presence of tones in one-third-octave band spectra, etc. However, it seems that even these enhanced metrics fail to properly capture the large variance in annoyance responses. On top of all this, there is a correlation between demographic factors and the reported noise-induced annoyance, such as age, gender, background, individual noise sensitivity [8], as well as between visual factors and annoyance [9], which hinders even more the isolation of the main contributors toward the perceived annoyance.

Research in the field of psychoacoustics has bridged a significant portion of the knowledge gap concerning the subjective perception of sound through the emergence of so-called sound quality metrics (SQMs) [10, 11], which are computed based on the human's auditory system's characteristics, such as its varying sensitivity to different frequency bands. Starting from the five individual sound quality metrics, namely loudness, sharpness, tonality, roughness, and fluctuation strength, several psychoacoustic annoyance models have been developed throughout the years, such as those of Zwicker and Fastl [12], Di *et al.* [13], and More [14]. This study aims at identifying the sound metrics with the best annoyance predictive performance. To this end, a psychoacoustic listening experiment was conducted featuring 60 flyover recordings of conventional turbofan aircraft. The obtained annoyance raitings were subsequently employed in an exhaustive correlation analysis, starting from a wide range of conventional noise certification metrics, sound quality metrics, and psychoacoustic annoyance models.

This paper presents a short summary of the MSc thesis of Buzeţelu [15], which also dealt with the use of machine learning and artificial intelligence for aircraft noise annoyance prediction. The interested reader is referred to that document for further information.

2. METHODOLOGY

2.1 Aircraft flyover recordings

The data used throughout the research consists of 60 aircraft flyover recordings (30 take-offs and 30 landings) measured at Schiphol Amsterdam Airport using a microphone array. For this study, only data from one of the central microphones from the array was considered. A sampling frequency of 48 kHz was used. More details about the experimental setup can be found in [4, 16].

The initial overhead altitudes ranged from a minimum of around 100 m up to more than 400 m for take-offs and from approximately 40 m to just over 100 m for landings, which resulted in relatively high noise levels being measured. In order to limit the sound exposure for the participants in the listening experiments, all signals were scaled to an overhead altitude of 1500 m, which translates to equivalent A-weighted sound pressure levels ($L_{A,eq}$) ranging from 49.4 dBA to 70 dBA per audio file. The scaling was performed considering spherical spreading and atmospheric absorption corrections. Due to the relatively smaller distance between the measured aircraft and the microphone array, the time interval of interest for landings was inherently shorter than for take-offs. The selected audio files for the listening experiment had a duration of 16 s for take-offs and 10 s for landings.

2.2 Listening experiment campaign

A listening experiment campaign was conducted in order to obtain short-term annoyance responses from the aircraft flyover recordings considered. The experiments took place in the Psychoacoustic Listening Laboratory (PALILA) at the Faculty of Aerospace Engineering of Delft University of Technology, which is an extremely quiet, highly insulated facility [17]. A graphical user interface [1] ensured a smooth way of presenting the recordings to the participants while collecting the annoyance responses using the touchscreen of a laptop. An ICBEN 11-point scale was used to answer the following question: "*What grade from 0 to 10 best shows how much you would be bothered, disturbed, or annoyed by the sound of the aircraft in this recording?*". Before starting the experiment, the participants were requested to imagine that they were hearing the aircraft flyovers while at home or somewhere in their (hypothetical) residential area in the vicinity of an airport.

A total of 30 people participated in the experiment, out of which 21 were male and 9 female, with an average age of 23 years old (and a standard deviation of 3 years). Moreover, 21 were students and the other 9 were employed at the time of the experiment. All participants were in good health condition and had good self-reported hearing. The mean duration of the individual experiments was just above 20 min and 30 s (with a standard deviation of 2 min and 49 s). Compulsory breaks were also given to the participants every 5 recordings to reduce fatigue. The

[1] https://zenodo.org/records/11546254

participants were requested to read and sign an informed consent form, which was previously approved by the Human Research Ethics Committee from Delft University of Technology (form number 3599).

The 60 flyover recordings were split into three subsets of 40 recordings with a 50% overlap, and these were rotated equally among all test subjects. In other words, each person listened to 40 (20 take-offs and 20 landings) of the 60 flyover recordings, and each individual recording was evaluated by 20 people. Half the participants started off with the 20 take-offs and ended with the 20 landings, and vice versa. Additionally, within each of these two sets (take-offs and landings) the order of the recordings was completely randomised for each subject. The latter two measures were taken to minimize any potential learning effects of the listening order on the received annoyance responses. Finally, all participants were compensated for their time with a 10 € universal voucher upon completing the experiment.

2.3 Correlation analysis of the annoyance ratings

The large pool of conventional noise metrics and SQMs was calculated using the open-access MATLAB Sound Quality Analysis Toolbox (SQAT) [2]. An overview of the software can be found in [18].

The level of correlation of each metric with the noise annoyance ratings from the listening experiments was assessed by calculating both Pearson's and Spearman's correlation coefficients. The former relates to linear correlation, whereas the latter coefficient captures more complex, non-linear dependencies. Their formulations are provided in Eqs. (1) and (2), respectively, where x_i and y_i are, respectively, the i-th data points of variables x and y; $\bar{x}$ and $\bar{y}$ are the mean of the variables in consideration; $\mathrm{d_i}$ denotes the difference between the two ranks of each observation, and n the number of observations. Furthermore, an overview of the extracted conventional metrics and SQMs using SQAT is provided in Tab. 1 [3] and Tab. 2, respectively. In total, 173 individual metrics (including their statistical variations) were computed per aircraft flyover recording.

$$R_{\text{Pearson}} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{1}$$

$$R_{\text{Spearman}} = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{2}$$

Table 1. Psychoacoustic metrics extracted from SQAT.

Metric	Explanation
N	Loudness (ISO532-1) [19], [sone]
K	Tonality, as per Aures [20], [t.u.]
S	Sharpness (DIN45692) [21], [acum]
R	Roughness, as per Daniel and Webber [22], [asper]
FS	Fluctuation Strength, as per Osses *et al.* [23], [vacil]
$\mathrm{PA_{Zwicker}}$	Psychoacoustic Annoyance - Zwicker's model [12]
$\mathrm{PA_{Di}}$	Psychoacoustic Annoyance - model of Di *et al.* [13]
$\mathrm{PA_{More}}$	Psychoacoustic Annoyance - More's model [14]

Table 2. Conventional noise metrics extracted from SQAT.

Metric	Explanation
EPNL	Effective Perceived Noise Level, [EPNdB]
PNLM	Maximum Perceived Noise Level, [PNdB]
PNLTM	Maximum Tone-Corrected Perceived Noise Level, [PNTdB]
$\mathrm{L_{A_{eq}}}$, $\mathrm{L_{B_{eq}}}$, $\mathrm{L_{C_{eq}}}$, $\mathrm{L_{D_{eq}}}$, $\mathrm{L_{Z_{eq}}}$	A, B, C, D, and Z weighted equivalent Sound Pressure Level, [dB]
$\mathrm{LAF_{max}}$, $\mathrm{LBF_{max}}$, $\mathrm{LCF_{max}}$, $\mathrm{LDF_{max}}$, $\mathrm{LZF_{max}}$	Maximum values of A, B, C, D, and Z weighted fast-time weighting SPL, [dB]
$\mathrm{SEL_A}$, $\mathrm{SEL_B}$, $\mathrm{SEL_C}$, $\mathrm{SEL_D}$, $\mathrm{SEL_Z}$	A, B, C, D, and Z weighted Sound Exposure Level, [dB]

As a further proxy for the predictive power of the considered metrics, several functions were fitted between them and the annoyance responses. The functions are

[2] `https://github.com/ggrecow/SQAT`

[3] It should be noted that, for the SQMs and psychoacoustic metrics, several variations from the standard metrics were computed, including the 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95th percentiles, as well as the maximum, minimum, standard deviation, and mean values. In the case of the psychoacoustic models, the scalar values (values calculated based on the 5th percentile values of the SQMs used in their respective calculations) were also considered.

given in Eqs. (3) through (6) [4]. The first is the simple linear function (Eq. (3)) as normally employed in multiple studies from the literature. The second is the 10-base logarithmic function (Eq. (4)) because of the logarithmic nature of many of the noise certification metrics [24]. Furthermore, the logistic (Eq. (5)) and hyperbolic tangent power (Eq. (6)) functions were also fitted. The logistic function was observed to be particularly well-suited for relating psychoacoustic annoyance models to the percentage of highly annoyed people [25], whereas the hyperbolic tangent has a similar S-like shape that could also mimic typical results from psychoacoustic listening experiments. All four functions are relatively simple since the fits only require two parameters for tuning.

Linear:

$$\mathrm{PA}_{\mathrm{exp}} = bx + a \tag{3}$$

Logarithmic (base 10):

$$\mathrm{PA}_{\mathrm{exp}} = b \log_{10} x + a \tag{4}$$

Logistic:

$$\mathrm{PA}_{\mathrm{exp}} = \frac{10}{1 + \mathrm{e}^{-k(x - x_0)}} \tag{5}$$

Hyperbolic tangent power:

$$\mathrm{PA}_{\mathrm{exp}} = 10 |\tanh(k \cdot x)|^{b} \tag{6}$$

3. RESULTS & DISCUSSION

The annoyance ratings per operation (take-off and landing) averaged within all participants of the listening experiment are summarized in the form of a violin plot in Fig. 1. The take-offs were perceived, on average, as slightly more annoying than the landings, with overall average annoyance ratings of 6.11 and 5.75, respectively. It was also noticed that the spread in the obtained annoyance ratings is larger in the case of take-offs, whereas most of the responses are generally concentrated between 5 and 7, indicating that aircraft noise caused medium to high annoyance in the conditions tested within this research. The longer duration of the take-offs (16 s vs 10 s) may have influenced the obtained results to some extent. Nevertheless, the results are mostly in line with the expectations, since the take-off audio files considered were also generally louder than the landings. Previous studies have shown that metrics related to the magnitude of noise - as is the case for loudness - tend to be crucial predictors for annoyance, as far as environmental noise is concerned [12].

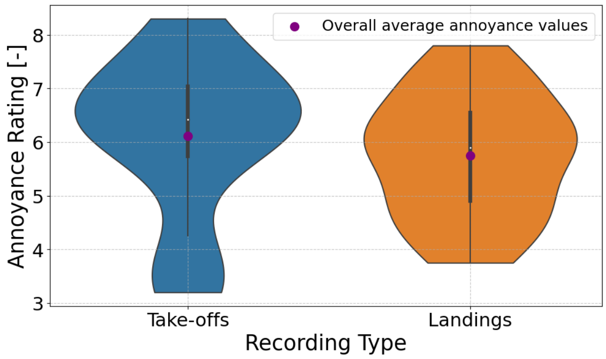

Figure 1. Violin plot of the listening experiment results. The width of the plot reflects the density of the averaged annoyance data (per recording) at each point. The purple dots denote the overall average annoyance ratings. The box plot within the violin plots consists of a white dot (median value), the box (which is the interquartile range), and the whiskers (lines extending from the box to show data points within $\pm$ 1.5 times the interquartile range).

Table 3 lists the Pearson and Spearman correlation coefficients for the metrics wich correlations higher than 0.80 [5]. All correlations considered in this paper had a p-value < 0.05. In general, metrics derived from psychoacoustics, along with the more complex conventional metrics, such as EPNL, PNLM, PNLTM, and some frequency weightings of the SPL show the greatest predictive potential for the annoyance ratings. Interestingly, apart from

[4] Note that, in this case, $\mathrm{PA}_{\mathrm{exp}}$ denotes the mean psychoacoustic annoyance rating reported in the listening experiment, while in the context of Tab. 1 it denotes metrics obtained from psychoacoustic annoyance models.

[5] In the case of loudness and the metrics resulting from the psychoacoustic models, there were multiple percentile values with correlation factors above 0.80 (from the 1st to the 30th percentile). However, these metrics were all considered to be very similar and their correlation factors were almost equal up to the 10th percentile, hence only the 5th percentile values (the ones normally used in literature), were kept for further analysis and denoted by the subscript "5". The standard deviation of N was also considered and denoted as N_{std}. The parameters $\mathrm{PA}_{\mathrm{Zwicker}}$, $\mathrm{PA}_{\mathrm{Di}}$, $\mathrm{PA}_{\mathrm{More}}$ are the scalar values of Zwicker's, Di's, and More's models, respectively, and their maximum values are denoted by the subscript "max".

loudness, N, none of the other four SQMs presented a strong correlation on their own. These metrics are, however, taken into account within the three psychoacoustic annoyance models, so their contribution is indirectly considered and perform, in general, better than N alone. This finding perhaps explains that, on their own, these other SQMs do not have a great predictive potential, but, when combined, they are able to explain a large portion of the variance. It is also a further confirmation of the fact that environmental noise annoyance is difficult to quantify and requires more sophisticated models, supporting the possibility of bridging this gap with the use of more complex tools like machine learning and artificial intelligence [15].

Table 3. Pearson and Spearman correlation coefficients (R), sorted in decreasing order (considering the average value). All cases had a p-value < 0.05).

Metric	Pearson Correlation	Spearman Correlation
PNLM	0.94	0.91
$PA5_{Zwicker}$	0.93	0.92
$PA5_{Di}$	0.93	0.92
N_5	0.93	0.91
$PA_{Zwicker}$	0.93	0.91
$PAmax_{Di}$	0.93	0.91
$PA5_{More}$	0.92	0.91
PA_{Di}	0.92	0.91
N_{max}	0.92	0.90
$PAmax_{Zwicker}$	0.93	0.90
PNLTM	0.93	0.89
LDF_{max}	0.92	0.86
LAF_{max}	0.91	0.87
$L_{D_{eq}}$	0.86	0.84
$PAmax_{More}$	0.86	0.83
EPNL	0.84	0.83
$L_{A_{eq}}$	0.84	0.84
PA_{More}	0.84	0.84
N_{std}	0.82	0.82
LZF_{max}	0.82	0.81
LBF_{max}	0.83	0.81
LCF_{max}	0.81	0.80
SEL_D	0.80	0.81

The superiority of psychoacoustic metrics is further confirmed by the results shown in the figures below, which present the coefficient of determination R^2 obtained from fitting the linear (Fig. 2), logarithmic (Fig. 3), logistic (Fig. 4), and hyperbolic tangent power functions (Fig. 5), respectively (see Eqs. (3)-(6)). The functions were fitted to the average annoyance ratings per recording (i.e. using the responses of the corresponding 20 participants per case), using the metrics from Table 3 as the variable x in Eqs. (3)-(6). Overall, the (variations of) metrics derived from the PA models of Zwicker, Di *et al.*, and More show the highest R^2 values, which means that they explain the largest amount of variance in the obtained annoyance responses. More's PA model performs slightly worse on the data compared to Zwicker's and Di's models, which is unexpected since this model was created based on aircraft noise characteristics. Nevertheless, the aircraft considered when developing this model were relatively old (mostly Boeing 757, MD-80, and Beechcraft 1900) [14]. The performance of the fits were also evaluated using the root-mean square error (RMSE) in the prediction [15]. For conciseness, only the RMSE values for the best performing cases in Tables 4 and 5 are reported.

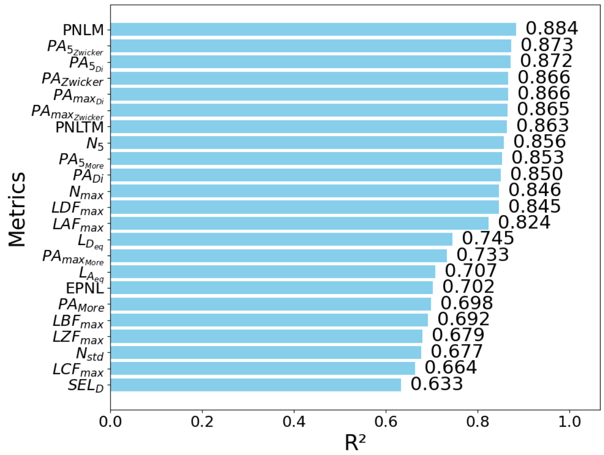

Figure 2. R^2 values per metric considered for the linear function fits.

In comparison, many of the more conventional noise certification metrics show a weaker predictive potential (although they are still strongly correlated with the annoyance ratings, as per Table 3), such as the EPNL and the Sound Exposure Level (SEL). These results also convey that even relatively simple functions are quite versatile and show a solid potential for relating the perceived annoyance to a multitude of metrics. An overview of the

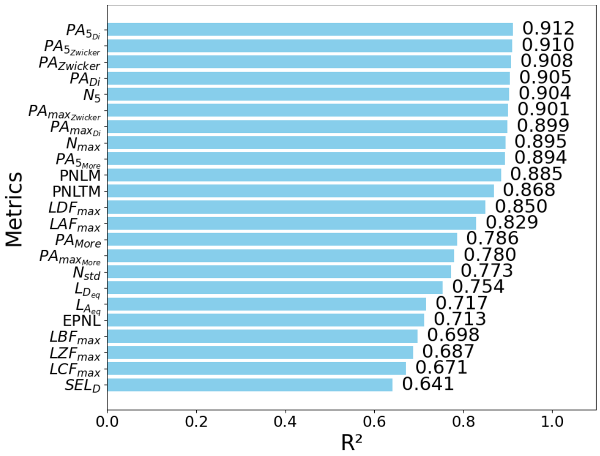

Figure 3. R^2 values per metric considered for the logarithmic function fits.

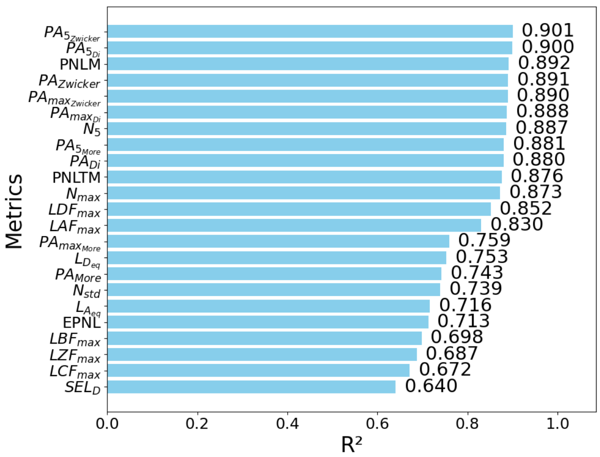

Figure 4. R^2 values per metric considered for the logistic function fits.

coefficients and the RMSE values of the functions corresponding to the respective best individual fits is provided in Table Tab. 4. Hence, it seems that when using one single metric to fit the annoyance ratings, the logarithmic and hyperbolic tangent power functions show marginally better performance compared to the logistic and linear functions. The best-performing fit obtained (hyperbolic tangent power for $\mathrm{PA}_{5_{\mathrm{Di}}}$) can be visualised in Fig. 6.

The analysis is further expanded to fitting the same functions to relate sound metrics to the percentage of highly annoyed people (%HA, typically defined as the

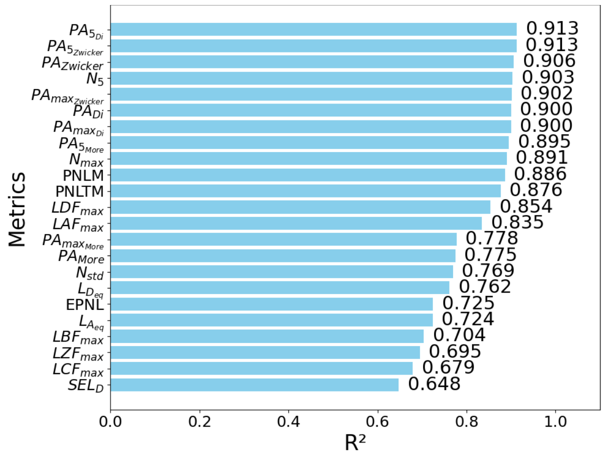

Figure 5. R^2 values per metric considered for the hyperbolic tangent power function fits.

Table 4. Parameters of the functions corresponding to the best fits, for the average annoyance ratings.

Function	R^2	RMSE	Metric	Coefficients
Linear	0.8835	0.468	PNLM	a = -15.14 b = 0.26
Logarithmic	0.9116	0.408	$\mathrm{PA}_{5_{\mathrm{Di}}}$	a = -7.60 b = 9.80
Logistic	0.9007	0.432	$\mathrm{PA}_{5_{\mathrm{Zwicker}}}$	x_0 = 19.84 k = 0.08
Tanh Power	**0.9128**	**0.405**	$\mathrm{PA}_{5_{\mathrm{Di}}}$	k = 0.029 b = 1.07

Table 5. Parameters of the functions corresponding to the best fits, for the percentage of highly annoyed people (%HA).

Function	R^2	RMSE (%)	Metric	Coefficients
Linear	0.8479	10.329	$\mathrm{PA}_{\mathrm{max}_{\mathrm{Di}}}$	a = -45.67 b = 3.08
Logarithmic	0.8654	9.719	$\mathrm{PA}_{\mathrm{Di}}$	a = -224.5 b = 191.06
Logistic	0.8557	10.061	PNLM	x_0 = 81.18 k = 0.27
Tanh Power	**0.8727**	**9.450**	$\mathrm{PA}_{5_{\mathrm{Di}}}$	k = 0.0599 b = 8.43

percentage of respondents with an annoyance rating larger than or equal to 7), since this metric is more commonly used in legislation regarding environmental noise annoyance. As such, in Table 5 the best fits are provided, in terms of the R^2 value and RMSE (in percentage). The

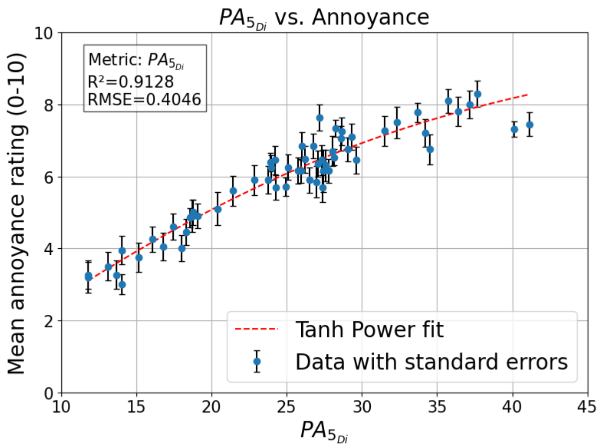

Figure 6. Best obtained fit with standard errors- hyperbolic tangent power function fitted using $\mathrm{PA}_{5_{\mathrm{Di}}}$.

findings confirm the fact that the logistic function is well-suited for this particular purpose ($R^2 = 0.8557$), but they also emphasize the slightly better performance achievable via the logarithmic and hyperbolic tangent power functions (over 86% and 87% of the variance is explained, respectively). Interestingly, the hierarchy among these functions remains the same as for the annoyance rating case in Table 4, although their corresponding best case scenario metrics slightly differ.

4. CONCLUSIONS & RECOMMENDATIONS

Quantifying aircraft-induced noise annoyance is a crucial challenge that needs to be addressed in order to mitigate the consequences of environmental noise pollution on the affected communities. This task is anything but trivial since the associated mechanisms are highly complex. Hence, identifying the main predictors of psychoacoustic annoyance is critical for this purpose.

The results from a listening experiment campaign featuring 60 aircraft flyover recordings showed that psychoacoustic metrics and their statistical variations are, in general, much better correlated to the reported annoyance ratings than conventional sound metrics. This shows that metrics that account for the characteristics of the human auditory system to a deeper level, such as the increased sensitivity to certain frequency ranges, amplitude modulations, or the presence of tones, are of paramount importance for the task of annoyance quantification. Furthermore, leveraging the non-linear, S-shaped functions of both the annoyance ratings and the percentage of highly annoyed people (%HA) is possible by using the logarithmic, logistic, and hyperbolic tangent power functions. In general, more than 90% of the observed variance could be explained among the annoyance ratings obtained from the 60 aircraft flyover recordings, and up to 87% in the case of %HA. Once again, almost all of the most promising fits were obtained using metrics derived from PA models, showcasing their strong predictive potential.

The main limitation of the study is the lack of a larger dataset to validate the current results further. Thus, it is highly recommended for future work to increase the number of available annoyance ratings corresponding to a larger dataset of aircraft flyover recordings and participants and to apply the methodology presented in this paper to assess its validity.

5. ACKNOWLEDGMENTS

The authors would like to thank Ir. Irina Besnea for kindly offering the flyover recordings used in this research and to Ir. Josephine S. Pockelé for the help given during the listening experiment campaign. Special thanks are also extended to the 30 participants in the listening experiment campaign. This publication is part of the project *Listen to the future* (with project number 20247) of the research program Veni 2022 (Domain Applied and Engineering Sciences) granted to Roberto Merino-Martinez which is (partly) financed by the Dutch Research Council (NWO).

6. REFERENCES

[1] World Health Organization, *Burden of disease from environmental noise: Quantification of healthy life years lost in Europe*. World Health Organization. Regional Office for Europe, 2011.

[2] L. Fredianelli, S. Carpita, and G. Licitra, "A procedure for deriving wind turbine noise limits by taking into account annoyance," *Science of the total environment*, vol. 648, pp. 728–736, 2019.

[3] W. Dobrzynski, "Almost 40 years of airframe noise research: what did we achieve?," *Journal of aircraft*, vol. 47, no. 2, pp. 353–367, 2010.

[4] R. Merino-Martínez, I. Besnea, B. von den Hoff, and M. Snellen, "Psychoacoustic analysis of the noise emissions from the airbus a320 aircraft family and its nose landing gear system," in *30th AIAA/CEAS Aeroacoustics Conference (2024)*, p. 3398, 2024.

[5] R. Merino-Martinez, L. Bertsch, M. Snellen, and D. G. Simons, "Analysis of landing gear noise during approach," in 22nd *AIAA/CEAS Aeroacoustics Conference, May 30 – June 1 2016, Lyon, France*, 2016. AIAA paper 2016–2769.

[6] R. Merino-Martinez and M. Snellen, "Implementation of tonal cavity noise estimations in landing gear noise prediction models," in 26th *AIAA/CEAS Aeroacoustics Conference, June 15 – 19 2020, Virtual Event*, 2020. AIAA paper 2020–2578.

[7] G. Malaval, "Approach to noise regulation of unmanned aviation in the european union," 2024.

[8] H. M. Miedema and H. Vos, "Demographic and attitudinal factors that modify annoyance from transportation noise," *The Journal of the Acoustical Society of America*, vol. 105, no. 6, pp. 3336–3344, 1999.

[9] T. J. Cox, "The effect of visual stimuli on the horribleness of awful sounds," *Applied acoustics*, vol. 69, no. 8, pp. 691–703, 2008.

[10] R. Merino-Martinez, R. Pieren, and B. Schäffer, "Holistic approach to wind turbine noise: From blade trailing–edge modifications to annoyance estimation," *Renewable and Sustainable Energy Reviews*, vol. 148, pp. 1–14, May 2021.

[11] R. Merino-Martinez, R. M. Yupa-Villanueva, B. von den Hoff, and J. S. Pockelé, "Human response to the flyover noise of different drones recorded in field measurements," in 3rd *Quiet Drones conference, September 8 – 11 2024, Manchester, United Kingdom*, 2024.

[12] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*, vol. 22. Springer Science & Business Media, 2013.

[13] G. Di, X. Chen, K. Song, B. Zhou, and C.-M. Pei, "Improvement of zwicker's psychoacoustic annoyance model aiming at tonal noises," *Applied Acoustics*, vol. 105, pp. 164–170, 2016.

[14] S. R. More, *Aircraft noise characteristics and metrics*. Purdue University, 2010.

[15] V. Buzetelu, "Aircraft-Induced Psychoacoustic Annoyance Quantification Using Artificial Intelligence," Master's thesis, Delft University of Technology, March 2025.

[16] R. Merino-Martinez, M. Snellen, and D. G. Simons, "Functional beamforming applied to imaging of flyover noise on landing aircraft," *Journal of Aircraft*, vol. 53, pp. 1830–1843, November–December 2016.

[17] R. Merino-Martínez, B. von den Hoff, and D. G. Simons, "Design and acoustic characterization of a psycho-acoustic listening facility," in *Proceedings of the 29th International Congress on Sound and Vibration, ICSV 2023*, Society of Acoustics, 2023.

[18] G. F. Greco, R. Merino-Martínez, A. Osses, and S. C. Langer, "Sqat: a matlab-based toolbox for quantitative sound quality analysis," in *Inter-Noise and Noise-Con Congress and Conference Proceedings*, vol. 268, pp. 7172–7183, Institute of Noise Control Engineering, 2023.

[19] "ISO norm 532–1 – Acoustics – Method for calculating loudness – Zwicker method," Tech. Rep. 1, International Organization for Standardization, 2017.

[20] W. Aures, "Procedure for calculating the sensory euphony of arbitrary sound signal. In German: Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale," *Acustica*, vol. 59, no. 2, pp. 130–141, 1985.

[21] G. von Bismark, "Sharpness as an attribute of the timbre of steady sounds," *Acta Acustica united with Acustica*, vol. 30, no. 3, pp. 159–172, 1974.

[22] P. Daniel and R. Webber, "Psychoacoustical Roughness: Implementation of an Optimized Model," *Accustica – acta acustica*, vol. 83, pp. 113–123, 1997.

[23] A. Osses, R. García León, and A. Kohlrausch, "Modelling the sensation of fluctuation strength," in 22nd *International Congress on Acoustics (ICA), September 5 – 9 2016, Buenos Aires, Argentina*, 2016.

[24] R. Merino-Martinez, R. Pieren, B. Schäffer, and D. G. Simons, "Psychoacoustic model for predicting wind turbine noise annoyance," in 24th *International Congress on Acoustics (ICA), October 24 – 28 2022, Gyeongju, South Korea*, 2022.

[25] G. Di, C. Cong, Y. Yao, D. Li, and W. Jian, "A study on the conversion relationship of noise perceived annoyance and psychoacoustic annoyance—a case of substation noise," *Journal of Low Frequency Noise, Vibration and Active Control*, vol. 41, no. 2, pp. 810–818, 2022.