EXPERIMENTAL STUDY ON IMPULSE CHARACTERISTICS AFFECTING PERCEIVED IMPULSIVENESS AND ANNOYANCE

Felix Hochbaum[1], Lukas Billerbeck[1], André Fiebig[1]

[1] *Department of Engineering Acoustics, Technische Universität Berlin, Einsteinufer 25, 10587 Berlin, Germany*

Abstract

Impulsive components can substantially increase noise annoyance, motivating the inclusion of impulse penalties in environmental noise assessment. While existing research and impulse penalty procedures mainly focus on level- and onset-related characteristics, the roles of the spectral content of impulses and temporal irregularity of impulse occurrences have received less attention. To explore these aspects further, this study investigated how selected impulse characteristics affect perceived impulsiveness and annoyance in a controlled laboratory experiment. Thirty participants rated synthetically generated stimuli with impulses systematically varied in level difference, onset rate, irregularity, and spectral content of the impulse. Ratings were compared with model predictions to assess the explanatory power of different approaches. For the stimuli considered in this study, a more complex method incorporating onset-related characteristics provided the most accurate prediction of perceived impulsiveness, whereas a loudness-based metric predicted annoyance well. A simple level percentile-based approach predicted both well. Level difference and the spectral content had statistically significant effects on both rating criteria, while onset rate and regular vs. irregular impulse occurrences showed no significant influence. However, methodological constraints, particularly difficulties in impulse-level scaling and loudness differences between the impulses due to the different spectral content, limit generalizability of the observed effects. The study thus provides relevant considerations for future experiments and highlights the potential roles of psychoacoustic loudness and onset detection in modelling perceived impulsiveness.

Keywords: *Impulsiveness, impulsive sound, environmental noise, psychoacoustics, impulse perception, annoyance*

***Corresponding author**: felix.hochbaum@tu-berlin.de.*

1. Introduction

Noise exposure is one of the most widespread environmental burdens in modern societies and is associated with adverse health effects [1]. While environmental noise assessment is commonly based on the energy-equivalent continuous sound pressure level indicator, this metric cannot fully account for the temporal structure of sounds [2]. A limitation that is particularly relevant for impulsive sounds, such as hammer blows, door slams, or construction noise, which, despite their short duration and limited contribution to the energy-equivalent continuous level, are often perceived as highly salient and more annoying than continuous noise at the same energy-equivalent sound level [3]. Although the concept of impulsive sound is not uniformly defined in detail, it is generally associated with brief high sound-pressure events with a sudden onset. The ISO/PAS 1996-1 further indicates that a single impulsive event typically lasts less than one second [4]. Impulsiveness refers to the perceived strength of such components and may depend on their magnitude, number, temporal distribution, and spectral characteristics [5].

Several standardized and non-standardized procedures have been proposed to identify, classify, or quantify impulsive components (e.g., [5], [6], [7], [8], [9]). They differ in their intended application, computational complexity, and underlying signal representation. While many approaches derive penalties for noise assessment to account for the increased annoyance caused by impulsive components, others more directly quantify impulsiveness as a perceptual attribute. Conceptually, the methods can be categorized according to whether they are level-based, for example using A-weighted sound pressure levels with Fast time weighting, or auditory-model-based, for example using time-varying loudness. They also differ in whether impulsive content is identified from statistical properties of the signal, such as percentile-based quantities, or through the explicit detection of individual impulse events based on onsets (steep rising segments in the sound pressure level time history). Based on these differences, a set of representative models was selected for the present study. The method standardized in ISO/PAS 1996-3 [7], which builds on the earlier Nordtest method NT ACOU 112 [6], evaluates impulsiveness at the level of individual sound events. Candidate impulses are identified from their onsets and characterized by the onset rate, describing the steepness of the level rise in decibels per second, and the level difference between the beginning and end of the onset. These quantities are combined into a measure of impulse prominence, which is converted linearly into an impulse adjustment once a specified threshold is exceeded. Based on the resulting adjustment, sound sources are classified as non-impulsive, regular impulsive, or highly impulsive. The method is intended for impulsive sound sources other than high-energy impulsive sounds, such as gunfire or explosions. Because it focuses on local rise characteristics of individual events, other temporal properties, including impulse duration, decay, and repetition rate, are not explicitly taken into account. The method specified in the German standard DIN 45645-1 [8] and used for noise assessment evaluates the magnitude of impulsiveness from the temporal variation of the A-weighted sound pressure level. The measurement period is divided into consecutive sampling intervals, typically of 5 s duration, and the highest Fast time-weighted level is determined within each interval. These interval maxima are then energetically averaged and compared with the equivalent continuous A-weighted sound pressure level over the full measurement period. The resulting impulse adjustment corresponds to the positive level difference between these two quantities and is therefore applied only when the averaged short-term maxima clearly exceed the energy-equivalent level (by at least 3 dB). Hongisto et al. [10] evaluated, among other approaches, a simple percentile-based model that quantifies impulsiveness from the difference between the A-weighted sound pressure levels exceeded during 1% and 99% of the measurement time. This difference is converted linearly into an impulse adjustment, without requiring the explicit detection of individual impulses. The loudness-based measure proposed by Willemsen and Rao [5] quantifies impulsiveness from time-varying loudness calculated using the Zwicker-based procedure specified in DIN 45631/A1 [11]. At each time step, the instantaneous loudness is compared with a local baseline defined by the loudness exceeded during 95% of a moving 1 s window, representing the non-impulsive component. The resulting descriptor represents the mean excess loudness above this baseline and therefore accounts for both the magnitude and temporal extent of impulsive components. In addition, a calculation method based on the Sottek Hearing Model (SHM) [9] is available in commercial software. The method incorporates impulsive sound characteristics into a comprehensive model of auditory processing and thus evaluates impulsiveness as a perceptual attribute rather than as an isolated signal-based descriptor. While the effects of level difference, onset rate, and spectral content on annoyance have been addressed by previous research (e.g. see [3], [12]), the influence of temporal regularity — whether impulses occur regularly or irregularly — has, to our knowledge, not yet been systematically investigated. To address this gap and broaden the empirical basis for understanding the perception and annoyance of impulsive sounds, a listening experiment was conducted using synthetically generated stimuli with factorial variations in level difference, onset rate, spectral content, and temporal regularity. Modified real-world stimuli and synthetic reference noises were also included, and all sounds were rated with respect to perceived impulsiveness and annoyance. Several existing calculation methods for assessing impulsiveness are considered in the subsequent analysis and discussion.

2. Method

2.1. Stimuli

The original stimulus set consisted of 27 stimuli: 16 synthetic stimuli, eight modified real-world stimuli, and three pink-noise references without impulse components. The synthetic stimuli followed a full-factorial $2 \times 2 \times 2 \times 2$ design, varying onset rate (OR: 100 and 1000 dB/s), level difference between peak and background level (LD: 15 and 25 dB), temporal irregularity (IR: regular (0.0) and irregular (0.6)), and noise form (NF: pink and blue noise). For the two modified real-world sounds, Hammer and Smith, a 2×2 design varying LD and IR was intended. Due to a stimulus-generation error, the four nominally regular conditions did not exhibit the intended temporal characteristics and were therefore excluded from the main analyses. All stimuli had a duration of 5 s and were generated at a sampling rate of 48 kHz.

Each impulsive stimulus consisted of stationary pink-noise background and four superimposed impulse components, with the general stimulus structure informed by [3]. For the synthetic stimuli, each impulse component was generated by multiplying a noise signal, either pink or blue noise depending on NF, with a temporal envelope. The envelope consisted of a linear rise to a maximum followed immediately by an exponential decay of the sound pressure level over 0.3 s, corresponding to 14 400 samples at 48 kHz. The resulting impulse waveform was then added to the background noise at predefined temporal positions, as illustrated in Fig. 1.

Noise was generated with predefined seeds. Pink noise was generated with a spectral slope of -3 dB/oct, corresponding to equal energy per octave. Blue noise was generated with the complementary spectral characteristic of $+3$ dB/oct, resulting in increasing energy with frequency. The same pink noise was used as background noise for all impulsive stimuli, with its amplitude kept constant across conditions. Impulse amplitudes were scaled relative to the maximum amplitude of the background noise according to the intended LD condition. Because the impulse components were scaled on a peak-amplitude basis without subsequent

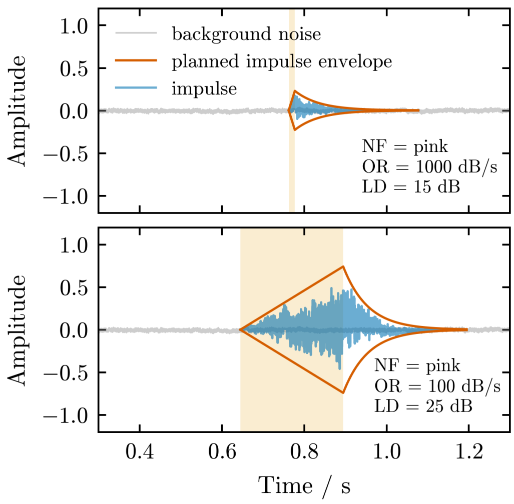

Figure 1: Exemplary visualization of two synthesized single impulses. The background noise is shown in grey, the planned impulse envelope in orange, and the resulting impulse component in blue. The corresponding parameter values are indicated in each panel.

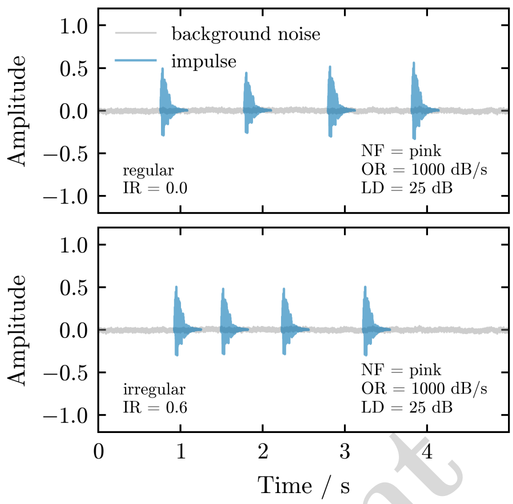

Figure 2: Exemplary comparison of regular and irregular impulse placement. Background noise is shown in grey and the isolated impulse component in blue.

level adjustment, the resulting overall sound levels varied slightly between stimuli. Background-noise references contained no impulse components. The baseline reference used the same pink-noise realization and amplitude as the background component of all impulsive stimuli, while two additional reference noise sounds were adjusted to the mean L_{eq} of the respective LD15 and LD25 conditions. These mean levels were 58.9 dB for LD15 ($SD = 0.69$ dB) and 67.1 dB for LD25 ($SD = 1.6$ dB). Temporal irregularity (IR) was manipulated by varying the temporal positions of the four impulse occurrences. In the regular condition, impulse centres were placed at $t = [0.9, 1.9, 2.9, 3.9]$ s. In the irregular condition, suitable impulse centres were determined once using a random algorithm and then kept identical across all sound conditions, resulting in $t = [0.95, 1.52, 2.26, 3.25]$ s. This ensured that irregularity was introduced consistently across stimuli, as illustrated in Fig. 2.

2.2. Apparatus

The experiment was carried out in an audiometry booth at TU Berlin. Stimulus presentation and response collection were controlled by a custom test interface. The signals were played back via an RME Fireface UCX II audio interface and presented over Sennheiser HD 600 headphones. Playback levels were calibrated using a HEAD acoustics HMS III artificial head measurement system. The energy-equivalent sound pressure level of the base background-noise was measured as $L_{eq} = 47.7$ dB and $L_{Aeq} = 43$ dB(A); these values were used as the reference for all subsequent level calculations.

2.3. Procedure

Participants provided informed consent and completed a pre-experimental questionnaire covering age, gender, self-reported hearing impairments, whether they considered themselves experienced or trained listeners, and previous experience with listening experiments. All data were anonymized using an individual participant code. Before requesting ratings of sounds, participants provided their own brief definition of impulsive sound and were then presented with a working definition for generating a consistent frame of reference for all participants. In each part, all 27 nominal stimuli were presented three times in randomized order. Participants first rated perceived impulsiveness and subsequently annoyance on unipolar scales from 0 to 10 using the questions *"How impulsive do you perceive the signal?"* and *"How annoying do you find the signal?"*, respectively.

2.4. Participants

Participants were recruited among students and staff at TU Berlin. A total of 30 participants with a mean age of $M = 32.6$ ($SD = 5.82$) participated in the experiment. Eleven participants identified as female, 18 as male, and one participant preferred not to provide gender information. Eleven participants reported previous experience with listening experiments, and 12 participants classified their hearing abilities as trained. Twenty-eight of the 30 participants self-reported no hearing impairments. Two participants reported slight tinnitus, which was not considered a severe limitation for participation in the experiment.

3. Results

The repeated judgments were averaged for each participant and stimulus, resulting in one impulsiveness rating and one annoyance rating per participant–stimulus combination. Agreement across participants in the stimulus ratings was assessed separately for both rating dimensions using an average-measures intraclass correlation coefficient, ICC(2,k), based on a two-way random-effects model with absolute agreement [13].

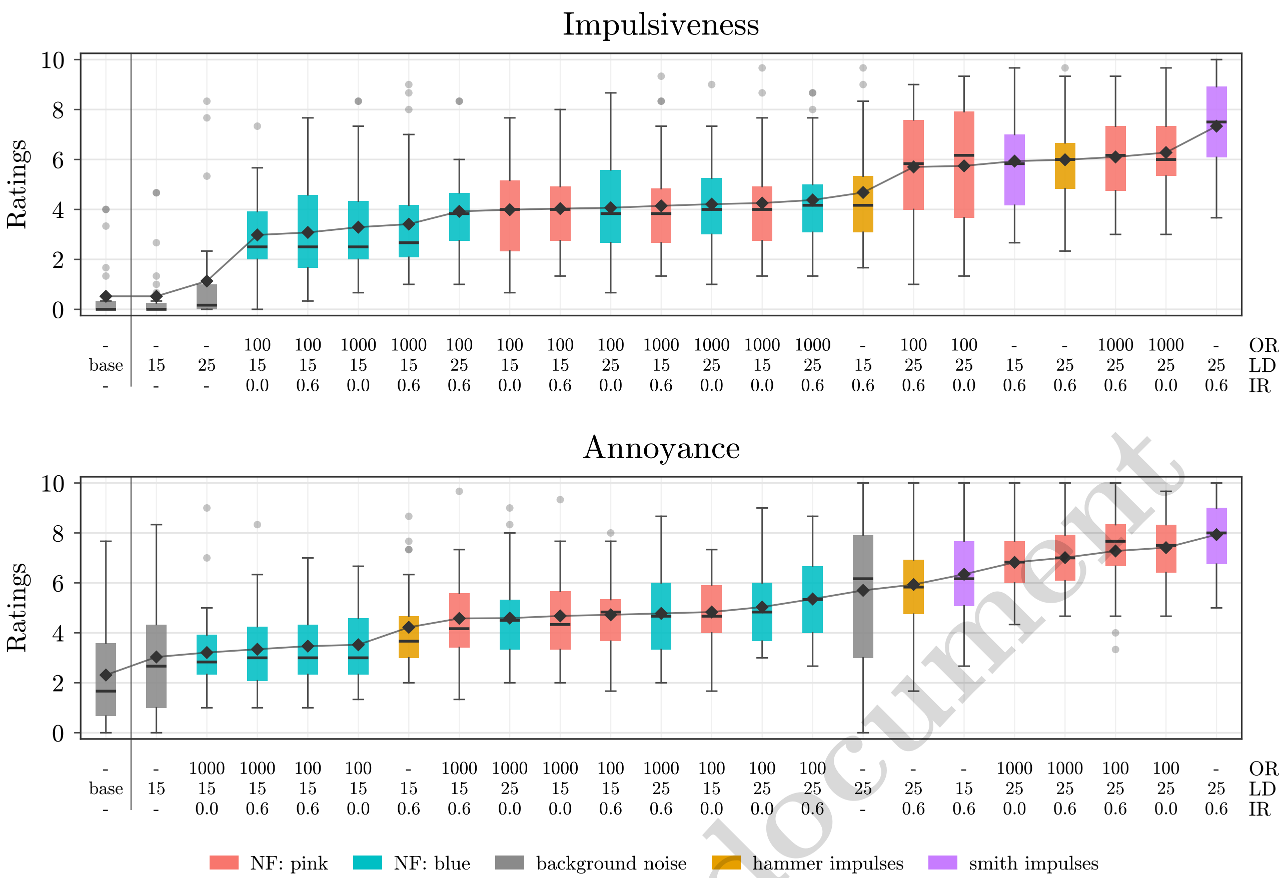

Figure 3: Participant ratings for perceived impulsiveness (top) and annoyance (bottom), ordered by ascending stimulus mean separately within each panel. Boxes show medians and interquartile ranges, whiskers extend to 1.5 times the interquartile range, and circles indicate outliers. Diamonds denote stimulus-wise means connected by a grey guide line. OR, LD, and IR denote onset rate, level difference, and temporal irregularity; dashes indicate non-applicable parameters. For the background-noise references, LD15 and LD25 denote levels matched to the mean L_{eq} of the respective stimulus conditions. Colours distinguish NF conditions, background-noise references, Hammer impulses, and Smith impulses.

The average ratings showed high inter-participant reliability for both impulsiveness, $\mathrm{ICC}(2,k) = .907$, 95% CI $[.841, .951]$, and annoyance, $\mathrm{ICC}(2,k) = .903$, 95% CI $[.836, .949]$. Following the reliability assessment, the four nominally regular Hammer and Smith conditions were excluded from subsequent analyses due to the stimulus-generation error, leaving 23 stimuli. Subsequent analyses used the participant-wise mean ratings for each retained stimulus, whose distributions for impulsiveness and annoyance are shown in Fig. 3. Two full-factorial $2 \times 2 \times 2 \times 2$ repeated-measures ANOVAs were fitted, one for each dependent variable: perceived impulsiveness and annoyance. Noise form (NF), level difference (LD), onset rate (OR), and temporal irregularity (IR) were included as within-subject factors, along with all possible interactions. The results are reported in Tables 1 and 2, respectively. For perceived impulsiveness, significant main effects of level difference and noise form were observed, as well as a significant interaction between level difference and noise form. Onset rate and temporal irregularity showed no significant main effects. For perceived annoyance, significant main effects of level difference, noise form, and onset rate were observed. These effects were qualified by significant interactions between level difference and noise form and between level difference and onset rate. Effects involving noise form should be interpreted with caution, as they may partly reflect unintended level differences resulting from the peak-based scaling procedure and the absence of subsequent level matching

Table 1: Repeated-measures ANOVA for impulsiveness in the full-factorial synthetic stimulus set. All main effects and significant two-way interactions are reported.

Effect	$F(1,29)$	p	η_p^2
LD	54.62	$< .001^{***}$	.653
OR	2.08	.160	.067
IR	0.01	.943	$< .001$
NF	95.90	$< .001^{***}$	.768
LD × NF	26.07	$< .001^{***}$	.473

Note. LD = level difference; OR = onset rate; IR = temporal irregularity; NF = noise form. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. All remaining two-way and higher-order interactions were non-significant.

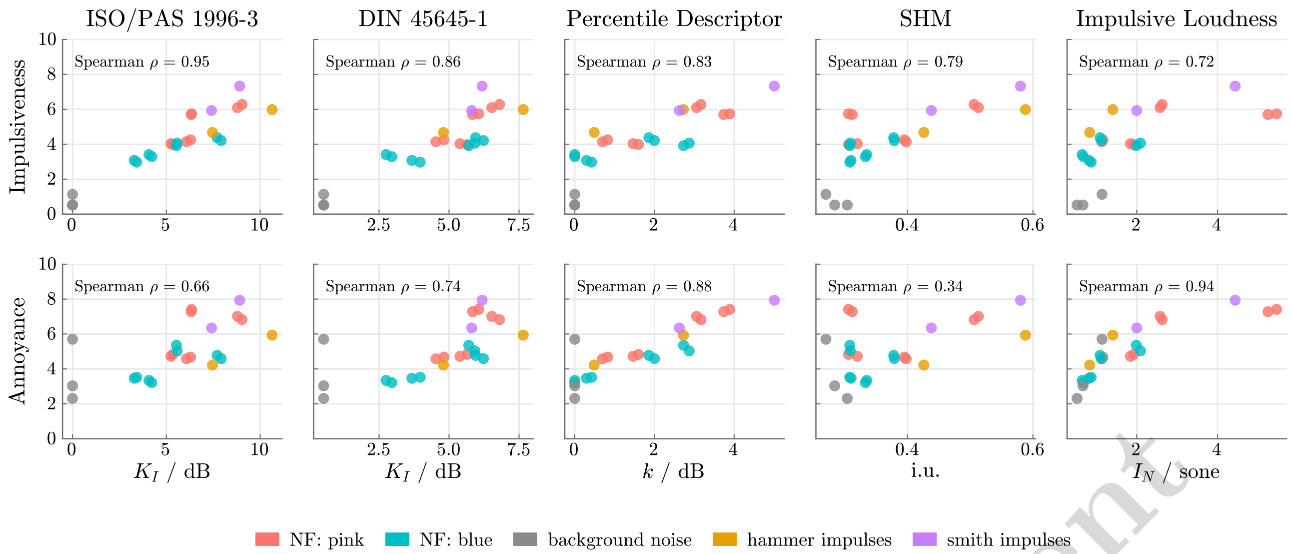

Figure 4: Relationships between the model predictions and the mean participant ratings for perceived impulsiveness (top row) and annoyance (bottom row). Each point represents one stimulus, with colours indicating the respective stimulus category. Spearman rank-order correlations were calculated across stimuli between the mean participant ratings and the corresponding predictions of ISO/PAS 1996-3 [7], DIN 45645-1 [8], the Percentile Descriptor [10], SHM [9], and the Impulsive Loudness measure [5]. The resulting Spearman correlation coefficient ρ is reported in each panel.

Table 2: Repeated-measures ANOVA for annoyance in the full-factorial synthetic stimulus set. All main effects and significant two-way interactions are reported.

Effect	$F(1, 29)$	p	η_p^2
LD	202.90	$< .001^{***}$	.875
OR	6.92	$.014^{*}$	.193
IR	0.45	.507	.015
NF	87.69	$< .001^{***}$	.751
LD × OR	4.58	$.041^{*}$	.136
LD × NF	26.66	$< .001^{***}$	.479

Note. LD = level difference; OR = onset rate; IR = temporal irregularity; NF = noise form. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. All remaining two-way and higher-order interactions were non-significant.

to account for small level variations. For both Hammer and Smith, paired-samples t-tests restricted to the retained irregular conditions showed significantly higher impulsiveness and annoyance ratings for LD25 than for LD15 (all $t(29) \geq 6.58$, Bonferroni-corrected $p < .001$). Spearman correlations were calculated between the mean participant ratings and the corresponding model predictions (Figure 4). Perceived impulsiveness showed the strongest association with ISO/PAS 1996-3 ($\rho = .95$), followed by DIN 45645-1 ($\rho = .86$). Annoyance correlated most strongly with the Impulsive Loudness measure ($\rho = .94$), followed by the Percentile Descriptor ($\rho = .88$).

4. Discussion and Conclusion

Impulsiveness and annoyance were assessed in a listening experiment with 30 participants using synthetic and modified signals that were systematically varied in onset rate, spectral content of the impulse, temporal irregularity of impulse appearance, and impulse-to-background level difference.

The modified Hammer and Smith sounds were generally perceived as both more impulsive and more annoying than the synthetic stimuli. Within the full-factorial synthetic stimulus set, level difference (LD) and noise form (NF), reflecting the spectral content of the impulse components, showed the strongest effects. Synthetic stimuli with an LD of 25 dB were rated significantly more impulsive and annoying than those with an LD of 15 dB. Likewise, stimuli with pink-noise impulse components, containing more low-frequency energy, received higher ratings than the corresponding blue-noise stimuli. These effects should, however, be interpreted with caution because the peak-based scaling procedure introduced unintended differences in impulse sound levels between conditions. On average, the pink-noise stimuli exceeded the corresponding blue-noise stimuli by 3.31 dB, indicating that future stimulus-generation procedures should include subsequent level normalization. Nevertheless, the significant effect of LD observed for the modified real-world stimuli suggests that the influence of LD cannot be attributed solely to this scaling-related confound.

The influence of onset rate was less clear than expected from previous findings. OR had no significant effect on perceived impulsiveness based on the collected data. For annoyance, higher onset rates were associated with slightly lower ratings, indicating a trend opposite in direction to the increase in annoyance penalty reported previously [12]. This may be attributed to the longer overall duration of the impulses in the lower OR condition due to the stimulus creation process (see Fig. 1). Temporal irregularity had no significant effect on either rating dimension. As only one regular and one randomly generated irregular pattern were tested (see Fig. 2), and the temporal contrast between them may not have been sufficiently pronounced, its perceptual relevance should be examined using a broader range of temporal configurations in longer sequences. The continuous-noise references were rated as essentially non-impulsive across all level conditions (see Fig. 3), but were nevertheless rated as annoying to a certain extent, while annoyance ratings increased markedly with sound pressure level, as expected. Relative to the noise references representing the mean L_{eq} of the respective LD conditions, the impulsive stimuli often received higher annoyance ratings, consistent with earlier findings [3] that energy-equivalent sound level does not fully account for the increased annoyance associated with impulsive sounds.

In the model comparison, the impulsiveness ratings were best predicted in terms of rank correlation by the procedure specified in ISO/PAS 1996-3, which detects individual impulse onsets, followed by DIN 45645-1, the Sottek Hearing Model, and the simple level percentile-based descriptor. Impulsive loudness showed the weakest correspondence with perceived impulsiveness but provided the best prediction of annoyance. This result is plausible because loudness constitutes a major contributor to psychoacoustic annoyance (see [14]) and loudness calculation emphasized the longer OR100 impulses more in terms of instantaneous loudness and thereby contributed to its closer correspondence with annoyance. Metrics that detect individual impulsiveness assign little or no penalty to continuous noise by design, even when its level and perceived annoyance increase.

Overall, the results emphasize the relevance of psychoacoustic loudness for predicting annoyance while also demonstrating the potential of onset-related metrics for characterizing perceptual impulsiveness. Although level or loudness percentile-based method performed well for both impulsiveness and annoyance for the comparatively short signals considered in this study, onset-related approaches may be better suited to detecting individual impulses within longer and temporally more complex recordings. Future studies should apply consistent level or loudness normalization during stimuli generation, include longer signal durations and a broader range of real-world sounds to investigate in detail whether combined loudness- and onset-based models can improve the prediction of both perceived impulsiveness and annoyance.

Author contributions

Felix Hochbaum wrote this manuscript, supervised the study, and reanalysed and visualised the data. Lukas Billerbeck planned, conducted, and evaluated the experiment, including stimulus generation and initial data analysis. André Fiebig conceived the original idea, supervised the study and contributed to this manuscript. All authors approved the final manuscript.

Acknowledgements

The authors would like to thank the HEAD-Genuit Foundation for supporting this research (P-16/05-W).

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, Tech. Rep., 2011.

[2] J. M. Wunderli et al., “Intermittency ratio: A metric reflecting short-term temporal variations of transportation noise exposure,” *J. Expo. Sci. Environ. Epidemiol.*, vol. 26, pp. 575–585, 2016.

[3] V. Rajala and V. Hongisto, “Annoyance of impulsive sounds—a psychoacoustic experiment involving synthetic sounds,” in *Proc. 23rd Int. Congr. Acoust. (ICA 2019)*, 2019, pp. 7697–7702.

[4] International Organization for Standardization, *ISO 1996-1:2016*, 2016.

[5] A. M. Willemsen and M. D. Rao, *Characterization of sound quality of impulsive sounds using loudness based metric*, 2010.

[6] Nordtest, *Nordtest method NT ACOU 112.*

[7] International Organization for Standardization, *ISO/PAS 1996-3:2022, Acoustics—Description, Measurement and Assessment of Environmental Noise—Part 3: Objective Method for the Measurement of Prominence of Impulsive Sounds and for Adjustment of* $L_{\mathrm{A,eq}}$, 2022.

[8] Deutsches Institut für Normung, *DIN 45641:1990-06, Mittelung von Schallpegeln*, 1990.

[9] R. Sottek, P. Vranken, and G. Busch, “Ein Modell zur Berechnung der Impulshaltigkeit,” in *Proc. DAGA 1995*, 1995, pp. 423–426.

[10] V. Hongisto, V. Rajala, and P. Saarinen, “Prediction of subjective annoyance penalty of impulsive noise based on objective measurements,” *Appl. Acoust.*, vol. 245, p. 111 186, 2026.

[11] Deutsches Institut für Normung, *DIN 45631/A1:2010-03, Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum—Verfahren nach E. Zwicker—Änderung A1: Berechnung der Lautheit von zeitvarianten Geräuschen*, 2010.

[12] V. Rajala and V. Hongisto, “Annoyance penalty of impulsive noise—the effect of impulse onset,” *Build. Environ.*, vol. 168, p. 106 539, 2020.

[13] P. E. Shrout and J. L. Fleiss, “Intraclass correlations: Uses in assessing rater reliability,” *Psychol. Bull.*, vol. 86, pp. 420–428, 1979.

[14] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*, 3rd ed. Berlin: Springer, 2007.