# Annoyance penalty of impulsive noise – the effect of impulse onset

Ville Rajala[1] and Valtteri Hongisto[1]

[1] Turku University of Applied Sciences, Turku, Finland.
ville.rajala@turkuamk.fi; valtteri.hongisto@turkuamk.fi

## Abstract

Impulsive sound can be perceived more annoying than a steady-state sound having the same A-weighted equivalent sound pressure level, $L_{Aeq}$. The difference in perceived noise annoyance can be compensated by adding a penalty or an adjustment $k$ to $L_{Aeq}$ (rating level). Many legislations apply a constant penalty value, such as 5 dB or more, but the validity of this procedure has been questioned. Nordtest method NT ACOU 112 identifies an impulse from the time profile of sound pressure level by using two measures describing the onset of an impulse: level difference ($D_L$) and onset rate ($R_{on}$). The purpose of this study was to determine how the annoyance penalty depends on $D_L$ (5–40 dB) and $R_{on}$ (5–800 dB/s) and to compare obtained results to the penalty prediction model of Nordtest method. A psychoacoustic laboratory experiment of 32 participants was conducted. Synthetic and periodic impulsive sounds were studied with two alternative spectra. The sounds were presented at 55 dB $L_{Aeq}$. Steady-state sounds at levels 49 – 70 dB were used to derive the penalty of impulsive sounds. The observed penalty values ranged between 0 and +8 dB. The penalty values depended somewhat on spectrum. The penalty deviated from zero when $D_L$ > 10 dB or $R_{on}$ > 15 dB/s and increased with increasing $D_L$ and $R_{on}$. The penalty predicted by Nordtest method usually overestimated the observed penalty when $R_{on} \geq 200$ dB/s. The results are against constant penalty values and they can be used to develop future penalty schemes.

## 1 INTRODUCTION

National regulations for environmental noise have been prescribed to achieve sufficiently comfortable and healthy living environments. A typical limit value for the A-weighted equivalent sound pressure level (SPL) of environmental noise inside residential buildings is 35 dB $L_{Aeq,07–22}$ during daytime and 30 dB $L_{Aeq,22–07}$ during night time, e.g. [1]. For comparison, regulated level for building services is 28 dB $L_{Aeq}$ in living rooms and 33–38 dB in other rooms within a dwelling, e.g. [2]. Very similar target values are used in the regulations of many European countries.

The regulations do not concern only the values of $L_{Aeq}$. Many regulations involve a penalty, $k$ [dB], which is added to the measured or predicted $L_{Aeq}$ in order to counteract the negative effect of a *specific feature* of the sound on annoyance. Frequently used synonyms for penalty are e.g. adjustment, sanction, bonus, allowance, and surplus. The outcome is also called rating level, $L_r = L_{Aeq} + k$ in e.g. ISO 1996-2 standard [3]. The most typical specific features are *impulsiveness* and *tonality*. E.g. in Finland, the penalty values due to impulsive sound vary between 3 and 10 dB depending on the scope of the regulation [1, 2, 4]. The penalty places special noise control requirements especially for residential buildings where the noise regulations are the tightest but also to other building types, such as schools, hospitals, offices, and accommodation buildings. Denmark [5] and Sweden [6] apply penalty of 5 dB for impulsive environmental noise. Instead, Italy [7] applies a constant penalty of 3 dB. For comparison, Switzerland [8] applies four alternatives (0, 2, 4, and 6 dB), Germany [9] applies two alternatives (3 or 6 dB), and Great Britain [10] applies four alternatives (0, 3, 6, or 9 dB). ISO 1996-1 [3] suggests a constant penalty of 5 dB for regular impulsive noise and 12 dB for highly impulsive noise. However, the application of constant penalties are not always supported by laboratory studies [11].

Many regulations do not describe exact quantitative criteria for identifying the impulsiveness in sound. For example, ISO 1996-1 standard [12] does not present a mathematical method to identify and categorize impulses. At least three kinds of methods are used to identify impulsiveness in sound: subjective methods, simple rating methods, and sophisticated rating methods. Subjective methods are based on subjective assessment of the measurer. If the sound contains rapid audible level changes, the sound is impulsive. Simple ratings are based on built-in time and frequency weightings of standard sound level meters. One well-known rating quantity is $L_{AI}$ - $L_{AS}$ where S and I denote the Slow and Impulse time weightings. The larger the value, the stronger is the impulsivity. A European council directive [13] suggested that noise is impulsive if $L_{AI}$ - $L_{AS}$ ≥ 4 dB. However, the review of Rice [14] suggests that the quantity is not able to detect all impulses. Furthermore, I weighting is no longer supported in new sound level meters. Rice [14] suggested that quantities based on the analysis of time series of the signal using time weighting $L_{Aeq,10ms}$ showed promising results in the indication of impulsiveness. Furthermore, Rice [14] suggested that the maximum positive differences between successive values of $L_{Aeq,10ms}$ should be determined.

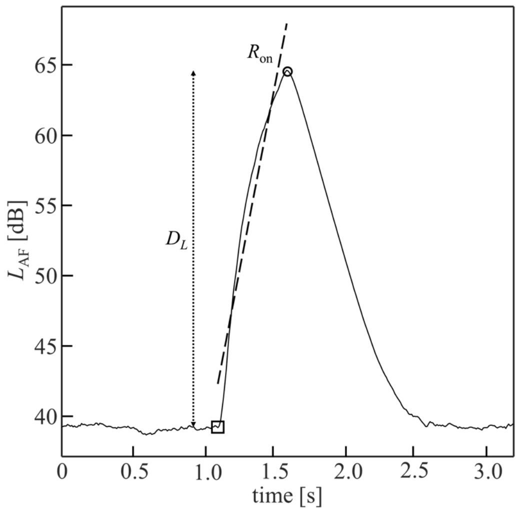


**Fig. 1**. An example of the determination of $R_{on}$ and $D_L$ of an impulse (*sound* I52, see **Table 1**). The solid line is the A-weighted SPL versus time using Fast time weighting, $L_{AF}$. The white square is the starting point of the onset and the white circle is the end point of the onset. The $D_L$ (visualized by the dotted arrow) is the difference between the SPL of the end point and the SPL of the starting point. The dashed line shows the linear fit to the SPL values between the starting point and the end point of the onset. The $R_{on}$ is the slope of the fitted line. In this case, $D_L$ = 25 dB and $R_{on}$ = 52 dB/s.

A more sophisticated rating method, which is partially meeting the suggestions of Rice [14], was introduced in NT ACOU 112 method [15]. It describes how the subjective prominence of impulsive sound could be objectively assessed by analyzing both the strength and the growth of the impulse. Nordtest method [15] describes an impulse onset by using two measures: level difference, $D_L$, [dB] and onset rate, $R_{on}$ [dB/s]. The measures are described in **Fig. 1** where the A-weighted SPL time profile is analyzed with Fast time weighting, $L_{AF}$. The method suggests that the sampling of $L_{AF}$ is made using 10–25 ms time window, which conforms with the suggestion of Rice [14]. The starting point of an impulse is the first point where the first order linear regression slope between two consecutive SPLs is over 10 dB/s. The end point of the impulse is the first point where the slope between two consecutive SPLs is below 10 dB/s. $D_L$ is calculated by subtracting the SPL of the end point from the SPL of the starting point. The $R_{on}$ is calculated by fitting a first order regression line to SPLs between the starting and the ending point. The $R_{on}$ is the slope of the regression line. It is noticeable that the method defines an impulse to be the onset of an impulsive sound, avoiding taking into account the decay of the SPL.

Nordtest method [15] defines the unitless predicted prominence, $P$, of the impulse by

$$P = 3\cdot\lg(R_{on}) + 2\cdot\lg(D_L). \quad (1)$$

Penalty $k$ [dB] is determined according to

$$k = 1.8\cdot(P - 5), \text{ for } P > 5 \quad (2)$$

if $R_{on}$ > 10 dB/s. Otherwise, $k$ = 0 dB.

According to the criteria of Nordtest method [15], impulsiveness is not limited to gunfire or shooting noise, where many previous studies on impulsiveness penalty are focused, e.g. [11, 16]. Impulsiveness exists in many kinds of environmental sounds, such as construction noise, music, shooting, explosion, wind turbine noise [17], and several logistic and industrial operations. Impulsivity also exists in numerous daily sounds whenever two items impact each other, such as door closing, hand clapping, dropping of items, log splitting, or walking. Rapid fluctuations in traffic noise have been considered to be impulsive [18]. 'Human voice contains also strong temporal variations.' For example, a recording containing the previous spoken sentence in quotations includes altogether 23 onsets where $R_{on}$ > 10 dB. Eleven of them also lead to a penalty according to **Eq. (2).** The penalties varied between 0.6 and 9.0 dB.

There is reasonably little scientific research concerning the adequate penalty of impulsive sound. Vos and Smoorenburg [11] found that the penalty for impulsive gunfire noise or impulsive metal construction noise was 12.5 dB when the background noise was absent, 10 dB when background noise was 35 dB $L_{Aeq}$, and 5.5 dB when background noise was 55 dB $L_{Aeq}$. Vos [16] studied further the penalty for gunfire noise and summarized that the penalty reduces with increasing equivalent SPL of impulses. The penalty is 10 dB for impulses at 35 dB $L_{Aeq}$ and 0 dB for impulses at 65 dB $L_{Aeq}$. The review of Rice [14] covered several other studies and it supports the level dependent penalty: the penalty varies between 10 dB (low levels, e.g. 50 dB $L_{Aeq}$) and 0 dB (high level, e.g. 80 dB $L_{Aeq}$). The impulses in the previous studies were very prominent (high values of $D_L$ and $R_{on}$) and the findings may not be directly applicable for impulsive sounds having a significantly slower onset rate, such as those mentioned above.

To our knowledge, only Pedersen [19] has studied experimentally how the penalty depends on the onset of the impulse, i.e. $D_L$ and $R_{on}$. His model was adopted to Nordtest method [15]. However, Pedersen [19] suggested that a more thorough study is desirable since the penalty model was partially based on general experience and using only industrial noise types. Better knowledge of the variables affecting the penalty of impulsive sounds could improve our understanding about the annoyance of environmental noise and building service noise in residential environments.

The investigation of the effect of impulse onset on penalty is topical since a new standardization working group (ISO TC 43 / SC 1 / WG 45) has been launched in 2019. It attempts to develop a standard for the objective identification of impulsive sounds. Nordtest method [15] is already implemented in a British standard of assessment of environmental sound [10].

Impulsive sounds carry also spectral information. Hongisto et al. [20] found in a psychoacoustic experiment that wide-band sounds having strong emphasis on high frequencies were more annoying than wide-band sounds having strong emphasis on low frequencies. They conducted the study at a constant level of 42 dB $L_{Aeq}$. Because the impact of spectrum on annoyance was drastic, the spectrum may also have an impact on the annoyance of impulsive sounds. Contrary to the findings of Hongisto et al. [20], Vos [21] found that low-frequency impulses from firearms were more annoying than high-frequency impulses. We are not aware of previous studies investigating spectrally different impulses while keeping the $D_L$ and $R_{on}$ constant.

The purpose of this study is to determine how the annoyance penalty of impulsive sounds depends on the values of $D_L$ and $R_{on}$ and to compare the observations to the penalties predicted by Nordtest method [15]. This study focuses on periodic impulses and constant overall level of 55 dB $L_{Aeq}$. The study was conducted using two different spectra of impulsive sounds to investigate a possible effect of spectrum on annoyance. This study also investigates the smallest threshold values of $D_L$ and $R_{on}$ above which the impulse penalty could be suggested. Therefore, this study covers an exceptionally large range of $D_L$ and $R_{on}$ values. The experiment has a generic nature so that the results can be applied to all kinds of built environments.

# 2 MATERIALS AND METHODS

## 2.1 Participants

The participants were recruited via university mailing lists and Turku University of Applied Sciences. The requirements for the participant were: age within 20 – 45 years, Finnish as a native language, and normal hearing. It was instructed that one should not participate the experiment during a flu or any other illness. Thirty-two voluntary persons (13 men and 19 women) participated in the experiment. The participants were native Finnish speakers and their age ranged from 20 to 44 years (mean 29, standard deviation 7). The participant received a 20 euro gift token as a compensation for their participation. None of the participants was professionally related to our research group.

## 2.2 Overall design

The study was a psychoacoustic experiment with voluntary participants. Each participant listened and rated the *annoyance* of sounds in a laboratory. The study consisted of 74 synthetically created *experimental sounds*. All participants rated all experimental sounds. The experimental sounds included both *impulsive sounds* and *reference sounds* (non-impulsive sounds). The impulsive sounds had varying $D_L$ and $R_{on}$ and the reference sounds had varying $L_{Aeq}$. The independent variable was the *experimental sound* and the dependent variable was the subjective measure *annoyance*. The reference sounds were used to determine the penalty associated with the impulsive sounds. The penalty was calculated by comparing the mean *annoyance* rating of an impulsive sound to a regression line derived from the mean *annoyance* ratings of the reference sounds. The method is described in **Sec. 2.8**.

## 2.3 Experimental sounds

### 2.3.1 Description of experimental sounds

The experimental sounds are listed in **Table 1**. They consisted of reference sounds and impulsive sounds. The experimental sounds were synthetic and they were presented within the one-third octave bands from 20 to 20 000 Hz.

The *reference sounds* consisted of wide-band noise. The one-third octave spectra of the reference sounds are shown in **Fig. 2**. This was shaped according to suggestion of Beranek [22] for the masking sounds to be used in open plan offices. Such a spectrum of wideband sound is expected to annoy little and therefore ventilation products placed close to users, such as cooling convectors, are often designed to follow this spectrum. Therefore, we chose this spectrum shape for reference sounds and for impusive sounds with spectrum S1. This choice does not imply that our study deals specifically with open plan offices but annoyance of impulsive sounds in general. The reference sounds were presented at levels from 49 to 70 dB $L_{Aeq}$ in 3 dB steps.

*Impulsive sounds* consisted of two components: steady-state wide-band background sound (spectrum S1) and seven impulses (either spectrum S1 or S2, **Fig. 3**). The background sound was continuously presented behind the impulses. The length of an impulsive sound was always 18.5 seconds. The impulses were presented periodically: an impulse began always 2.5 seconds after the beginning of the previous impulse. The first impulse occurred one second after the beginning of the impulsive sound. The duration of the onset was limited to one second because the typical maximum duration of an impulse is less than one second according to ISO 1996-1 standard [12]. All impulsive sounds were presented at an overall level of 55 dB $L_{Aeq}$. The SPL of the background sound under the impulsive sound varied experimental sounds to maintain SPL of 55 dB $L_{Aeq}$.

The acoustic descriptors describing the onset of the impulsive sounds were $D_L$ (7 nominal levels), $R_{on}$ (10 nominal levels), and *spectrum* of the impulse (S1 and S2). S1 corresponds to the spectrum of the reference sounds. Spectrum S2 corresponded to white noise. It is expected to be more annoying than the spectrum S1 based on Hongisto et al. [20].

It is notable that experimental sounds I1 and I34 (see **Table 1**) did not fulfill the criteria of Nordtest method [15], that $R_{on}$ should exceed 10 dB/s. The purpose of this study was to explore the criteria when the impulsiveness affects annoyance. It is not clear if $R_{on}$ > 10 dB/s is a suitable lower boundary for an impulse to be associated with an annoyance penalty. Therefore, the range of variables was extended beyond those of Nordtest method [15].

Impulses having large $D_L$ and small $R_{on}$ could not be used because the duration of the onset would have exceeded one second which is not in accordance with ISO 1996-1 [12]. On the other hand, impulses having concurrently small $D_L$ and large $R_{on}$ were not achievable. This is because the Nordtest method [15] uses Fast time weighting for the analysis of A-weighted SPL. Fast time weighting equals to integration time of about 100 ms. Impulses having small $D_L$ and high $R_{on}$ would have onset duration significantly less than 100 ms. Onsets with duration considerably less than the time weighting lead to low $R_{on}$ values. This is because the calculation of SPL includes signal from the onset and the decline of an impulse.

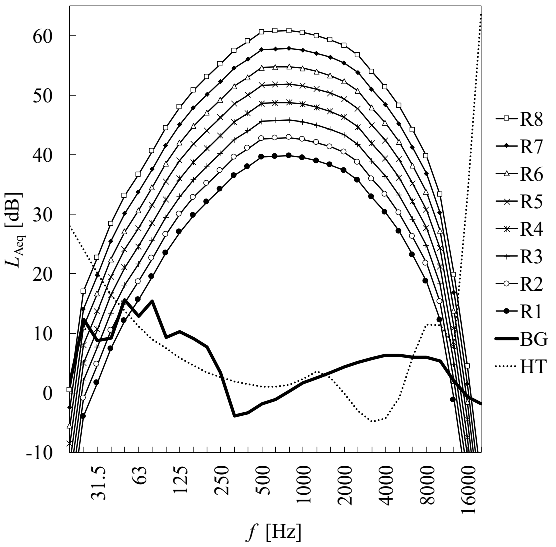


**Fig. 2.** The A-weighted equivalent SPL, $L_{Aeq}$, as a function of frequency, *f*, for the reference sounds (R1 – R8), the background noise level in the experimental room (BG), and the standardized hearing threshold (HT) according to ISO 389-7 [23].

We wanted in the first place that our study would cover a wide range of $D_L$ and $R_{on}$ values. The reasons explained above prevented adopting a full factorial design where all $D_L$ values are permutated with all $R_{on}$ values. Including a narrower range of $D_L$ and $R_{on}$ values would have allowed a full factorial design. However, in that case, the experiment would not have represented as wide range of $D_L$ and $R_{on}$ as it now did. In addition, the reference sounds (which vary with respect of $L_{Aeq}$ and not with respect of $D_L$ and $R_{on}$) would not fit to the full factorial design.

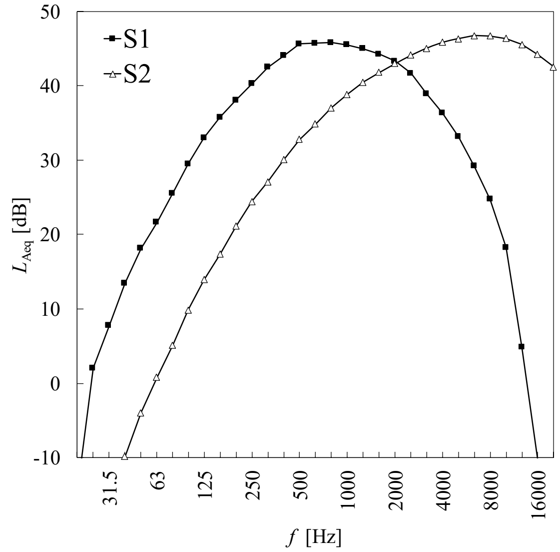


**Fig. 3.** The A-weighted equivalent SPL, $L_{Aeq}$, as a function of frequency, *f*, for spectra S1 and S2. Both spectra are presented at the overall level of $L_{Aeq}$ = 55 dB.

### 2.3.2 Creation of sounds

The sounds were created digitally (MATLAB R2017b, MathWorks Inc., Natick, MA, USA). The reference sounds were shaped from pseudo-random noise in MATLAB by using graphicEQ from Audio System Toolbox. The impulses were created by multiplying pseudorandom noise with a function. For the onset of an impulse the sound pressure increased linearly. The onset was created by multiplying the noise with the function:

$$y_{\text{onset}}(t) = \frac{f_s}{n_{\text{on}}} \cdot 10^{\frac{D_L}{20}} \cdot t, \qquad (3)$$

where $f_s$ [Hz] is the sampling frequency, $D_L$ [dB] is the set level difference, $n_{on}$ is the number of the samples included in the onset, and $t$ [s] is time. The time was discretized by

$$t = (0, 1, 2, \dots, n_{\text{on}}) \cdot \frac{1}{f_s}. \qquad (4)$$

The number of samples in the onset is

$$n_{\text{on}} = \frac{D_L}{R_{\text{on}}} \cdot f_s, \qquad (5)$$

where $R_{on}$ [dB/s] is the set onset rate.

After achieving the top of the onset the sound pressure was set to decay exponentially. The exponential decay mimics the decay of an impulse generated by a collision of two solid objects. The decay of sound pressure was created by multiplying the noise with the function:

$$y_{\text{decay}}(t) = 10^{\frac{D_L}{20}} \cdot e^{-B \cdot t}, \qquad (6)$$

where $B$ [1/s] is the damping factor describing the speed of the decay. The time was discretized by

$$t = \left(0, 1, 2, \dots, n_{\text{decay}}\right) \cdot \frac{1}{f_s}, \qquad (7)$$

where $n_{decay}$ is the number of samples included in the decay of the impulse. The number of samples can be calculated from the equation

$$n_{\mathrm{decay}} = -\ln(0.2 \cdot 10^{-\frac{D_L}{20}}) \cdot \frac{f_s}{B}. \quad (8)$$

The applied damping factor varied between the impulsive sounds. The damping factor was selected based on the subjective evaluation of three researchers so that the decay sounded as a natural pair for the onset and the decay was expected to affect the *annoyance* ratings as little as possible. It was decided that the impulses with long onset time had a smaller damping factor and impulses with shorter onset time had a greater damping factor. The damping factors measured from the experimental sounds are given in **Table 1**.

The created impulse was summed to the background sound. Because the impulses were made by modifying random noise it was not possible to exactly achieve the nominal values of $D_L$ and $R_{\mathrm{on}}$ (target values). However, it was maintained that the difference between measured values of $D_L$ and $R_{\mathrm{on}}$ was no more than 5 – 12 % from the nominal value. Therefore, **Table 1** involves both measured and nominal values for $D_L$ and $R_{\mathrm{on}}$. The nominal values are only used in Results. All seven impulses in an impulsive sound were generated by multiplying the created impulse and background sound six times. The sounds were high-pass filtered with a 46$^{th}$ order Butterworth filter with -3 dB point at 19 Hz and saved in standard waveform audio file format (.wav, 16 bit, $f_s$=40 kHz). **Fig. 4** presents an example of an impulsive sound.

**Table 1** a) The properties of the reference sounds. b) The properties of the impulsive sounds with spectrum S1. The background sound had a spectrum S1. c) The properties of the impulsive sounds with spectrum S2. The background sound had a spectrum S1. The $L_{\mathrm{Aeq}}$ of impulsive sounds was always 55 dB.

a)

| *Sound* | $L_{\mathrm{Aeq}}$ [dB] | Spectrum |
|---|---|---|
| R1 | 49 | S1 |
| R2 | 52 | S1 |
| R3 | 55 | S1 |
| R4 | 58 | S1 |
| R5 | 61 | S1 |
| R6 | 64 | S1 |
| R7 | 67 | S1 |
| R8 | 70 | S1 |

b)

| | Nominal | | Measured | | |
|---|---|---|---|---|---|
| *Sound* | $D_L$ [dB] | $R_{\mathrm{on}}$ [dB/s] | $D_L'$ [dB] | $R_{\mathrm{on}}'$ [dB/s] | $B$ [1/s] |
| I1 | 5 | 5 | 5 | 5 | 5 |
| I2 | 5 | 10 | 5 | 10 | 5 |
| I3 | 5 | 15 | 5 | 15 | 5 |
| I4 | 5 | 20 | 5 | 19 | 10 |
| I5 | 5 | 50 | 5 | 48 | 15 |
| I6 | 10 | 10 | 10 | 10 | 5 |
| I7 | 10 | 15 | 10 | 15 | 5 |
| I8 | 10 | 20 | 10 | 20 | 10 |
| I9 | 10 | 50 | 10 | 47 | 15 |
| I10 | 10 | 100 | 10 | 97 | 20 |
| I11 | 15 | 15 | 15 | 15 | 5 |
| I12 | 15 | 20 | 15 | 20 | 10 |
| I13 | 15 | 50 | 15 | 48 | 15 |
| I14 | 15 | 100 | 14 | 96 | 20 |
| I15 | 20 | 20 | 20 | 20 | 10 |
| I16 | 20 | 50 | 19 | 48 | 15 |
| I17 | 20 | 100 | 20 | 97 | 20 |
| I18 | 20 | 200 | 19 | 212 | 25 |
| I19 | 25 | 50 | 25 | 48 | 15 |
| I20 | 25 | 100 | 24 | 95 | 20 |
| I21 | 25 | 200 | 24 | 190 | 25 |
| I22 | 25 | 400 | 24 | 383 | 30 |
| I23 | 30 | 50 | 29 | 49 | 15 |
| I24 | 30 | 100 | 29 | 97 | 20 |
| I25 | 30 | 200 | 30 | 191 | 25 |
| I26 | 30 | 400 | 29 | 392 | 30 |
| I27 | 30 | 600 | 28 | 620 | 30 |
| I28 | 40 | 50 | 39 | 50 | 15 |
| I29 | 40 | 100 | 39 | 96 | 20 |
| I30 | 40 | 200 | 39 | 199 | 25 |
| I31 | 40 | 400 | 39 | 417 | 30 |
| I32 | 40 | 600 | 38 | 637 | 30 |
| I33 | 40 | 800 | 38 | 776 | 30 |

c)

| | Nominal | | Measured | | |
|---|---|---|---|---|---|
| *Sound* | $D_L$ [dB] | $R_{\mathrm{on}}$ [dB/s] | $D_L'$ [dB] | $R_{\mathrm{on}}'$ [dB/s] | $B$ [1/s] |
| I34 | 5 | 5 | 5 | 5 | 5 |
| I35 | 5 | 10 | 5 | 9 | 5 |
| I36 | 5 | 15 | 5 | 15 | 5 |
| I37 | 5 | 20 | 6 | 22 | 10 |
| I38 | 5 | 50 | 5 | 53 | 15 |
| I39 | 10 | 10 | 10 | 10 | 5 |
| I40 | 10 | 15 | 10 | 15 | 5 |
| I41 | 10 | 20 | 10 | 20 | 10 |
| I42 | 10 | 50 | 10 | 52 | 15 |
| I43 | 10 | 100 | 10 | 98 | 20 |
| I44 | 15 | 15 | 15 | 15 | 5 |
| I45 | 15 | 20 | 15 | 20 | 10 |
| I46 | 15 | 50 | 16 | 52 | 15 |
| I47 | 15 | 100 | 15 | 100 | 20 |
| I48 | 20 | 20 | 21 | 20 | 10 |
| I49 | 20 | 50 | 21 | 51 | 15 |
| I50 | 20 | 100 | 20 | 105 | 20 |
| I51 | 20 | 200 | 20 | 191 | 25 |
| I52 | 25 | 50 | 25 | 52 | 15 |
| I53 | 25 | 100 | 26 | 100 | 20 |
| I54 | 25 | 200 | 24 | 203 | 25 |
| I55 | 25 | 400 | 24 | 389 | 30 |
| I56 | 30 | 50 | 30 | 49 | 15 |
| I57 | 30 | 100 | 30 | 95 | 20 |
| I58 | 30 | 200 | 29 | 197 | 25 |
| I59 | 30 | 400 | 29 | 399 | 30 |
| I60 | 30 | 600 | 29 | 613 | 30 |
| I61 | 40 | 50 | 40 | 49 | 15 |
| I62 | 40 | 100 | 40 | 97 | 20 |
| I63 | 40 | 200 | 40 | 193 | 25 |
| I64 | 40 | 400 | 39 | 384 | 30 |
| I65 | 40 | 600 | 38 | 640 | 30 |
| I66 | 40 | 800 | 39 | 763 | 30 |

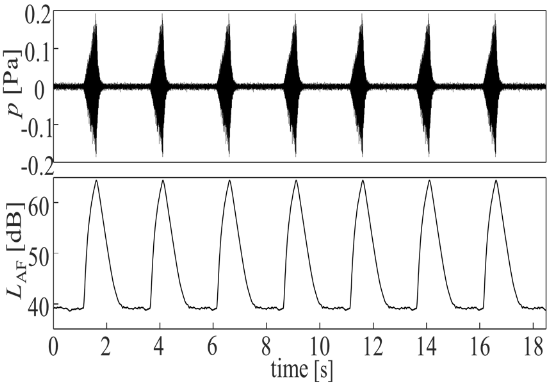


**Fig. 4.** Upper panel) The sound pressure, *p*, versus time of *sound* I52 (see **Table 1**). Lower panel) The A-weighted SPL with Fast time weighting, $L_{AF}$, versus time of *sound* I52.

## 2.4 The playback and verification of the experimental sounds

The experimental sounds were played and the data was collected by using a program coded with MATLAB. The sounds were played with a computer by using a sound card (D-audio USB Pre-Amp, Duran Audio Ltd., The Netherlands), a headphone amplifier (Brüel&Kjær ZE 0769, Denmark), and headphones (Sennheiser HD 580, Sennheiser GmbH & Co., Germany). Headphones were used instead of loudspeakers to avoid any effects of the room on perceived impulse onset. The SPL and the spectrum of each sound was verified by recording the sounds by using a head-and-torso simulator (Brüel&Kjær 4100, Denmark), a microphone power supply (Bruel&Kjaer 2804, Denmark), and a portable multitrack recorder (TASCAM DR-680MKII, Montebello, USA). MATLAB was used to measure and adjust the 1/3 octave band spectrum of the sounds to match the nominal target spectrum. The frequency dependent diffuse-field correction was applied (Brüel&Kjær Pulse Sound Quality 15.1.0, Denmark), which compensates the amplification of SPL at high frequencies caused by the artificial ear of the head-and-torso simulator.

## 2.5 Measurement of *annoyance*

The *annoyance* was measured according to the eleven-step response scale of the technical specification ISO/TS 15666 [24]. The specification presents a question suitable for socio-acoustic surveys in residential environments. The question was modified to suit the requirements of the current laboratory study. The exact question presented for the participants was similar as previously used by Oliva et al. [25] in a psychoacoustic experiment related to tonal noise: "*How much the sound bothers, disturbs, or annoys you*?" Scale value 10 was labeled as "Extremely" and scale value 0 as "Not at all". The participants were instructed to use the full scale and try to make their answers as consistent as possible. The participant had to listen to each sound for 18.5 seconds before being able to give the rating. The sound continued until the participant had responded.

## 2.6 The experimental room

The experiment was conducted in the psychophysical test room Tuuli at Turku University of Applied Sciences, Turku, Finland. The internal room dimensions were 4.5 × 2.7 m and the height was 2.5 m. The room was equipped with a desk, a chair, a monitor, a wireless keyboard, and a mouse. The computer was located outside of the room to avoid any increase in background noise level. The temperature of the room was between 23 – 24 °C during the experiments. Adequate air quality in the room was maintained with inlet ventilation rate of 27 l/s. The background noise level of the room was 23 dB $L_{Aeq}$ and the spectrum is shown in **Fig. 2**. The background noise was measured by using a precision sound level meter (Norsonic NOR150, Norsonic Co., Norway), a microphone preamplifier (Norsonic Nor1209, Norsonic Co., Norway), and a microphone (Norsonic Nor1225, Norsonic Co., Norway).

## 2.7 The experimental procedure

The experiment was conducted in May and June 2018. One participant at a time conducted the experiment by using a computer and MATLAB based software with a graphical user interface. The experiment consisted seven phases shown in **Table 2**.

Before entering the experimental room, the participants read and signed the information consent form (phase 1). The participant was informed that the scope of the study was to examine how noise should be measured so that the measurements would better correspond to the annoyance perception. Participants were informed that the sounds will not be loud and there is no risk of hearing damage or getting frightened.

After signing the consent form, the participant entered the experimental room. The participant filled an initial questionnaire (phase 2) which gathered their background information. The questionnaire inquired the participant's age, gender, and noise sensitivity, which was measured by using Weinstein's 21-item scale [26]. The noise sensitivity data is not analyzed in this article.

Each participant's hearing ability was tested (phase 3) to ensure that the hearing was normal. The hearing test was performed by using Hughson-Westlake method in frequencies 250, 500, 1000, 2000, and 4000 Hz for both ears (Madsen electronics Micromate 304). The requirement for normal hearing was that the hearing threshold was not higher than 20 dB in any of the tested frequencies. Each participant fulfilled the requirement.

In the familiarization (phase 4), the participant listened six experimental sounds that were in order R1, R8, I1, I66, I15, and I55 (see **Table 1**). The length of each sound was 18.5 seconds followed by a one second long pause. The participant did not judge the sounds in the familiarization phase. After the familiarization the participant rehearsed the rating process (phase 5). The participant listened and judged ten sounds that were in order R1, R8, I1, I33, I39, I43, I3, I20, I13, and I59 (see **Table 1**). The rehearsal phase was included to make the participant familiar with the rating procedure and to allow them to ask any questions regarding the judgment.

The ratings given in the rehearsal phase were not analyzed.

The annoyance rating phase (phase 6) represents the actual experiment. The experimental sounds were presented for the participant in one of the five predetermined pseudorandom order. The five orders were decided so that the same sound was never presented at same point in different orders. The reference sounds were distributed so that they were never consecutive. The rating phase started with two dummy sounds R1 and R8, which were not analyzed.

After the experiment (phase 7) the participant received a gift token and a short introduction of the goals and impacts of the conducted experiment. The participant had a chance to ask any questions related to the experiment. The participants stayed in the laboratory on average 65 min.

The researcher was in the experimental room with the participant during initial questionnaire, hearing test, familiarization, rehearsal, and ending. The researcher leaved the room during phase 6 and followed the experiment from a monitor showing a cloned view of the participant's display. The participant was informed that the researcher had a cloned view of the participant's monitor.

The ethical board of Finnish Institute of Occupational Health has accepted the research plan concerning this experiment (ETR 6/2015 5.11.2015).

**Table 2** The phases of the experiment and their typical durations.

| Phase | Duration [min] | Description |
|---|---|---|
| 1 | 5 | Information consent form |
| 2 | 5 | Initial questionnaire |
| 3 | 10 | Hearing test |
| 4 | 3 | Familiarization |
| 5 | 5 | Rehearsal |
| 6 | 35 | Annoyance rating |
| 7 | 2 | Gift token and feedback |

## 2.8 Analysis

The data of three participants out of 32 was excluded from the further analyses. One participant mentioned spontaneously about the absence of some sounds after the experiment. The data analysis showed that the participant had responded zero to several sounds while the other responses correlated well with other participants′ mean responses. This confirmed that this specific session suffered from technical problems. In addition, two of the participants gave annoyance responses which correlated badly with the mean of all responses. Furthermore, these two participants also gave more than 13 responses which were rated as outliers. **Supplementary data** provides a detailed description of the method used for excluding participants' data. The statistical analyzes were conducted for the data of twenty-nine participants. The normality distribution test, correlation coefficients and related $p$-values were determined using SPSS (IBM SPSS Statistics for Windows, Version 25.0. IBM Corp., Armonk, NY, USA).

Seventy-six percent of the annoyance responses (56 out of 74 sounds) were normally distributed ($p$ > 0.05, Shapiro-Wilk test). However, the absolute values of both skewness and kurtosis were below 2 for all 74 sounds, so the distributions of annoyance responses did not deviate much from normal distribution.

The *annoyance* penalty caused by impulsivity was determined by the method developed by Oliva et al. [25]. It is specifically developed for experimental designs where the investigated sounds have a constant $L_{\mathrm{Aeq}}$ and the reference sounds have varying level: the penalty for a specific sound is determined by searching the equally annoying $L_{\mathrm{Aeq}}$ of a reference sound. Thus, a full factorial design is not needed unlike with ANOVA.

The method is depicted in **Fig. 5**. The penalty of an impulsive sound was determined by using a regression line fitted over the mean annoyance ratings of reference sounds. The penalty value $k$ of an impulsive sound ($L_{\mathrm{Aeq}}$ = 55 dB) was the number of decibels that should be added to the SPL of the reference sound R3 ($L_{\mathrm{Aeq}}$ = 55 dB) so that the reference sound would be perceived equally annoying as the impulsive sound.

Next, the required variables for the annoyance penalty calculation are defined. The *annoyance* ratings for reference sound $i$ = 1… $n_r$ are defined as $\boldsymbol{y}_{\mathbf{R}i} = (y_{\mathrm{R}i1}, y_{\mathrm{R}i2}, \dots, y_{\mathrm{R}in})$, where $y_{\mathrm{R}in}$ is the rating given for the $i$:s reference sound by the participant $n$. The A-weighted SPLs of the reference sounds are defined as $\boldsymbol{L}_{\mathbf{RAeq}} = (L_{\mathrm{RAeq1}}, L_{\mathrm{RAeq2}} \dots, L_{\mathrm{RAeq}n})$, where $L_{\mathrm{RAeq}n}$ is the A-weighted SPL of the reference sound $n$. The *annoyance* ratings for an impulsive sound are defined as $\boldsymbol{y} = (y_1, y_2, \dots y_n)$, where $y_n$ is the rating given for the impulsive sound by the participant $n$.

The penalty calculation is executed in five steps:

1. The mean annoyance $\overline{y_{\mathrm{R}i}}$ for every reference sound $i$ is calculated from all ratings given for the reference sound.
2. The mean annoyance $\bar{y}$ for every impulsive sound is calculated from all ratings given for the impulsive sound.
3. *Annoyance A* is calculated by making a first order linear fitting to $\overline{y_{\mathrm{R}i}}$ and $\boldsymbol{L}_{\mathbf{RAeq}}$, $i$=1…$n_R$: $A = a \cdot x + b$, where $x$ is the A-weighted SPL, and $a$ and $b$ are the coefficients of the fitted line.
4. The A-weighted SPL corresponding $\bar{y}$ (the apparent SPL) is determined using the coefficients from the fitted line: $L_{\mathrm{Aeq}}{}' = (\bar{y} - b)/a$.
5. The penalty $k$ is achieved by reducing the actual A-weighted SPL of the impulsive sound (55 dB) from the apparent A-weighted SPL: $k = L_{\mathrm{Aeq}}{}' - 55$ dB.

The 95% confidence intervals were calculated for the *annoyance* responses and penalty values. The penalty $k$ was considered statistically significant if $k$ = 0 dB was not included within the 95% confidence intervals. The penalty analysis was conducted by using MS Excel (Version 2016, Microsoft Corporation, USA).

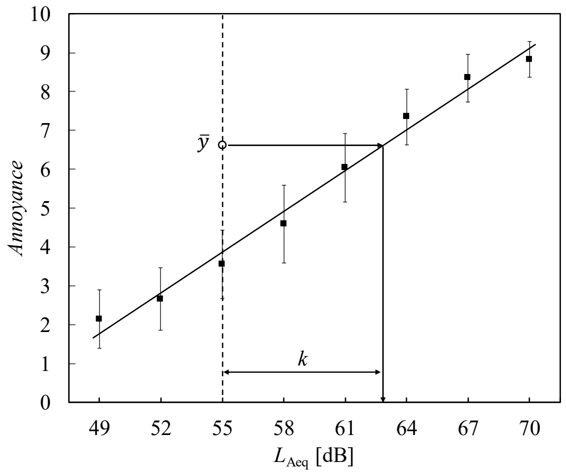


**Fig. 5.** An example of the determination of the penalty value *k* for *sound* I32 (see **Table 1**). The mean *annoyance* over all participants for *sound* I32, $\bar{y}$, is expressed as the white circle. The mean *annoyance* of the reference sounds R1–R8 (see **Table 1**) are expressed as the black squares and the whiskers indicate the 95% confidential intervals. The solid line shows the linear fit to the mean *annoyance* of the reference sounds. The equation of the fit is $y = 0.350 \cdot x - 15.367$. The squared Pearson's correlation coefficient is $r_p^2=0.986$. The $L_{Aeq}$ value corresponding to the same *annoyance* on the fitted reference line as $\bar{y}$, i.e. the apparent SPL, is marked with the down pointing arrow. The penalty value *k* for sound I32 is the difference between the apparent SPL and the actual SPL (55 dB) of the impulsive sound. In this case, $k = 7.9$ dB.

# 3 RESULTS

**Fig. 5** shows the mean *annoyance* and confidence intervals for the reference sounds as a function of $L_{Aeq}$ and the regression line fitted to the mean *annoyance* of the reference sounds. The observed mean *annoyance* of the reference sounds increased monotonically with increasing $L_{Aeq}$, thus providing a robust reference for calculating the penalty of the impulsive sounds. The mean *annoyance* and the 95% confidence intervals for *sounds* I1 – I66 are shown in **Fig. 6**. Overall, higher values of $D_L$ and $R_{on}$ yield to higher mean *annoyance* than low values of $D_L$ and $R_{on}$. The penalty values *k* and the corresponding 95% confidence intervals for *sounds* I1 – I66 are shown in **Table 3**. The values were determined as described in **Sec. 2.8**. **Fig. 7** shows the penalty values for *sounds* I1 – I66 as a function of $R_{on}$. The penalty values show increasing trend with increasing $D_L$ and $R_{on}$. Finally, the penalty observed for spectrum S1 is compared to the penalty predicted by Nordtest method [15] in **Fig. 8**. The penalty prediction method of Nordtest method [15] usually led to larger values than those observed in **Table 3**.

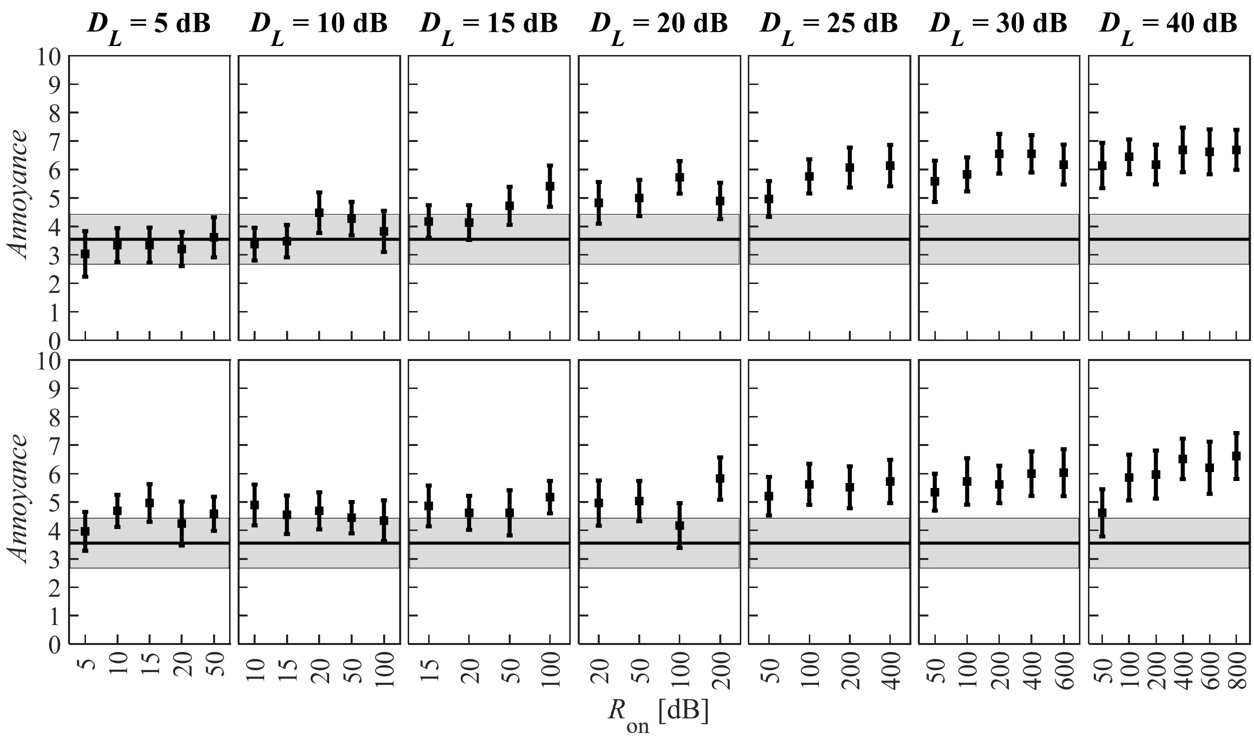


**Fig. 6.** Upper panel) The mean *annoyance* and 95% confidence intervals for *sounds* I1–I33 (impulse spectrum S1, see **Table 1**). For comparison, the mean annoyance of *reference sound* R3 (55 dB) is shown with a horizontal black line (Mean *annoyance* = 3.6). The grey area represents the 95% confidence intervals for the mean annoyance of *reference sound* R3. Lower panel) The mean *annoyance* and 95% confidence intervals for *sounds* I34–I66 (impulse spectrum S2, see **Table 1**).

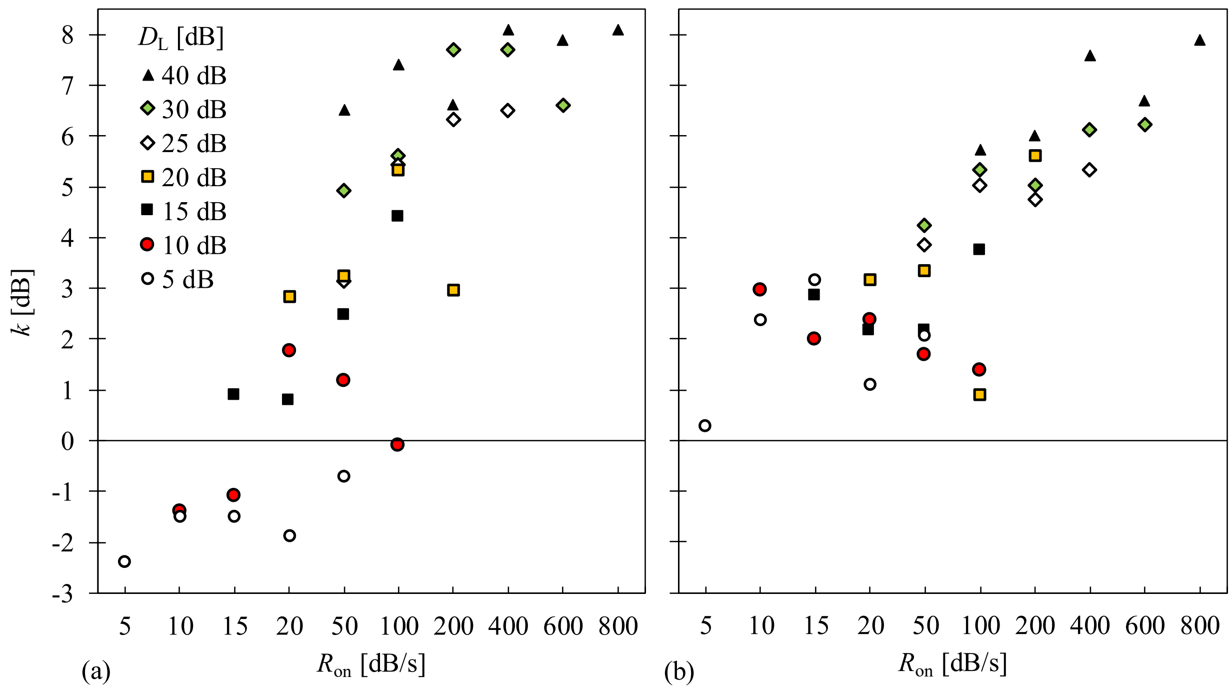


**Fig. 7.** The observed penalty $k$ as a function of onset rate of the impulse, $R_{on}$, and level difference, $D_L$, for (a) impulsive *sounds* I1–I33 (impulse spectrum S1, background spectrum S1) and (b) impulsive *sounds* I34–I66 (impulse spectrum S2, background spectrum S1).

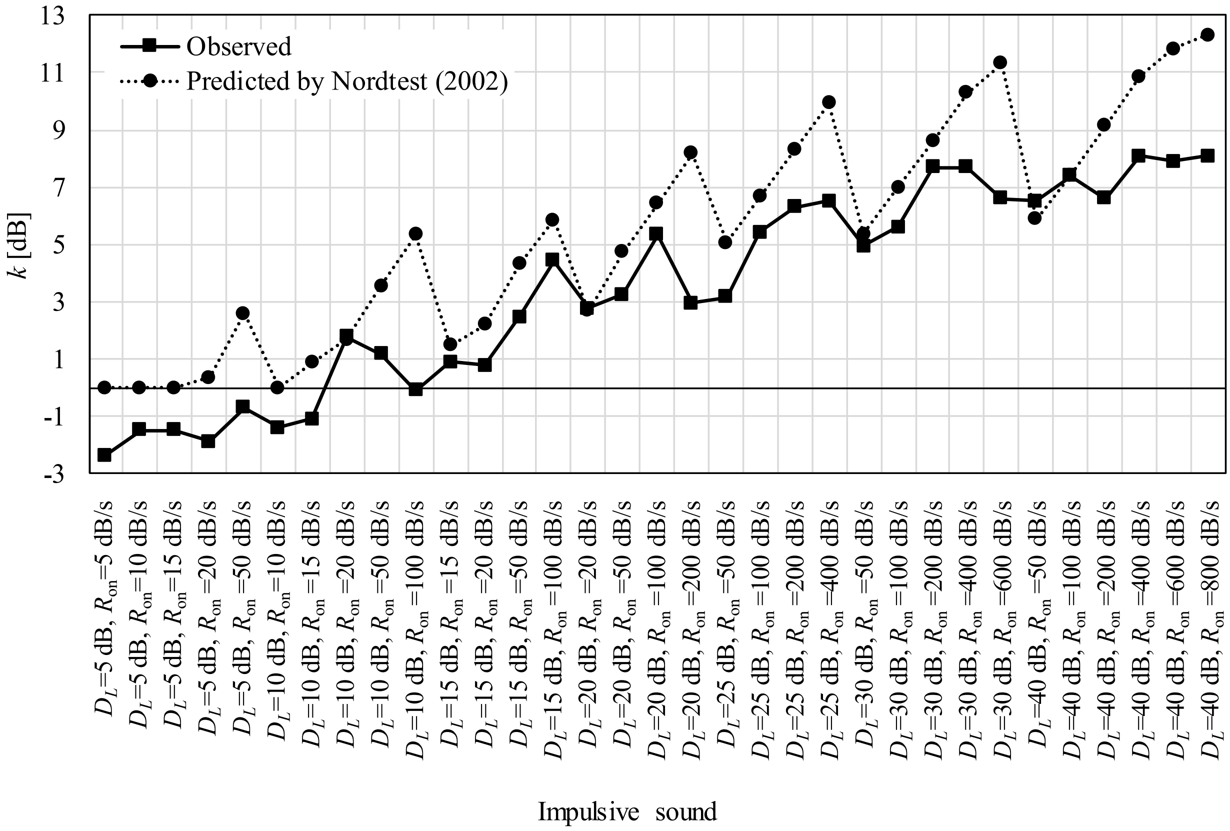


**Fig. 8.** The observed and the predicted penalty, $k$, for the impulsive *sounds* with spectrum S1. Prediction by Nordtest (2002) [15] was conducted according to Equations (1) and (2).

**Table 3** a) The penalty values and the 95% confidence intervals (CI) for *sounds* I1–I33 (impulse spectrum S1). The upper line is the penalty value *k* and the lower line is the corresponding 95% CI. b) The penalty value *k* for *sounds* I34–I66 (impulse spectrum S2) and the 95% CI. The statistically significant penalty values are bolded.

| a) | $R_{on}$ [dB/s] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $D_L$ [dB] | 5 | 10 | 15 | 20 | 50 | 100 | 200 | 400 | 600 | 800 |
| 5 | **-2.4**<br>-4.7 – -0.1 | -1.5<br>-3.2 – 0.2 | -1.5<br>-3.2 – 0.3 | **-1.9**<br>-3.6 – -0.2 | -0.7<br>-2.7 – 1.3 | | | | | |
| 10 | | -1.4<br>-3.0 – 0.3 | -1.1<br>-2.7 – 0.5 | 1.8<br>-0.3 – 3.8 | 1.2<br>-0.5 – 2.9 | -0.1<br>-2.2 – 2.0 | | | | |
| 15 | | | 0.9<br>-0.8 – 2.5 | 0.8<br>-1.0 – 2.5 | **2.5**<br>0.6 – 4.4 | **4.4**<br>2.4 – 6.5 | | | | |
| 20 | | | | **2.8**<br>0.7 – 4.9 | **3.3**<br>1.4 – 5.1 | **5.3**<br>3.7 – 7.0 | **3.0**<br>1.1 – 4.8 | | | |
| 25 | | | | | **3.2**<br>1.4 – 5.0 | **5.4**<br>3.7 – 7.1 | **6.3**<br>4.3 – 8.3 | **6.5**<br>4.4 – 8.6 | | |
| 30 | | | | | **4.9**<br>2.9 – 7.0 | **5.6**<br>3.9 – 7.3 | **7.7**<br>5.7 – 9.7 | **7.7**<br>5.8 – 9.6 | **6.6**<br>4.6 – 8.6 | |
| 40 | | | | | **6.5**<br>4.2 – 8.8 | **7.4**<br>5.7 – 9.1 | **6.6**<br>4.6 – 8.6 | **8.1**<br>5.8 – 10.3 | **7.9**<br>5.6 – 10.2 | **8.1**<br>6.1 – 10.1 |

| b) | $R_{on}$ [dB/s] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $D_L$ [dB] | 5 | 10 | 15 | 20 | 50 | 100 | 200 | 400 | 600 | 800 |
| 5 | 0.3<br>-1.7 – 2.3 | **2.4**<br>0.7– 4.0 | **3.2**<br>1.3 – 5.1 | 1.1<br>-1.1 – 3.3 | **2.1**<br>0.4 – 3.8 | | | | | |
| 10 | | **3.0**<br>0.9 – 5.0 | 2.0<br>0.0 – 3.9 | **2.4**<br>0.5 – 4.2 | **1.7**<br>0.1 – 3.3 | 1.4<br>-0.7 – 3.4 | | | | |
| 15 | | | **2.9**<br>0.8 – 4.9 | **2.2**<br>0.5 – 3.9 | 2.2<br>-0.1 – 4.4 | **3.7**<br>2.1 – 5.4 | | | | |
| 20 | | | | **3.2**<br>0.9 – 5.4 | **3.4**<br>1.3 – 5.4 | 0.9<br>-1.4 – 3.1 | **5.6**<br>3.5 – 7.8 | | | |
| 25 | | | | | **3.8**<br>1.9 – 5.8 | **5.0**<br>3.0– 7.1 | **4.7**<br>2.6 – 6.8 | **5.3**<br>3.1 – 7.5 | | |
| 30 | | | | | **4.2**<br>2.4 – 6.1 | **5.3**<br>3.0 – 7.7 | **5.0**<br>3.1 – 6.9 | **6.1**<br>3.9 – 8.4 | **6.2**<br>3.9 – 8.6 | |
| 40 | | | | | 2.2<br>-0.2 – 4.5 | **5.7**<br>3.4 – 8.0 | **6.0**<br>3.6 – 8.4 | **7.6**<br>5.6 – 9.6 | **6.7**<br>4.1 – 9.3 | **7.9**<br>5.6 – 10.2 |

# 4 DISCUSSION

The observed penalty increased with increasing $D_L$ and $R_{on}$. This is in agreement with **Eq. (2)** and Nordtest method [15]. However, the dependence of penalty on $D_L$ and/or $R_{on}$ was not a simple function of $R_{on}$ and $D_L$. The results give strong evidence that the onset of the impulse, i.e. both onset rate, $R_{on}$, and level difference, $D_L$, have an effect on *annoyance* and *annoyance* penalty of impulsive sound. The results are against the constant penalty values for impulsive noise suggested by ISO 1996-1 standard [12] and many national regulations. ISO 1996-1 [12] suggests a constant penalty of 5 dB for regular impulsive noise and 12 dB for highly impulsive noise.

**Eq. (2)** suggested a penalty for the impulsive sounds of this study when $R_{on} > 15$ dB/s and $D_L > 10$ dB. These limits are in good accordance with the results of this study for impulsive sounds having spectrum S1. However, the penalties were achieved for spectrum S2 with lower values of $D_L$ and $R_{on}$ than **Eq. (2)** predicts. In general, the *annoyance* and the penalty of impulsive sounds seemed to depend on spectrum. Statistically significant positive penalty values were obtained for lower $D_L$ values for impulses with spectrum S2 than for impulses with spectrum S1. This may be partially explained by the spectrum of reference sounds which followed the spectrum S1. In fact, most of impulses deserved a penalty for spectrum S2 suggesting that the spectrum S2 itself may be more annoying than spectrum S1. Hongisto et al. [20] studied the annoyance of different wide-band spectra and found that white noise (spectrum S2) was rated among the most annoying spectra of their study while spectra which were close to spectrum S1 were among the least annoying. If the reference sounds would have had a spectrum S2, the penalty values might have been different. In addition, spectra S1 and S2 were just two examples of an infinite number of spectra. Therefore, this study does not provide a proper understanding of the effect of the spectrum of the impulse on annoyance. Thus, following conclusions concentrate on sounds having spectrum S1. We assume that the results for spectrum S1 more trustworthy describe the effect of $D_L$ and $R_{on}$ on annoyance because the spectrum was similar for both the reference sounds and the impulsive sounds. Further research is needed to understand the effect of spectrum on *annoyance* and related *annoyance* penalty.

The penalty values suggested by **Eq. (2)** were larger than our experimental findings when $R_{on} \geq 200$ dB/s. Because there is little psychoacoustic evidence behind **Eq. (2)**, it is suggested that the finding of this study could be taken into account in future development of impulsivity penalty schemes and further evaluation of Nordtest penalty model [15]. A follow-up study would be useful where a prediction model is developed to predict the experimental findings of this study better than **Eq. (2)**. The prediction model could have similar form as, for example, tonality penalty model developed by Hongisto et al. [27], which is based on the experimental work of Oliva et al. [25].

**Fig. 7** suggests a general trendline that the penalty increases with increasing $R_{on}$ and $D_L$. However, individual points deviate sometimes strongly from this general trendline. The 95% confidence interval of the penalties was typically around ±2 dB. Therefore, deviations smaller than 2 dB from the trendline can be ignored by statistical reasons. The trendlines for each $D_L$ set (each set included four or five points) are not shown (to avoid unnecessary generalization of the results) but the points deviated from these trendlines less than 2 dB except for one case (spectrum S2, $R_{on}$ = 100 dB, $D_L$ = 20 dB). Furthermore, the trendline for each $D_L$ set showed a positive slope except for spectrum S2 with set $D_L$ = 10 dB where the penalty reduced with increasing $R_{on}$. Fortunately, the penalty values for this $D_L$ set were small and the concern about this finding diminishes. In conclusion, the results are logical in general. They suggest that the investigation of the effect of the impulse onset on penalty needs a large number of $R_{on}$ and $D_L$ combinations in order to develop a proper understanding about the penalty schemes.

The penalties for spectra S1 and S2 aligned in many points but there are a couple of examples where the opposite was found. A large difference in penalty between spectra S1 and S2 (more than 4 dB) was found for $R_{on}$ = 100 dB/s and $D_L$=20 dB (impulses I17 and I50) and $R_{on}$ = 50 dB/s and $D_L$=40 dB (impulses I28 and I61). The difference is so large that it is unlikely caused by statistical reasons. The authors listened these two pairs of sounds afterwards to see if the findings could be explained. The latter finding could be explained so that the impulse with S1 spectrum may be associated with approaching avalanche or a car. Similar explanation could not be found for the former finding. These analyses are speculative but it is possible that certain onset rates with a certain spectrum may be associated with some real-life sounds and these associations may affect the annoyance ratings. Because of these associations, penalty prediction models with perfect match to all experimental sounds may be difficult to develop.

Negative penalty values were observed for spectrum S1 (low frequency sound) with $D_L$ of 5 or 10 dB and $R_{on}$ < 100 dB/s. Most of the negative penalties were not statistically significant but it is an interesting finding and may call for additional research. Negative penalty was also observed by Vos [16] for gunfire sounds. Negative penalties were also found by Oliva et al. [25] for low-frequency tonal sounds and Virjonen et al. [28] for amplitude-modulated sounds having a modulation frequency 0.25 Hz. The current study also involved weak and slow low-frequency impulses. They may be associated with sounds which are not perceived annoying, such as peaceful sea waves on the coast or wind gusts.

The largest observed penalty value was 8 dB while Nordtest [15] predicted penalty values up to 12 dB (**Fig. 8**). The difference is large. Ceiling effect [16], which is also called range effect [11], can limit the obtained annoyance values if the sounds are very annoying. It is unlikely that e.g. range effect would have limited the observed penalty values to 8 dB since the most annoying impulsive sound I33 (mean *annoyance* = 6.69) was significantly less annoying than the loudest reference sound R8 (mean *annoyance* 8.83), see **Figs. 5-6**. According to Vos [16], the impulsivity penalty of gunfire sound reduces with increasing $L_{Aeq}$. They found a penalty of 10 dB for impulses presented at 35 dB $L_{Aeq}$ and 0 dB for impulses presented at 65 dB $L_{Aeq}$. The largest penalty value of the current study was 8 dB while our equivalent level was 55 dB $L_{Aeq}$. The value is larger than expected according to Vos [16]. It would be useful to conduct the present study by using two or more overall levels to verify the combined effect of $L_{Aeq}$, $R_{on}$, and $D_L$ on annoyance penalty of impulsive sound. Based on our study, it is justified to focus on impulses with $D_L$ > 10 dB or $R_{on}$ > 15 dB/s where impulses can cause a penalty, when the overall level is around 55 dB $L_{Aeq}$.

It is challenging to determine an ultimate penalty value for impulsive sound in general, because impulsivity is not a stationary phenomenon. The present experiment was a laboratory study with homogenous synthetic sounds. A future experiment would be useful to investigate the penalties of real sounds. It is interesting if the penalty should be determined by using other or more measures than $D_L$ and $R_{on}$.

We also found some practical issues concerning the impulse detection procedure of Nordtest method [15]. The method does not state a minimum value for $D_L$ to be notified as an impulse. The method leads to a significant amount of onsets having very small level differences. Furthermore, the method defines that an onset must have $R_{on}$ > 10 dB/s and that an impulse is "a sudden onset of a sound". The definition of an impulse does not include a quantitative minimum value for an onset to be classified as an impulse.

The present study involves some limitations which are typical for laboratory experiments. First, the participants were relatively young and the results may not apply for other age groups. However, if these penalties exist for people with normal hearing, it is expected that the penalties are not larger for people with slightly reduced hearing ability. Second, synthetic impulses created from wide-band noise were used to obtain a desirable coverage of different values of $D_L$ and $R_{on}$. Different results may emerge by using real impulses with the same values of $D_L$ and $R_{on}$. Third, the present study involved only periodic impulses with beginning points separated by 2.5 seconds. This choice had to be made since the number of $D_L$ and $R_{on}$ values was large and the experiment duration shall usually not be longer than 60 minutes. The appearance of the impulses, such as duration and random period between impulses, may also affect annoyance. Vos and Smoorenburg [11] found rather similar penalties for randomly presented gunfire impulses having almost the same mean interval between the impulses but different standard deviation of the interal distributions. Fidell et al. [29] found that the noisiness of impulses is the same when the interval between the two impulses is between 33 and 1000 ms. The largest interval was 2350 ms in the present study. Although these factors were beyond the scope of the present study, they may play an important role on the annoyance of real impulses. Fourth, the 95% confidence intervals of the mean *annoyance* ratings and penalties were relatively large. Individual ratings varied quite a lot even after the removal of the three outlier participants. Oliva et al. [25] studied the penalty of tonal sounds and their confidence intervals were close to the current study, that is ±2 dB. It might be useful to present each sound twice to reduce uncertainty. However, in the current study it was not possible because the number of experimental sounds was very large. Hongisto et al. [30] studied the repeatability of direct *annoyance* ratings in a psychoacoustic experiment where six experimental sounds out of 60 sounds were presented twice. They did not find significant differences between repeated ratings, which suggests that one rating seems to be sufficient.

# 5 CONCLUSIONS

Nordtest method NT ACOU 112 [15] defines a method to identify impulses in a sound using A-weighted sound pressure level time profile with Fast time weighting. An onset is identified from the time profile when the onset rate exceeds $R_{\mathrm{on}} = 10$ dB/s. The predicted prominence of the impulse depends also on level difference $D_L$, i.e. the strength of the impulse in decibels. Nordtest defines an annoyance penalty in decibels (i.e. an adjustment to $L_{\mathrm{Aeq}}$ to obtain the rating level) based on the predicted prominence. We conducted a psychoacoustic laboratory study to examine the effect of onset rate and level difference on the *annoyance* penalty of impulses. The present experiment is unique since similar systematic studies dealing with the effect of impulse onset on *annoyance* have not been done previously. It was found that the penalty depends on $R_{\mathrm{on}}$ and $D_L$. Larger values of $R_{\mathrm{on}}$ and $D_L$ usually led to larger penalties than low values of $R_{\mathrm{on}}$ and $D_L$. It was found that the penalty may occur when $D_L > 10$ dB or $R_{\mathrm{on}} > 15$ dB/s. The observed penalties for impulses having $R_{\mathrm{on}} \geq 200$ dB/s were smaller than the penalties predicted by Nordtest method [15]. Therefore, the revision of Nordtest penalty calculation scheme [15] seems to be justified. The results of the present study can be applied in the development of penalty schemes. The present study suggests that the spectrum of an impulse may also affect the penalty. In general, more research is needed to understand how much the spectrum of sound affects the annoyance of environmental sounds.

# ACKNOWLEDGEMENTS

This study is a part of a public research project “Anojanssi” (The metrics of noise annoyance). The project was mainly funded by Business Finland (Tekes grant 828/31/2015). The other funders were Turku University of Applied Sciences, the Ministry of the Environment, the Ministry of the Social Affairs and Health, Infra Assoc., Finnish Wind Power Assoc., Environment Pool c/o Adato Energy Ltd., TuuliWatti Ltd., APL Systems Ltd., Kone plc, Nokian Tyres plc, and Wärtsilä Finland Ltd.

# REFERENCES


[1] Finnish Ministry of the Environment, Government Decision on the Noise Level Guide Values (993/1992), Helsinki, Finland.

[2] Finnish Ministry of the Environment, Decree 796-2017 of the Ministry of the Environment on the acoustic environment of buildings. 24 November 2017, Helsinki, Finland. https://www.finlex.fi/fi/laki/alkup/2017/20170796 (In Finnish).

[3] ISO 1996-2:2017. Acoustics -- Description, measurement and assessment of environmental noise – Part 2: Determination of sound pressure levels. International Organization for Standardization, 2017, Geneve, Switzerland.

[4] Finnish Ministry of Social Affairs and Health, Decree 545-2015 of the Ministry of Social Affairs and Health on Health-Related Conditions of Housing and Other Residential Buildings and Qualification Requirements for Third-Party Experts, 2015, Helsinki, Finland. Translation from Finnish: http://www.finlex.fi/en/laki/kaannokset/2015/en20150545.pdf.

[5] Delta, Environmental noise regulation in Denmark. Danish reference laboratory for noise measurements c/o DELTA, Orientering nre. 45, 17 Jan 2012, Hørsholm, Denmark.

[6] SOU 2013:57, Samordnade bullerregler för att underlätta bostadsbyggandet, Statens offentliga utredningar, ISBN 978-91-38-23996-4, 2013, Stockholm, Sweden.

[7] DPCM, Limiti massimi di esposizione al rumore negli ambienti abitativi e nell' ambiente esterno. Con le modifiche introdotte dal D.P.C.M. 14.11.97. G.U.-8.3.1991-n. 57, 1991, Council of Ministers Presidential Decree, Italy.

[8] The Swiss Federal Council, Lärmschutz-Verordnung (Noise abatement ordinance), 814.41, 15 Dec 1986, Der Bundesrat, Schweitzerische Eidgenossenschaft, Switzerland.

[9] TA Lärm, Sechste Allgemeine Verwaltungsvorschrift zum Bundes-Immissionsschutzgesetz (Technische Anleitung zum Schutz gegen Lärm - TA Lärm), 26 August 1998 (GMBl Nr. 26/1998 S. 503). VSGA 03/2000, 18 pp., Germany.

[10] BS 4142:2014, Methods for rating and assessing industrial and commercial sound. BSI Standards Ltd., 2014, London, United Kingdom.

[11] J. Vos, G.F. Smoorenburg, Penalty for impulse noise, derived from annoyance ratings for impulse and road-traffic noise. J. Acoust. Soc. Am. 77(1) (1985) 193–201.

[12] ISO 1996-1:2016. Acoustics -- Description, measurement and assessment of environmental noise – Part 1: Basic quantities and assessment procedures. International Organization for Standardization, 2016, Geneve, Switzerland.

[13] EEC, Council directive of 19 December 1978 on the approximation of the laws of the Member States relating to the determination of the noise emission of construction plant and equipment (79/113/EEC). Official Journal of the European Communities, Vol. 22, 1979, No. L 33/15–30, , European Community.

[14] C.G. Rice, Human response effects of impulse noise. J. Sound Vib. 190(3) (1996) 525–543.

[15] Nordtest method NT ACOU 112:2002 – Acoustics – Prominence of impulsive sounds and for adjustment of LAeq. Approved 2002-05, Taastrup, Denmark

[16] J. Vos, J, On the level-dependent penalty for impulse sound. J. Acoust. Soc. Am. 88(2) (1990) 883–893.

[17] C. Di Napoli, Wind turbine noise assessment in a small and quiet community in Finland. Noise Con. Eng. J. 59(1) (2011) 30–37.

[18] T. H. Pedersen, Objective method for measuring the prominence of impulsive sounds and for adjustment of LAeq. Proc. Internoise 2001, 27–30 August, 2001, The Hague, The Netherlands.

[19] T. H. Pedersen, Audibility of impulsive sounds in environmental noise. Proc. Internoise 2000, 27–30 August, 2000, Nice, France.

[20] V. Hongisto, D. Oliva, L. Rekola, Subjective and Objective Rating of Spectrally Different Pseudorandom Noises – Implications for Speech Masking Design, The Journal of the Acoustical Society of America, 137(3) (2015) 1344-1355.

[21] J. Vos, On the annoyance caused by impulse sounds produced by small, medium-large, and large firearms. J. Acoust. Soc. Am. 109(1) (2001) 244–253.

[22] L. L. Beranek, Criteria for Noise and Vibration in Communities, Buildings and Vehicles, Ch. 18, In: Noise and Vibration Control,

Ed. Beranek L. L., McGraw-Hill Book Company, 1971, New York, USA.
[23] ISO 389-7:2005 Acoustics – Reference zero for the calibration of audiometric equipment. Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions. International Organization for Standardization, 2005, Geneve, Switzerland.
[24] ISO/TS 15666:2003. Acoustics – Assessment of noise annoyance by means of social and socio-acoustic surveys. International Organization for Standardization, 2003, Geneve, Switzerland.
[25] D. Oliva, V. Hongisto, A. Haapakangas, Annoyance of low-level tonal sounds – Factors affecting the penalty, Build. Environ. 123 (2017) 404–414
[26] N.D. Weinstein, Individual differences in reactions to noise: a longitudinal study in a college dormitory. Journal of Applied Psychology 63(4) (1978) 458–466.
[27] V. Hongisto, P. Saarinen, D. Oliva, Annoyance of low-level tonal sounds – A penalty model. Appl. Acoust., 145 (2019) 358–361.
[28] P. Virjonen, V. Hongisto, J. Radun, Annoyance penalty of periodically amplitude-modulated wide-band sound. J. Acoust. Soc. Am., 146(6) (2019) 4159–4170.
[29] S. Fidell, K.S. Pearsons, M. Grignetti, D.M. Green, The Noisiness of Impulsive Sounds, J. Acoust. Soc. Am. 48(6:1) (1970) 1304-1310.
[30] V. Hongisto, D. Oliva, L. Rekola, Subjective and objective rating of the sound insulation of residential building façades against road traffic noise. J. Acoust. Soc. Am. 144(2) (2018) 1100–1112.

# Supplementary material

The main article reported that three participants out of 32 were rejected in the statistical analysis. This supplementary data clarifies the reasons for the rejection.

The participant's responses were analyzed according to the following four criteria: suspicion, Pearson's correlation coefficient, sum of squared errors, and number of outliers. Suspicion means that the participant reported something after the experiment raising a suspicion of possible invalid responses. Pearson's correlation coefficient, $r_P$, represents by a value between 0 and 1 how well the participant's response is associated with the mean of all participants' responses. It was determined for all 74 sounds between the mean of all participants' ratings and the current participant's rating. The mean value of 32 participants was $r_P = 0.60$ (**Table S1**). The violation criterion was set to $r_P$ with $p > 0.001$ (two-tailed). Four participants did not fulfill this criterion. Squared error, $E^2$, describes how much the participant's response differs from the mean of all participants' *annoyance* responses for a sound. It was determined separately for each 74 sound as a square of the difference of the participant's rating and mean of all participants' *annoyance* ratings. The sum of squared errors, $\Sigma E^2$, was calculated over all 74 sounds, thus describing the overall deviation of the participant's response from the mean of all participants' responses. The mean value of 32 participants was $\Sigma E^2 = 345$ (**Table S1**). The violation criterion was set to $\Sigma E^2 > 535$ because 27 participants out of 32 (84%) obtained values smaller than the criterion. An *annoyance* response was notified as an outlier response if the value deviated more than 3.5 from the mean response of 32 participants. The number of outliers per participant, $\Sigma O$, was determined. The total number of outlier responses was 230 out of 2368 *annoyance* responses (9.7%) (**Table S1**). The violation criterion was set to $\Sigma O > 13$ because 27 participants out of 32 (84%) indicated a smaller number of outliers than the criterion. It means that the participant reported an outlier in more than 18% of sounds, raising a doubt about e.g. lack of motivation or misunderstanding of the instructions.

The removal criteria were the following: suspicion criterion was fulfilled OR all three other violation criteria ($p > 0.001$ for $r_P$, $\Sigma E^2 > 535$, and $\Sigma O > 13$) were fulfilled.

Participant 632 reported spontaneously after the experiment that several sounds were inaudible. For that participant eight outlying *annoyance* responses were found in the data and all of them were zero ratings supporting the participant's self report. Thus, participant 632 was removed due to suspicion criterion only. Furthermore, participants 609 and 614 were removed because they did not fulfill the three other criteria. The responses of removed participants are shown in **Figure S1** together with respondent 623 whose responses were closest to the mean responses of all participant.

**Table S1.** The values of $r_P$, the p-value of $r_P$, $\Sigma E^2$, and $\Sigma O$ for the 32 participants. The bolded values mean that the criterion for the variable is violated.

| Participant | 601 | 602 | 603 | 604 | 605 | 606 | 607 | 608 | 609 | 610 | 611 | 612 | 613 | 614 | 615 | 616 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $r_P$ | 0.86 | 0.50 | 0.58 | 0.76 | 0.81 | 0.77 | 0.42 | 0.64 | **0.31** | 0.64 | 0.87 | 0.67 | 0.64 | **0.01** | 0.66 | 0.62 |
| $p$ | 5E-22 | 5E-06 | 6E-08 | 2E-15 | 1E-17 | 2E-16 | 1E-04 | 3E-12 | **0.006** | 1E-09 | 2E-22 | 2E-10 | 9E-10 | **0.924** | 2E-10 | 8E-09 |
| $\Sigma E^2$ | 138 | 276 | 151 | 215 | 351 | 107 | 148 | 184 | **1718** | **540** | 261 | **600** | 500 | **644** | 106 | 141 |
| $\Sigma O$ | 0 | 3 | 1 | 0 | 5 | 0 | 0 | 3 | **63** | **15** | 2 | **18** | 12 | **18** | 1 | 0 |

| Participant | 617 | 618 | 619 | 620 | 621 | 622 | 623 | 624 | 625 | 626 | 627 | 628 | 629 | 630 | 631 | 632 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $r_P$ | 0.72 | 0.78 | 0.62 | 0.82 | 0.78 | 0.41 | 0.88 | 0.51 | 0.38 | 0.53 | 0.76 | **0.25** | 0.63 | 0.68 | **0.18** | 0.45 |
| $p$ | 1E-12 | 1E-15 | 8E-09 | 2E-18 | 6E-16 | 0.0007 | 9E-25 | 3E-06 | 0.0009 | 2E-06 | 7E-15 | **0.026** | 6E-09 | 1E-10 | **0.087** | 6E-05 |
| $\Sigma E^2$ | 300 | 268 | 305 | 123 | 214 | 453 | 78 | 98 | 305 | 233 | **838** | 488 | 352 | 320 | 191 | 402 |
| $\Sigma O$ | 7 | 2 | 5 | 0 | 2 | 11 | 0 | 0 | 5 | 1 | **30** | 10 | 4 | 3 | 1 | 8 |

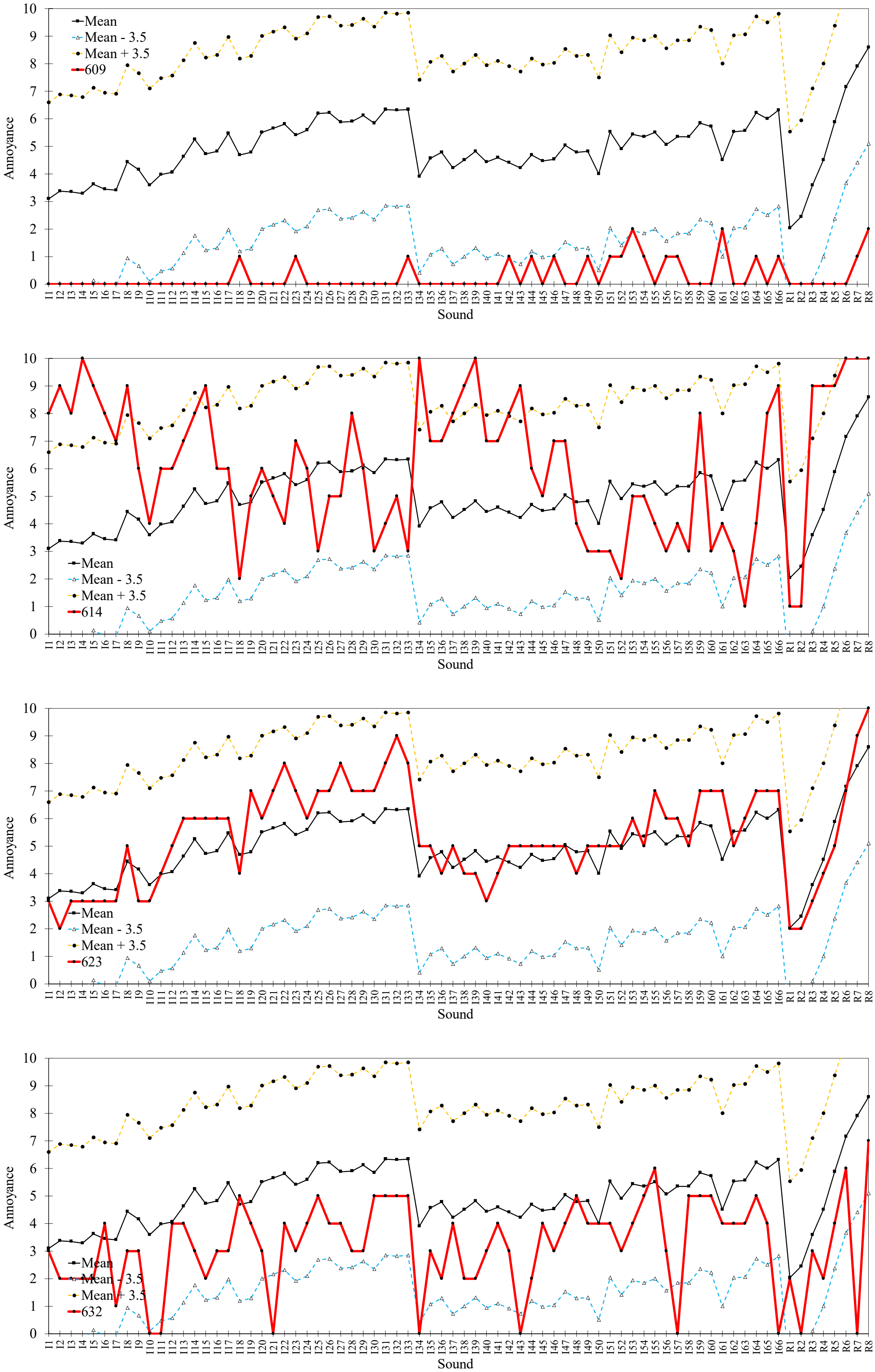


**Figure S1**. The annoyance responses of participants 609, 614, 623, and 632 for the 74 sounds. The comparison is made to the mean of 32 participants, the mean plus 3.5 annoyance units (upper outlier limit) and the mean minus 3.5 annoyance units (lower outlier limit).