inter.noise
21-24 AUGUST
SCOTTISH EVENT CAMPUS
GLASGOW 2022

Annoyance of helicopter-like sounds in urban background noise

Josef Schlittenlacher[1]
Department of Speech, Hearing and Phonetic Sciences
University College London
2 Wakefield Street
London WC1N 1PF
United Kingdom

Karine Wales
Division of Human Communication, Development and Hearing
University of Manchester
Oxford Road
Manchester M13 9PL
United Kingdom

ABSTRACT

Scenarios of urban air mobility see electric vertical take-off and landing aircraft (eVTOLs) operating within cities. Rotorcraft sounds are typically characterised by short bursts of noise, although eVTOLs offer more opportunities for a quieter sound design. We asked participants to compare the annoyance of a reference sequence of bursts of noise with a burst duration of 20 ms with that of a test sequence for which the burst duration was 1 or 5 ms. There were 20 bursts/s. A two-interval, two-alternative forced-choice task and a 1-up/1-down procedure was used. Both sequences were played in background noise that had either the same root-mean-square (RMS) level as the sequence of bursts or 10 dB less. The results were similar to those for loudness: On average, sequences with 1-ms bursts needed 6-8 dB less RMS level to sound equally annoying as the 20-ms bursts, and sequences with 5-ms bursts needed 2-4 dB less. This suggests that psychoacoustic annoyance is mainly explained by loudness and that the RMS level is an insufficient descriptor. Compared between the two background noise levels, the level difference for equal annoyance between short and 20-ms bursts was 1.5 dB larger in the louder background, which was statistically significant.

1. INTRODUCTION

Electric vertical takeoff and landing aircraft (eVTOLs) are about to become a novel sound source in urban environments. At the time of writing, there are about ten companies in an advanced stage of eVTOL development with funding of a few hundred million to a few billion US dollars each (Zorpette and Ackerman, 2022). A few hundred more companies with less funding add to the picture and may also deliver "air taxis" after scaling up from smaller drones. Some of the companies already operate more than one prototype and are in the process of aircraft certification and have several hundreds of orders each.

---

[1] also with University of Manchester, Division HCDH, josef.schlittenlacher@manchester.ac.uk

Although vehicle designs differ between manufacturers, eVTOLs share that they rely on rotors or similar propulsion for the desired vertical takeoff in urban air mobility. Early on, some in the industry focused on producing quiet vehicles (Josephson, 2017). In contrast to a helicopter with a single main rotor, eVTOLs allow for more opportunities in sound design in a combination of several rotors.

Most metrics for the assessment of aircraft noise are based on equivalent sound pressure level (L_{eq}). Examples that are used in regulations are the day-night-average sound level (DNL), the community noise equivalent level (CNEL) and the sound exposure level (SEL; ANSI/ASA S1.1-2013, sections 3.19, 3.20, 3.22). These metrics are problematic for rotorcraft and urban air mobility and incomplete for several reasons. First, these metrics do not consider that the sounds occur in background noise. Thus, they cannot reward a clever sound design that "hides" the vehicle sound in the urban background noise. Second, they do not consider the temporal shape of a sound. This is particularly important for eVTOLs because a more continuous and less impulsive sound is quieter at the same L_{eq} (Schlittenlacher and Moore, 2021). Third, the metrics ignore further aspects of perception like for example spectral loudness summation.

A metric that overcomes these limitations is the time-varying partial loudness model by Glasberg and Moore (2005; see also Glasberg and Moore, 2002; Moore et al. 2016, 2018). Among several other applications, it has been used for aircraft sounds (e.g., Marshall & Davies, 2011; Swift & Gee, 2011). Schlittenlacher and Moore (2021) investigated its performance for helicopter-like sounds in background noise.

Psychoacoustic annoyance or unpleasantness (Fastl and Zwicker, 2007, pp. 327 & 245; Fastl, 1998) is closely related to loudness. The emphasis of "psychoacoustic" underscores that those metrics aim at capturing the acoustic but not the social aspects. Ellermeier et al. (2004) asked participants to scale the annoyance of a variety of sounds and presented both original recordings and a version where the meaning of the sounds was neutralized. The calculated annoyance accounted for 86% of the variance of the evaluations of the neutralized sounds and still 73% for the original sounds.

The impact of a perceptual quantity on psychoacoustic annoyance can depend on the sound of interest. For example, tonality makes musical sounds more pleasant but environmental noises less pleasant. For this reason, formulae for psychoacoustic annoyance have been established for particular categories of sounds, for example gear units (Schlittenlacher and Ellermeier, 2013), photovoltaic converters (Ellermeier et al., 2014) or drone noise (Torija & Nicholls, 2022). In all studies, calculated loudness explained more than 75% of subjectively evaluated annoyance. Fastl and Stemplinger (2012) reported a Spearman correlation coefficient of 98% between loudness and annoyance for propeller noise.

This raises the question to what extent additional metrics like sharpness, roughness, tonality or impulsiveness improve annoyance predictions because they are an important contributor to psychoacoustic annoyance and to what extent the consideration of additional metrics leads to improvements due to correcting for the limitations of linear correlation models or minor discrepancies between calculated and subjectively evaluated loudness.

In the present study we asked participants to evaluate the annoyance of burst sequences in background noise with burst durations of 1, 5 and 20 ms. By comparing to the loudness evaluations of Schlittenlacher and Moore (2021) we could establish whether impulsiveness adds to annoyance or is already captured by loudness ratings. Furthermore, by using two different background levels we could quantify its impact on psychoacoustic annoyance.

2. METHOD

The method was almost identical with that used by Schlittenlacher and Moore (2021) but investigated annoyance instead of loudness. The experiment was done remotely, which still allowed us to study temporal features and relative levels. Measures were taken to control the absolute level, as described below.

2.1. Participants

Twenty participants completed the experiment. Their ages ranged from 22 to 62 years, with a median of 29 years and mean of 31 years. Nine of them identified as female, nine as male and two as non-binary. They reported to have normal hearing.

2.2. Apparatus

The experiment was run via the internet in the participants' home. They were asked to run it in a quiet room, to wear headphones and to use a computer or laptop. The audio files were stored in the wav file format and presented via the web browser, with underlying code in JavaScript and the Web Audio API.

Before the actual experiment started, the participants were asked to adjust their system settings so that the speech signal of a female speaker had a natural loudness. We assume that for most participants, this led to a level between 60 and 65 dB SPL for that signal. The root-mean-square (RMS) level of that speech sound was −36 dB relative to a full-scale sinusoid.

2.3. Stimuli

All stimuli consisted of a background noise and a helicopter-like sound, which was a sequence of noise bursts. To mitigate for the unknown frequency response of the headphones, all stimuli were bandpass-filtered with cut-off frequencies of 250 and 4000 Hz.

There were two background sounds. One was recorded at a busy street ("loud background") and the other in a suburban area ("soft background"). The loud background was always presented at the same level as the speech sound, the soft background at 10 dB less. Both background sounds had a duration of 2 s and were rather stationary parts that were cut out of longer recordings. The sounds had raised-cosine rise and fall times of 20 ms.

The helicopter-like signals were the same as those in Schlittenlacher et al. (2021). They were created from a stationary noise with the same average spectrum as a highway noise and a low crest factor. Bursts with durations of 1, 5 and 20 ms were cut out of that stationary noise using rectangular windows. These noise bursts were repeated at a rate of 20 Hz to result in a helicopter-like sound. Every second burst was the same as the preceding burst except for a phase shift of 180° to avoid any overall DC offset.

The sequences of noise bursts had durations of 1.6 s and started 200 ms after the background noise commenced. All levels are expressed as RMS levels over these 1.6 s, unless stated otherwise. All stimuli were presented diotically.

2.4. Procedure

The sequences with burst durations of 1 and 5 ms (target sounds) were each compared in annoyance to a sequence with burst duration of 20 ms (reference sound). This was done for each of the two background sounds, resulting in four experimental conditions (2 target sounds x 2 background sounds). The task was to indicate the more annoying of the two helicopter-like sounds in a two-interval, two-alternative forced-choice task. The target and reference sounds were presented in a random order and the two response buttons were highlighted in colour while the respective sound was being played.

Each target sound was compared to the reference sound in four runs of a 1-up/1-down procedure. In two runs the reference was fixed and the target was varied, while in the other two runs the reference was varied and the target was fixed. The fixed signal had the same level as the naturally loud speech sound. Thus, it had a level of 0 dB relative to the loud background and 10 dB relative to the soft background. The starting relative level of the variable sound was either +10 dB or −10 dB relative to the fixed signal. For each run, the 1-up/1-down procedure terminated after eight reversals. The step size was 5 dB until the second reversal occurred, then 3 dB until the fourth reversal occurred, and 1 dB thereafter. The mean level at the last four reversals was used to calculate the level difference needed for equal annoyance (LDEA) of the target and reference. The sixteen runs were performed in pseudo-random order.

3. RESULTS

To facilitate comparison of runs where the target was varied and runs where the reference was varied, the sign of the LDEA was inverted when the 20-ms bursts (reference) were varied in level and the 1 or 5-ms target bursts were fixed in level. A negative sign of the LDEA means that the sequence with the shorter bursts was perceived as being more annoying at an equal RMS level and needed a lower RMS level than the sequence with the 20-ms bursts for equal annoyance.

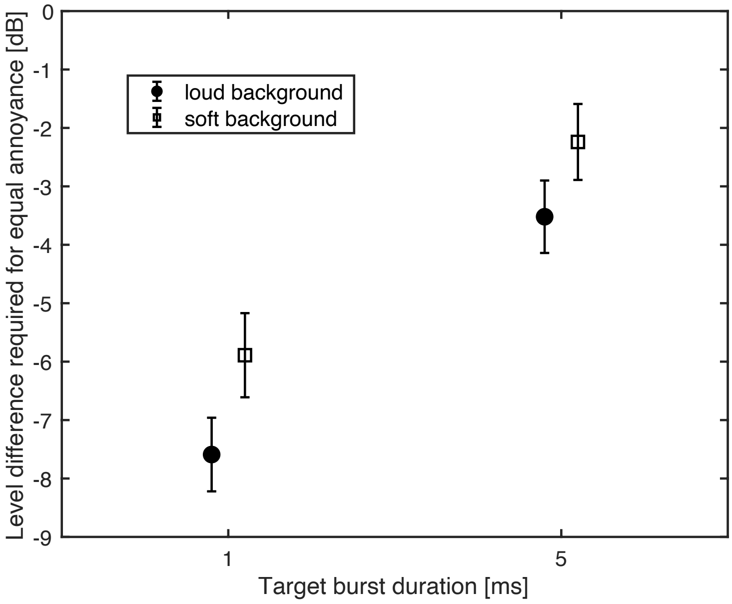

Figure 1: Level difference required for equal annoyance (LDEA). Negative values indicate that the short burst durations needed less level to sound equally annoying as the 20-ms reference.

Figure 1 shows the LDEAs between the sequences with burst durations of 1 and 20 ms (left), and 5 and 20 ms (right), averaged across participants, starting conditions and which of the two sounds was varied. Filled circles show the LDEAs in the loud background, open squares for the soft background. Error bars show the standard error of the mean.

A two-way 2x2 (target burst duration x background) within-subjects analysis of variances (ANOVA) showed a statistically significant main effect of target burst duration, $F(1,19)=178$, $p < 0.001$, $\eta_p^2 = 0.90$ and a statistically significant main effect of background noise level, $F(1,19) = 9.66$, $p < 0.01$, $\eta_p^2 = 0.34$. The interaction between the two was not statistically significant, $F(1,19) = 0.44$, $p = 0.52$, $\eta_p^2 = 0.02$.

The results for the loud background can be compared to the level differences for equal loudness (LDELs) measured by Schlittenlacher and Moore (2021). They used a similar background sound with the same level and the level of the fixed burst sequence was the same as that of the background noise in both studies. The experiment was also completed by 20 participants.

Table 1: Comparison of level differences for equal annoyance (LDEA) and loudness (LDEL) for the loud background noise.

Burst duration [ms]	**LDEA [dB]**	**LDEL [dB]**
1	-7.6	-6.6
5	-3.5	-3.1

Table 1 shows that the LDEAs were very similar to the LDELs and the difference between them amounted to at most 1 dB. A two-way 2x2 ANOVA on the level difference as dependent variable (LDEA or LDEL, respectively) with within-subjects factor "burst duration" (1 or 5 ms) and between-subjects factor "study" (the present study or Schlittenlacher and Moore, 2021) showed a statistically significant main effect of burst duration, $F(1,38) = 130$, $p < 0.001$, $\eta_p^2 = 0.77$. The main effect of study was not statistically significant and had an effect size close to zero, $F(1,38) = 0.682$, $p = 0.41$, $\eta_p^2 = 0.02$. This suggests that shorter burst durations lead to both more loudness and more annoyance at an equal RMS level.

4. CONCLUSIONS

The present study investigated the psychoacoustic annoyance of sounds in background noise as a function of impulsiveness, which was controlled by varying the durations of bursts. Sequences with shorter bursts were perceived as more annoying than longer bursts. A sequence with 1-ms bursts needed 6 to 8 dB less in L_{eq} to be equally annoying as a sequence with 20-ms bursts. This rather large difference is ignored by metrics that are based on equivalent sound pressure.

The effect of the level of the background noise amounted to about 1.5 dB in the LDEA for the difference of 10 dB in background noise level. This result advocates the use of perceptual models that incorporate the presence of background sounds (e.g., Glasberg and Moore, 2005) in scenarios like urban air mobility.

In general, predictions of psychoacoustic annoyance can be improved to some extent when considering the sensations of sharpness, roughness or tonality in addition to loudness (e.g., Fastl, 1998) even if loudness alone explains 75% or more of the variance in annoyance ratings (see introduction).

Although impulsiveness is an important sensation for judging sounds like noise from diesel engines (Hashimoto, 2000), its role for the annoyance of rotorcraft noise seems to be already captured by its effect on loudness. The present study showed that LDEAs were almost the same as the LDELs measured by Schlittenlacher and Moore (2021), and Fastl and Stemplinger (2012) found a near-perfect correlation between loudness and psychoacoustic annoyance for propeller noise.

6. REFERENCES

1. American National Standards Institute/Acoustical Society of America (ANSI/ASA), *Acoustical Terminology*, S1.1 2013 Edition.
2. Ellermeier, W., Zeitler, A. & Fastl, H. Predicting annoyance judgments from psychoacoustic metrics: Identifiable versus neutralized sounds. *Proceedings of Internoise 2004*, Prague, Czech Republic (2004).
3. Ellermeier, W., Kattner, F., Kurtze, L. & Bös, J. Psychoacoustic characterization of the noise produced by photovoltaic inverters. *Acta Acustica united with Acustica*, **100(6)**, 1120 – 1128 (2014).
4. Fastl, H. Psychoacoustics and sound quality metrics. *Proceedings of the 1998 Sound Quality Symposium*, 3 – 10 (1998).
5. Fastl, H. & Zwicker, E. *Psychoacoustics: Facts and Models*, 3rd edition, Springer, 2007.
6. Fastl, H. & Stemplinger, I. Psychoacoustic evaluation of noises produced by propellers with asymmetrical blade spacing. *Proceedings of Internoise 2012*, New York City, NY (2012).
7. Glasberg, B. R. & Moore, B. C. J. A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society,* **50(5)**, 331 – 342 (2002).
8. Glasberg, B. R. & Moore, B. C. J. Development and evaluation of a model for predicting the audibility of time-varying sounds in the presence of background sounds. *Journal of the Audio Engineering Society,* **53(10)**, 906 – 618 (2005).
9. Hashimoto, T. Sound quality approach on vehicle interior and exterior noise – Quantification of frequency related attributes and impulsiveness. *Acoustical Science and Technology (E)*, ***21*(6)**, 337-340 (2000).

10. Josephson, D. eVTOL community noise overview, *Uber Elevate Summit 2017*, Dallas, TX (2017).
11. Marshall, A. & Davies, P. Metrics including time-varying loudness models to assess the impact of sonic booms and other transient sounds. *Noise Control Engineering Journal,* **59(6)**, 681 – 697 (2011).
12. Moore, B. C. J., Glasberg, B. R., Varathanathan, A. & Schlittenlacher, J. A loudness model for time-varying sounds incorporating binaural inhibition, *Trends in Hearing*, **20**, 1 – 16 (2016).
13. Moore, B.C.J., Jervis, M., Harries, L. & Schlittenlacher, J. Testing and refining a loudness model for time-varying sounds incorporating binaural inhibition, *Journal of the Acoustical Society of America*, **143(3)**, 1504 – 1513 (2018).
14. Schlittenlacher, J. & Ellermeier, W. Psychoacoustic evaluation of gear noise using category ratings of multiple attributes, *Proceedings of Internoise 2013*, Innsbruck, Austria (2013).
15. Schlittenlacher, J. & Moore, B. C. J. Temporal integration of partial loudness of helicopter-like sounds. *Proceedings of Inter-noise 2021*, Washington D.C. (2021).
16. Swift, S. H. & Gee, K. L. Examining the use of a time-varying loudness algorithm for quantifying characteristics of nonlinearly propagated noise (L). *Journal of the Acoustical Society of America*, **129(5)**, 2753 – 2756 (2011).
17. Torija, A. J. & Nicholls, R.K. Investigation of metrics for assessing human response to drone noise. *International Journal of Environmental Research and Public Health*, **19(6)**, 3152 (2022).
18. Zorpette, G & Ackerman, E. What's behind the air-taxi craze? *IEEE Spectrum,* March 2022.